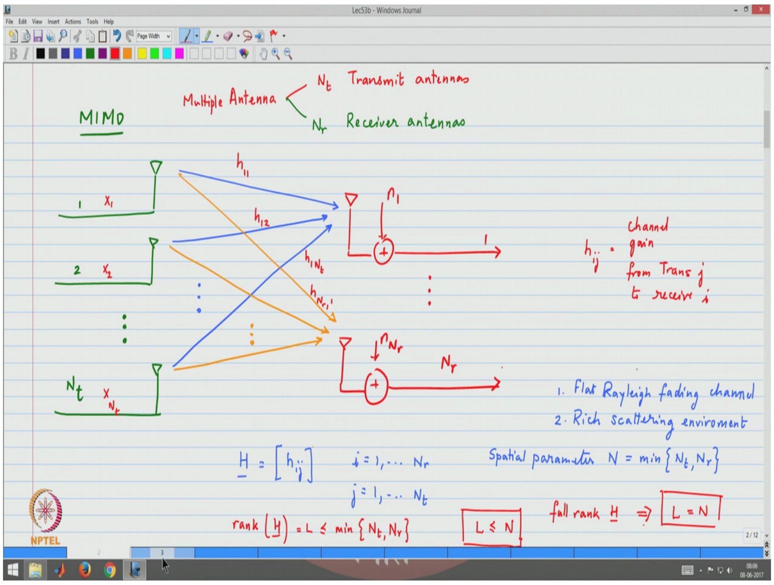
Introduction to Wireless and Cellular Communication
Prof. David Koilpillai
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture – 53

Good morning and welcome to lecture 53 in which we will be covering some of the key concepts in MIMO technology, multiple input multiple output which represents the multiple transmitters at multiple antennas that the transmitter and multiple antennas at the receiver. In the last lecture we have had an opportunity to review the basic MIMO problem formulation, we have also looked at if the in channel would were information were available at the transmitter then we could do a decomposition of the channel using singular value decomposition and then apply pre coding at the transmitter that lead us to the perspective of parallel decomposition of the channels.

We will start from there and build on our understanding of how to achieve capacity for a MIMO channel. Firstly, when the case when we can do parallel decomposition and then we will go on to the general case where we cannot do parallel decomposition, but still we would like to achieve the best in terms of the performance of the system capacity of the system and that will be based on our understanding of the entropy of the channel.

(Refer Slide Time: 01:23)

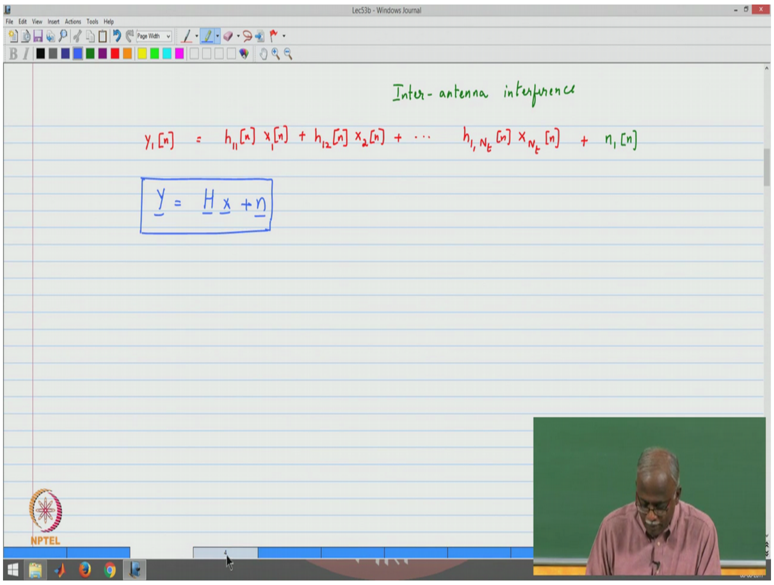


So, to begin with let us look at the MIMO context we have N t transmit antennas and we have N r transmit antennas and we have a channel that connects them. So, we have made certain assumptions about the channel. So, just for refreshing our minds I would like to write down the assumptions that we have made about the channel. The channel assumptions are that number one that it is a flat Rayleigh fading channels, it is a flat Rayleigh fading channel; that means, each of the channel transfer functions from the h ij from the jth transmit antenna to the ith receive antenna is a single Rayleigh distributed coefficient complex gain.

So, it is a flat Rayleigh fading channel. So, we have a matrix that the entire channel is represented by a matrix where each entry h ij is a really random variable. So, flat Rayleigh fading channel we have also made another assumption that it is a rich scattering environment, rich scattering environment is one in which each of because of the multiplicity of objects that create the reflections we have these channel coefficients are all independent of each other. So, they are independent and identically distributed complex Gaussians. So, a rich scattering environment and this is also a very important element of our analysis of the MIMO system.

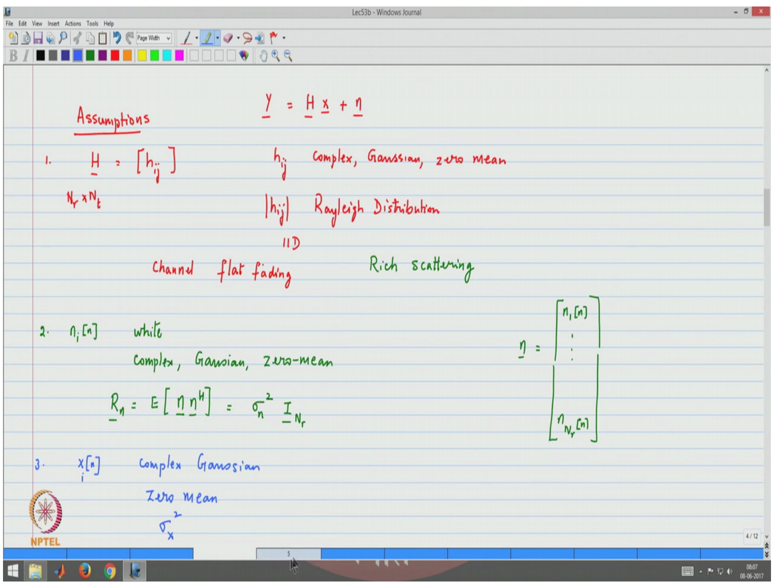
We also have not specifically introduce it, but it is important for us to recognize that the spatial parameter the one that gives us the advantage of the channels is can be indicated as N which is the minimum of the transmit and receive antenna number of transmit antennas and number of received antennas. So, that is an important indication how much of benefit we can achieve from such a system in a sense does depend on how much is our is the minimum between the transmit and receive antennas. And of course, when we write this as a channel matrix which consist of these elements h ij, i going from 1 to N r; j going from 1 to N t this matrix H has a rank. So, the rank of the matrix H is some; if we denote it as L and this will be less than or equal to the minimum of the number of channel the number of rows a number of columns which correspond to N t comma N r or in other words L is less than or equal to N and we like to consider the special case where we have full rank full rank; that means, the matrix full rank. So, therefore, full rank of the matrix H and this implies that the rank is equal to the minimum of the number of transmit antennas or receive antennas. So, again it is used for parameter that will come up in today’s discussion.

(Refer Slide Time: 04:47)



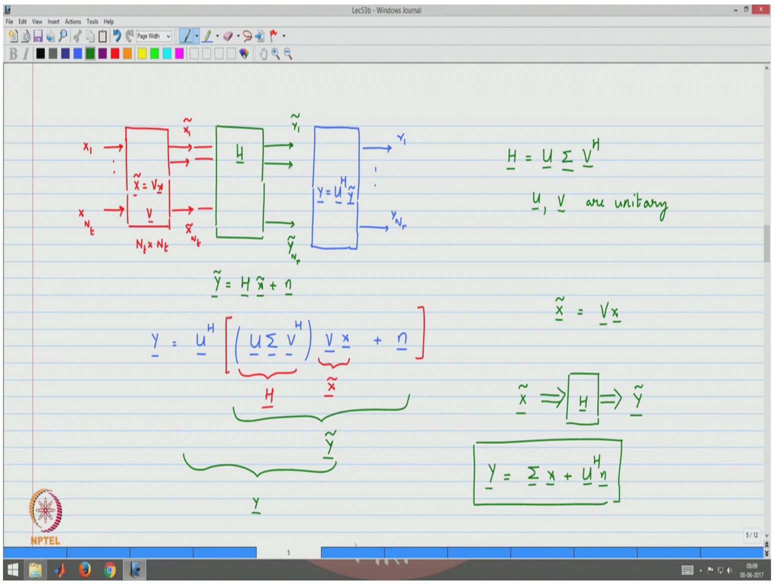
So, the expression for the each of the receive antennas I think this we have already looked at. So, it will consist of the channel gain from the corresponding transmit antenna to the receive antenna times the symbol or the signal transmitted by that we denoted as x 1 to denote the transmission from antenna 1, x 2 from antenna 2 and likewise from each of the transmit antennas and at the receiver we will also have the additive noise. So, the overall expression as we solved in the last lecture was given by Y is equal to H x plus n and we have make certain statements about the elements of H about the assumptions about the noise being white Gaussian noise and also about x being of the input.

(Refer Slide Time: 05:34)



So, the 3 assumptions you can review from yesterdays lecture let me move on to the case when we are able to do the singular value decomposition.

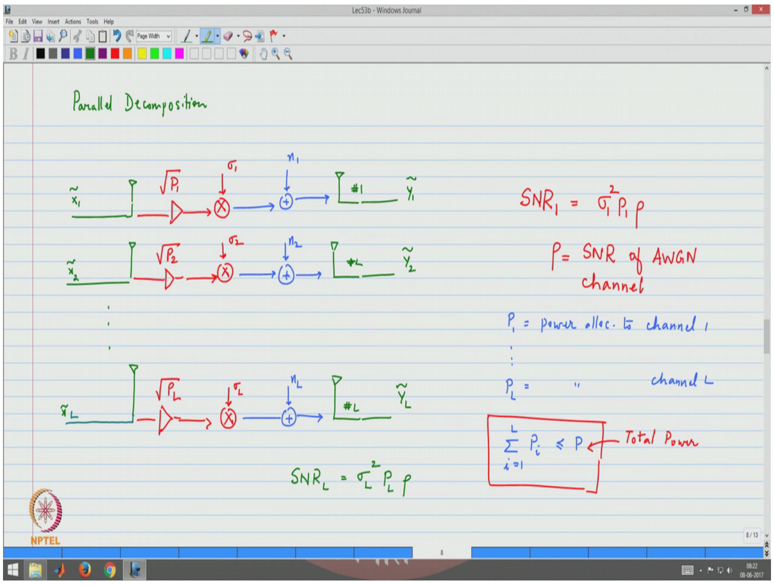
(Refer Slide Time: 05:41)



So, the singular value decomposition we write the matrix H in terms of the 2 unitary matrices u and v with the matrix sigma in between this is uv Hermitian where the matrices u and v are unitary and this is a very useful property for us because this is what helps us achieve the parallel decomposition of the channel as we saw in the last lecture. So, the modified input x tilde is given by V times x.

So, you can think of it as some transformation of the signals transmitted by the different antennas you have made a transformation a linear transformation now this x hat is passed through the channel. So, if were to think of it as a vector input to the channel H this is x tilde and produces the vector output y tilde and y tilde is given by H and H tilde plus N and then we process it subsequently using u Hermitian that gives us the final expression where we get y is equal to sigma times x plus u Hermitian times n. So, this is the final expression we recognize that these this corresponds to the expressions that are given in this form. The one correction to the diagram that we did in the last class let us call this as eta after the no actually this is correct.

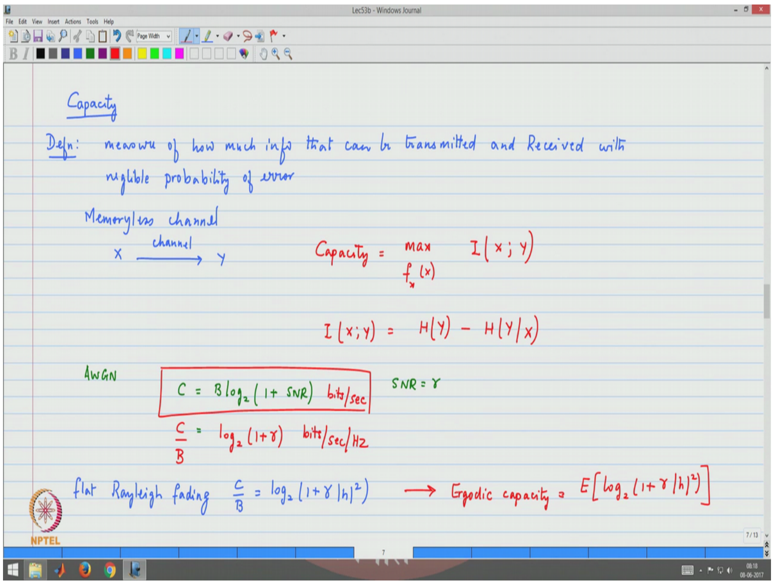
(Refer Slide Time: 07:20)



So, each of the L branches have got the corresponding noise terms being added and then once you get y 1 tilde y 2 tilde y L tilde the in other words the y tilde vector then we will process it by saying that the y vector the y y vector will be obtained by u Hermitian times y tilde.

So, that is the way we do that if we have assumed that each of these singular values have been arranged in descending order then we can also see that this correspond to channels with different SNRs where the SNR of the first channel where sigma 1 is the largest singular value basically you we find that each of them it is a ordering of channels in decreasing SNR. So, SNR 1 is greater than SNR to greater than or equal 2 and so on. So, this is our basic expression what we would like to do now is to build on this and in terms of the developments that we would like to describe. So, let me begin by quickly summarizing some of the key elements of the channel capacity.

(Refer Slide Time: 08:58)



So, we would now like to focus our attention on channel capacity. This is a parameter that we have already discussed and described it is a concept that we have already studied, but we would like to revisit this in the context of the MIMO system and introduce some additional insights into our understanding of how to achieve the maximum capacity of the wireless channel when you have the flexibility of multiples and tennis at the transmitter and at the receiver. So, channel capacity the basic definition would be that we are trying to assess or estimate what is the maximum rate at which we can transmit information. So, it is a measure of how much information of how much information that can be sent from the transmitter to receiver with negligible probability of error, how much information that can be transmitted, that can be sent from, that can be transmitted from the transmitter to the receiver with negligible probability of error. So, transmitted and received or sent from the transmitter to the receiver with negligible probability of error, with negligible probability of error.

So, that is our definition and I am sure when you have studied the basic concepts in digital communications and information theory this would have been a definition that was as studied, so negligible probability of error.

Now this is quantified in terms of quantity called the entropy. So, we would like to introduce that and build on that understanding. So, now, when we have a memory less channel where you have a whether signal is x is transmitted and y is received then we have a mem, in the context of a memoryless channel, memoryless channel where we transmit x through the channel and y is received this is through the channel. So, in such a situation we introduce the notion of capacity. So, we introduced capacity as the mutual information I being the mutual information between x and y. So, that is the information that can be transmitted through the channel and this is for all possible maximized over all possible probability distributions of x.

So, over all possible distributions of x what is the maximum mutual information that we can achieve and that is what we understand as the capacity of a memoryless channel. So, in other words mutual information between x and y can also be written in the context of the entropy where we write it as the entropy of the output H of y, entropy as you will recall is a measure of uncertainty. So, the measure of uncertainty in y minus the measure of uncertainty in y given x and this is the measure of the mutual information and this also gives us a way to compute the capacity under different channel conditions. So, given this basic understanding of capacity in terms of the mutual information and also expressed in terms of the entropy we can write down that the following results.

For an AWGN channel this is what we often referred to as Shannon capacity the capacity of a channel C is equal to B times logarithm base 2 1 plus SNR and if we denote as we have been using the notation SNR is equal to gamma this can be written as we can also you write it in terms of. So, the units for this will be bits per second and sometimes we find it more convenient to represent it in terms of a normal eyes capacity to C divided by B per unit bandwidth is given by logarithm base 2 1 plus gamma where the units will now become bits per second per hertz.

So, that is our basic definition of capacity and this is derived from first principles base using the definition of capacity in terms of mutual information in terms of entropy and this is what we often referred to as Shannon capacity. And this is the expression that we have used quite extensively when we were talking about the capacity and understanding of capacity of a wireless channel when we had 1 transmit antenna and 1 receive antenna. Now we are going to extend our understanding to the case where we have multiple transmitter multiple antenna.

So, the just to refresh your memory we also said that in Rayleigh fading supposing we had flat Rayleigh fading, flat Rayleigh fading. Then we can no longer talk about an average SNR it now becomes an instantaneous SNR. So, in terms of the Rayleigh fading if the instantaneous value of the channel is given by H then the capacity C by B is given by logarithm base 2 1 plus the instantaneous SNR and instantaneous SNR is the average SNR multiplied by the Rayleigh coefficient and the magnitude squares that affects the SNR and therefore, this is what we have. And of course, this leads us to this our understanding of argotic capacity where we rather than talking about an instantaneous value we talk about the average value or the expected value in a fading environment which is given by expected value of logarithm base 2 1 plus gamma mod h square and that was our expression for the argotic capacity.

So, given this particular framework let us 0 in on our understanding of the capacity of an AWGN channel, capacity of an AWGN channel where the capacity depends upon the SNR, the SNR in turn will depend upon the noise variance, it will also depend on the channel gain that we encounter and of course, what is the transmitted signal power. So, based on these assumptions we would now like to go back and revisit our parallel decomposition. So, in the context of the parallel decomposition if you for a moment omit the channel gain a sigma 1 and think of it as just an AWGN channel and the SNR would depend on the signal power that is transmitted over the noise power and if you denote that SNR of a AWGN channel, SNR of the AWGN channel basically without the sigma sigmas in place. So, if you denote this as rho, then the SNR of the first channel SNR 1 depend will become rho times sigma 1 squared, sigma 1 being a scalar gain and it will contribute to an amplification of the signal proportional to the energy proportional to sigma 1 squared, whatever was the original SNR in the absence of sigma 1 we now we will have rho time sigma 1 square. Similarly, you can write it down write down the SNRs for each of these.

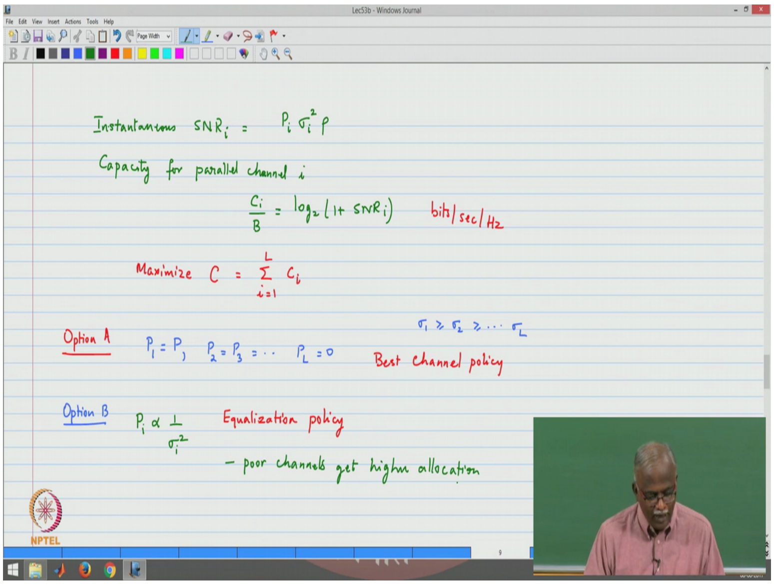
Now we would like to take it 1 step further where we also allocate power to each of the different of the L channels. So, if we allocate channel power P 1, P 1 is the power allocated power allocated to the channel 1 allocated to channel 1, P 2 to channel 2. So, on and then we have P L as the power allocated to channel L. Now this is always done under a total power constraint. So, we say that summation i is equal to 1 through L P subscript i this must be less than or equal to sum value P which is the total power that is permitted in our transmission. So, this represents the total power that we are allowed to transmit and what we are doing is we are dividing the total power into P 1 through P L where P subscript i denotes the transmitted in each of those channels. So, that is the total power that is we are allowed to transmit.

So, if you were to think of it as a gain where you are allowed to change the an additional control parameter is given to us where these values are in these values can be applied you can think of the SNR now in the fallowing way. You have the signal x you have the noise is N 1 that would with without anything else you would have had SNR of rho now when you have sigma 1 sigma 1 it becomes sigma 1 squared times rho and if you have square root of P 1 which is a scalar then you get and the total SNR become sigma 1 squared P 1 root P 1 square which is P 1 times rho and now where rho is the SNR of the AWGN channel.

So, we can now write down SNR of the channel number 1 and likewise we can write down for the for the Lth channel, SNR of the Lth channel can be given by sigma L squared P L times rho where we are we have made the assumption that he variances of all these noise terms are the same and if all of these are transmitted with the same signal power then rho is constant for all of them and what is different is only the difference is caused by the root P 1 and sigma 1 which are in the sigmas are introduced by the channel P’s are something that you can the user gets to modified based on our whatever is our criteria that we have chosen. But we must keep in mind that we must always satisfy the total power.

Now given this and given that we have an instantaneous SNR for a let us say for the ith channel, instantaneous SNR for the ith channel instant SNR i this is given by as we saw in the previous expression P i sigma i squared times rho.

(Refer Slide Time: 20:27)



Now, the capacity for this channel based on the Shannon capacity formula for channel i, for the parallel channel i, we have L parallel channels looking at 1 of them parallel channel I is given by C i divided by B is equal to logarithm base 2 1 plus SNR subscript and the units of course, will be since it is a normalize expression this will L be bits per second per hertz.

So, now our goal is to maximize, maximize the total capacity of the system which would mean maximize capacity C where C is the sum of the individual capacities of the L parallel channels. We have N t transmit antennas N t receive antennas using the singular value decomposition where L corresponds to the rank of H we have been able to divide them into parallel channels we have also done the power allocation P 1 to the first channel, P 2 to the second channel, P L subject to a total power constraint and then obtained instantaneous SNR use that to obtain the capacity of the each of the parallel channels and then we want to maximize the capacity. So, now, looking at this scenario let us look at the different options that we will have to achieve the capacities that we are looking for.

So, the first one is option a, what we call as option a, if we do the following then are we how are we doing in terms of the capacity. So, this is a allocation that we would like to do P 1 notice that we have made the assumption that sigma 1 greater than or equal to sigma 2 greater than or equal to sigma L, that is the channels are arranged in decreasing order of the SNR. So, P 1 means the power allocated to the best of the best of the parallel channels. Notice that I am going to allocate all the power to channel number 1 and of course, this would be in that P 2 equal to P 3 equal to P L will be equal to 0 . So, this is what we referred to as the best channel allocation best channel policy. So, we have L parallel channels, but we have deliberately chosen to give all the power to the best channel to we called the best channel policy.

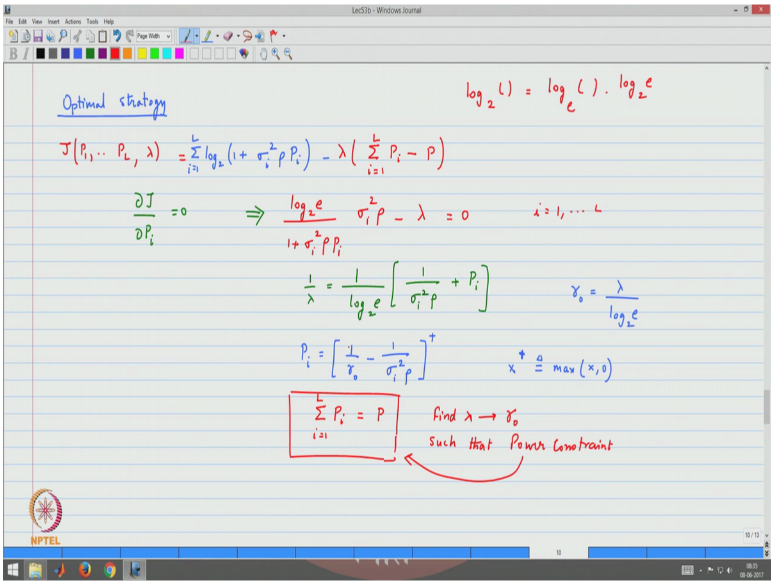
Now, this is also a way of saying that we are going to transmit only on the strongest eigen mode. Now previously from our discussion of capacity of wireless channels we know that though this has some attractive features it only 1 channel is transmitting, but we find that this may not achieve the best performance compared to allocation across all channels. So, option B where we tried to use all of the channels and come up with the following scheme where we say that I am going to transmit P i proportional to 1 by sigma i square.

So, in other words for those channels that are not so good where sigma is sigma is small we are going to allocate more power and so this would be referred to as an equalization policy you are trying to make each of the channels behave the same way the channel has inherently again sigma i squared, you are trying to compensate for it by adding less or more power. So, that the gain of the channel is cancelled by the allocation of the power. So, this is a equalization strategy or equalization policy and again there are advantages and disadvantages. The key point to note is that the poor channels that this the bad not so good quality channels poor channels get more power allocation get higher allocation of power.

Now is that a good strategy or a not so good one that we will have to analyze and understand, but if you recall we did discuss that though this seems like a very fair skin where you are trying to equalize the power we did see that when we were talking about a single input single output system that this may not be the best strategy, when the channel conditions are not so good it is better not to give too much power to that especially when you are going close to outage. In fact, we did see those situations where when a channel goes into outage we actually do not give any power to those channel. So, this would be something that we want to take into account or take cognizance of so that we can understand what is the best. So, poor channel get more power allocation now again we will have to see if this is the best strategy that we can do.

So, now, comes the third or the final or of the what we would like to call as the optimal strategy optimal strategy. Now the optimal strategy of course, says what is our criterion not use any heuristics like saying give all the power to the one, to the best channel or you know distribute the power.

(Refer Slide Time: 26:19)



So, this says that let us write down the objective function the objective function that we are trying to maximize is the total capacity. So, we are trying to maximize i equal to 1 through lone plus sigma I squared rho times P i that is the capacity of each of those parallel channels. So, we are trying to maximize the some capacity. Now this would be subject to a power constraint lambda times summation i equal to 1 through L P subscript i minus P where P is the total power. So, that would be the constraint and we would then write down the objective function the Lagrangian method where the objective function would mean that we want to find out these different power allocations P 1 through P L comma lambda. So, the objective function is written in terms of this.

Now, we want to maximize the object the some capacity. So, the approach would be to differentiate the objective function partial derivatives with respect to the power, different powers dou P i and set it equal to 0. Now if you remember we did something exactly the same steps when we derive the optimum power allocation under different SNR conditions for a SISO channel. So, basically following the same method, I have made a mistake this is not just summation it is log base 2 log base 2 of an summation I equal to 1 through L please note that if the capacity is not just the quantity within bracket is logarithm base 2 that is what gives us the capacity.

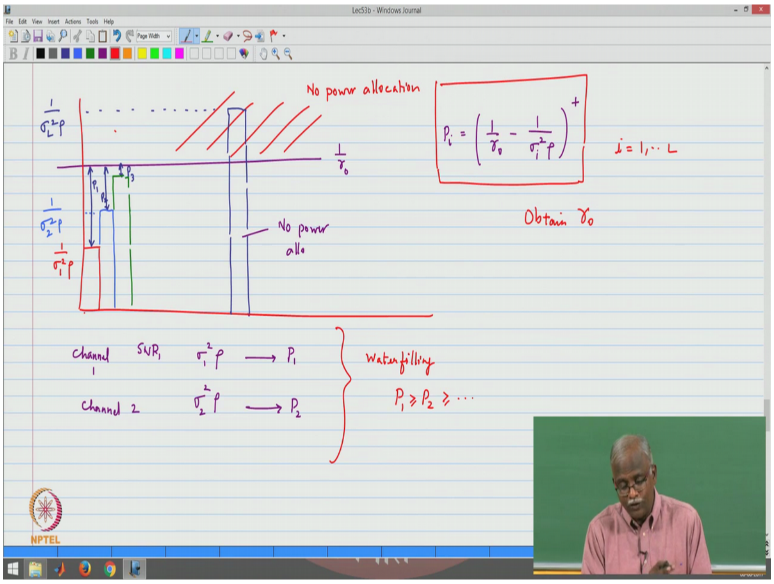
So, now I would have to differentiate each of these terms with respect to P i and notice when I differentiate with respect to P i I will get only one term from the first expression and one term from the second expression and also keeping in mind that we can always right logarithm of a quantity to the base 2 can be written in terms of the logarithm of the quantity base e multiplied by logarithm of e base 2. So, basically I would like to write it in terms of logarithm natural logarithm times log e base 2. So, this will give me log e base 2 that is a constant and when I differentiate the natural logarithm of 1 plus sigma i squared rho P i then I get 1 plus sigma I squared rho P i and then in the numerator sigma i squared times P rho sigma I times rho I am differentiating with respect to P i then on the other side we get minus lambda equal to 0. So, this is a similar expressions we will get for i equal to 1 all the way through L.

Now from this of course, we can get the values of the P i's, but how will what is it that we need to achieve we need to find out that value of lambda that will be use useful for us. So, if you were to write rewrite this expression just one step we can write 1 over lambda write rewriting this expression we can write it as 1 by logarithm e base 2, 1 by sigma I squared times rho plus P i that is you can write down an expression for lambda. And of course, we can write our intention is of course, to get the values for P i. So, we would like to write down and define a constant in the following form. So, write down call gamma naught as lambda divided by logarithm e base 2 then we can actually write down P subscript i can be written as 1 by gamma naught minus 1 by sigma I squared sigma I squared rho, sigma I squared times rho.

And of course, since this is a power allocation problem we would like to make sure that there is only positive powers that are located because the notation with the plus sign says that x plus is defined as maximum of x comma 0. So, if it is negative then you choose 0 as the threshold. So, this is a plus sign notation, because it is a power allocation problem. So, from the partial derivatives we get an expression for each of the P i's now with this P i's we want to find out the value of gamma not or in other words the value of lambda such that the condition is satisfied.

So, summation P i, i equal to 1 through L must be equal to p. So, find the value of gamma naught. So, find lambda or which is the same as finding gamma naught such that this is satisfied, such that the power constraint is satisfied, such that power constraint this is the power constraint is satisfied, again very very similar to the approach that we have taken in fact, identical approaches. So, what we find is that the power allocation is always done with respect to a constant value and it is also inversely proportional it also had dependent on the, with the reciprocal of the SNR where rho is the SNR sigma i squared times rho is the SNR.

(Refer Slide Time: 33:08)



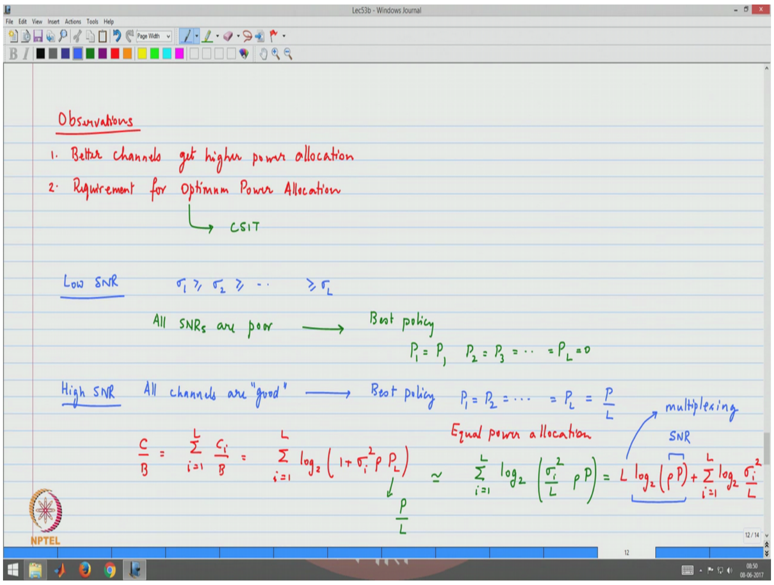
So, if you were to think of the in the following way if you were to draw a pictorial representation of these. So, if you were to draw the first parallel channel and its SNR 1 by sigma 1 squared times rho. And then the second channel which has got slightly smaller value of sigma i so that would be the second channel second channel when I take the reciprocal. So, what we have drawn here this is 1 by sigma 2 squared times rho then the third channel so on and until we come to the last channel where you would have the following the representation of the channel. So, what we are drawing here is the representation of 1 by the sigma i squared times rho. So, the characterization of the channels and this one the last one has got a power level or of a level 1 by sigma L squared times rho and so this is the first 1 is 1 by sigma 1 squared times rho.

Now, the water filling method says that there is a way of drawing the threshold which is given by 1 by gamma naught 1 by gamma naught and anything the any channel that lies above this where 1 by sigma L squared rho is greater than 1 by gamma naught will not get any power because if you remember the power allocation P i is given by 1 by gamma naught minus 1 by sigma I squared times rho. So, for those channels which exceed this 1 by gamma naught threshold, for those in this region those will not get any power allocation. So, these will not get any no power allocation, so no power allotted to these power allocation once you have cross the threshold now those channels which are below the threshold will get a power and the power allocation will be proportional to the difference from this if you remember the water filling approach that we have taken same thing we can we can do now.

So, basically if this is the channel now think of the difference from the or the height from the water level that corresponds to P 1 and likewise if you go to the second channel this becomes P 2 and P 3 and so on, but the minute you have cross the threshold then there is no more power allocation being done. So, it is completely analogous to what we looked at in terms of water filling under different SNR conditions. So, think of the channel that we have we are going to be working with as a scenario where we have L parallel channels each of them have got SNRs proportional to sigma 1 squared times the average SNR you are going to allocate power to each of these.

Now, based on the water filling strategy if the ratio that what we have just now what we have just now discussed that 1 by sigma I squared rho exceeds the threshold then we get a power allocation because of the plus sign at the top we find that there is no power allocation for those channels. So, for example, there is no power allocation here, but for those channels which fall below the 1 by gamma naught threshold we get the power allocation based on this. So, the intuition is the same as what we had previously where it says that the strongest channel which is represented by SNR 1, SNR 1 channel so this is channel 1 which has the best conditions sigma 1 squared times rho will get a power allocation P 1. Sigma 2 squared times rho channel number 2, channel 2 the parallel channel number 2 gets allocation P 2 and notice that from the water filing like that you get for each of the channels that is satisfying the threshold and the water filling principal says that P 1 will be greater than or equal to P 2 greater than or equal to so on and so forth until those channels that are falling within the threshold. Once you have crossed the threshold for power allocation those channels do not get any power.

(Refer Slide Time: 38:21)



So, the observations, so what we have done is we have looked at the power allocation through a scheme which is very similar to the where we have done the capacity calculations earlier. So, the observations, observations give us the important results the first one is that the better channels, channels with better SNR call that as the better channels get higher power allocation and this is completely consistent with our water filling strategy. So, a higher power allocation and of course, there are channels which will go below the threshold and if they go below the threshold then we get no power allocation. So, this is the strategy that we will follow and when we are presented with channels with the range of values then we do the optimum power allocation and then we get the maximum performance out of the system. So, that is the best way we can do when we are having when we have the luxury of doing power allocation at the transmitter.

Now, keep in mind that this does require. So, this is possible the requirement for water power allocation, requirement for optimum power allocation that is water filling is that you know the transmitter at the receiver or information has to be sent from the receiver to the transmitter so the transmitter has full knowledge of the channel. So, optimum power allocation this optimum power allocation is possible only if we have knowledge at the transmitter CSIT, it must have knowledge at the transmitter and based on that understanding of which channel has got how much SNR we are then able to do the power allocation.

So, let us now look at some interesting extreme cases and again these are useful for us to visualize and understand what these results mean. The first one is we look at the case where it is low SNR low SNR meaning that we have arrange the channels in the following manner sigma 1 greater than or equal to sigma 2 all the way to greater than or equal to sigma L and we find that they already sigma 1 itself is a small and of course, you can see that then sigma 2 all the way to sigma L has going to be even smaller.

Now, under this assumption where we are not going to get very good performance because already the channels under poor channel conditions. So, under this condition we find a very interesting result. So, the best channel allocation when all of them are poor, basically all SNRs are poor, so all SNRs are poor under this condition we find that the best channel allocation is the best eigen channel transmission, so the best policy under this condition is rather than trying to distribute your power between all the different channels where all of them are not so good the best option seems to be give all the power to the best of the channels that you have. So, basically P 1 equal to P, P 2 equal to P 3 equal to P L equal to 0. So, this is the what we called as the best eigen transmission policy or the best channel policy and the reason is clear because our ability to get the maximum data through the system is maximize by giving all the power to sigma 1 where sigma 1 by itself is not so good, but by giving additional power we are able to say that you know able to get the maximum out and. So, this is a intuitively something that you would feel it is the right thing to do.

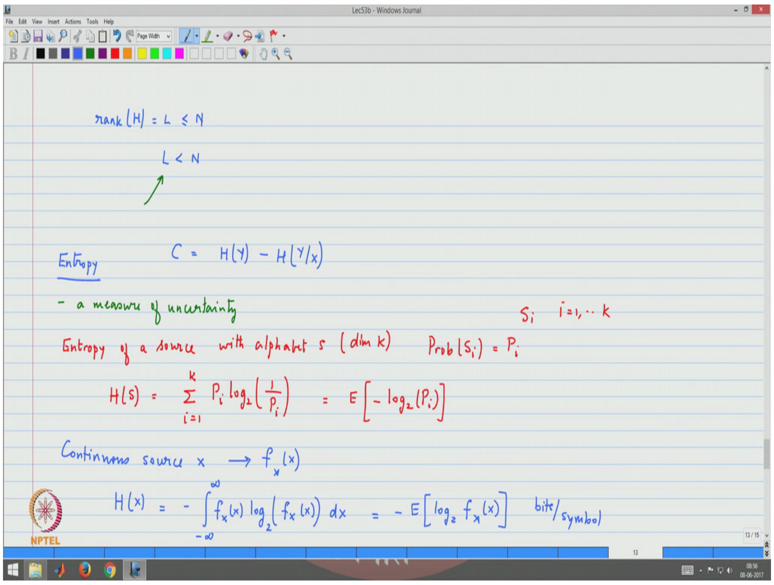
Now, what happens if all of them are high that is even all the way from sigma 1 to sigma L. So, that is the high N SNR extreme high SNR. So, which means all channels are good all channels, are good. So, under this condition again asymptotically it turns out that the best policy for this happens to be again something that you might be able to guess says that you know when all the channels are good really there is no need to fine tune your power allocation you are equally good performance when you say P 1 equal to P 2 equal to P L equal to P over L you allocate equal power. So, in other words do not be need not try to fine tune and achieve minor differences between the channel all of them are good just give them equal power and so this would be the equal power allocation strategy this is again another into intuitive result.

So, to summarize the best channel best method would be the water filling approach where we do the power allocation using the threshold for water filling. So, i equal to 1 through L where you are obtain you we have to obtain the value gamma naught, obtain gamma naught and then using that we compute the power to beat to be allocated to each of the channels. Now what we want to observe one more observation what would happen. So, when we have such equal power allocation let us do a quick calculation the capacity C divided by B would be equal to summation, i equal to 1 through L of the individual capacity C i divided by B that is given by summation i equal to 1 through L logarithm base 2 1 plus sigma i squared rho times P L, but we know that P L we have taken to be P divided by L as the expression. So, this can be rewritten as sigma i equal to 1 through L going to write it as in terms of the approximation it is logarithm of 1 plus the quantity, but since me know that all the SNRs are high values, with respect to log of 1 plus the quality we write it as log of the quantity ignoring the 1, this can be written as summation i equal to 1 through L logarithm base 2 sigma i squared divided by L rho times rho times P, let me just write it this is rho times p. So, this is our expression.

Now very interesting thing will come out one this can be rewritten in the following form where we write it as a write it in terms of 2 quantities the first one is a summation of logarithm of a constant. So, L times logarithm base 2 of the constant which is rho times P there are L such terms so we get L times this plus summation i equal to 1 through L logarithm base 2 of sigma i square divided by L just writing down that expression. So, I would like to highlight the following. So, if you were to, if you see that you have logarithm base 2 rho times P where rho times P is an expression of the SNR right, SNR rho times P proportional to the SNR.

Now, here we find that the capacity is going to grow linearly with the value L. So, this would what is what we would refer to as the multiplexing gain because we are transmitting the information along L parallel channels the total capacity that under conditions basically linearly with L. And again a very intuitively satisfying result because we have created as we created L parallel channels the capacity is dependent directly on L.

(Refer Slide Time: 47:50)



Now, the last observation is what happens when we find that the rank is not equal to the dimension of the rank of H. So, basically when we find that the rank of H rank of H is something that is less than full rank. So, I basically rank of H we said was equal to L we said is less than or equal to N. So, so what if L is less than N. So, in other words we get we do not have full rank and we have pure channels. So, in such a case then what you are doing is still transmitting using the L parallel channels where L is less than the rank of the matrix and therefore, the capacity goes down. So, you can see how if the rank of H goes down then the capacity also decreases. So, very very important that we identify that the rank of the matrix H is plays an important role in terms of the capacity that we can achieve.

So, I hope this gives us a good framework for us to understand the capacity of when we have when we are able to do the power allocation. Now, I would like to take us to a next level of understanding as or as the entropy and the context of the capacity is concerned. So, let us spend a few minutes on understanding the basic definitions of the quantity entropy which we have used in our definition of the capacity. Notice that we said the capacity of a channel capacity of a channel is given by the entropy of y minus the entropy of y given x. So, we are very interested in computing the quantity of called the entropy. So, entropy in it is by definition is a measure of uncertainty measure of uncertainty and we can define it in the following way.

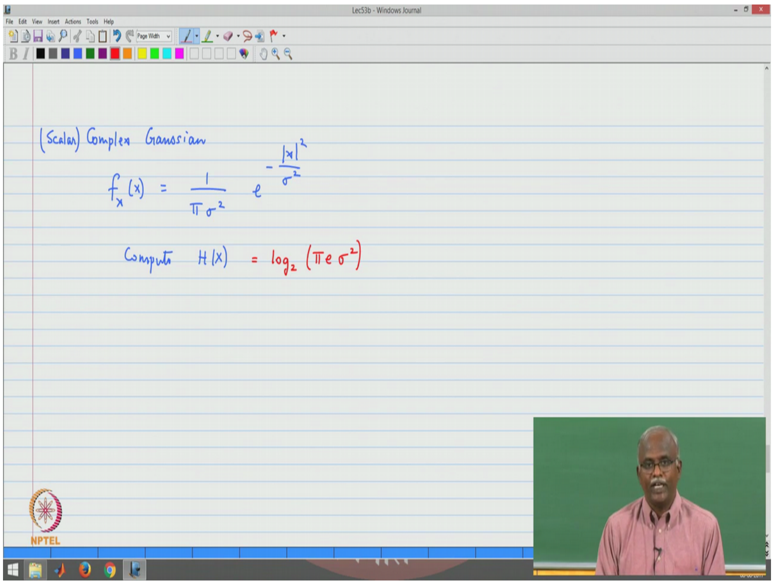
So, if you were to think of the entropy of a source, let us say that you think of a source with entropy of a source which produces symbols from an a finite alphabet. So, we can think of the entropy of a source this I am sure you would have studied in information theory with alphabet S with alphabet S of dimension K so; that means, there K symbols that are transmitted and we find and we have the probability that we have transmitted the symbol S K. So, S i where i is equal to 1 through k, i equal to 1 through k, probability that you have transmitted a symbol S i we call let us call that as P i. Then under such assumption, so basically we have a source with s alphabet of size S there are K such values that S can take and the probability of each of those K values is known it is given by P subscript i. Now the entropy of the source entropy of the source is given by summation i K equal to or equal to 1 through K, i equal to 1 through K of P sub i logarithm base 2 1 over P subscript i P i, logarithm base 2 1 over P i.

So, this is a very useful expression if you were to look at it carefully this can be written as the expected value of minus of logarithm base 2 of P i. So, this is the expression that that we will use and it is a very very useful expression for us this is the result that we have from information theory.

Now, if you were to think of it as a continuous source. So, for a where the variable of interest is a source x, so for a continuous source continuous source x then you would have the corresponding probability distribution function f subscript x of x and the entropy of this source H of x can be written as minus of minus infinity to infinity f x of x corresponding to P i and logarithm of 1 over f x of x, but that can be captured with a minus sign logarithm base 2 f x of x very very important that we are able to times dx. So, this would be our expression for the entropy of a continuous source. So, this can also be written just like we wrote last time it can be written as minus expected value of logarithm base 2 fx of x and this would be in terms of the bits per symbol, this would be the entropy of the source x.

So, this is a very very useful quantity for us to have and what we would now like to do is understand how to introduce this concepts in the context of the capacity calculations.

(Refer Slide Time: 54:26)



So, let me leave you with an exercise. So, if I have a complex Gaussian a scalar complex Gaussian source x, in that case we can write down the probability distribution function f x of x in the following way - 1 by P i times sigma squared where the sigma represents the variance of the real and imaginary parts sigma squared, e power minus mod x squared divided by sigma squared. So, if this is the expression and again you can refer 2 books like (Refer Time: 55:07) to confirm that this is the probability distribution function of a complex Gaussian variable.

So, f x of x is 1 by P i sigma squared e power minus x squared by sigma squared. So, what I would like you to do is compute the value H of x, compute the entropy of the variable x where x is scalar complex Gaussian and I would like you to verify that the expression that we will get is given by the following, is equal to logarithm base 2 P i e sigma square. So, that is the entropy of a complex Gaussian scalar source.

So, again I would like you to try attempt to do this we will pick it up from here in the next lecture and build our understanding of entropy, how it will contribute to our understanding of capacity and also how to estimate the capacity of a MIMO channel.

Thank you very much.