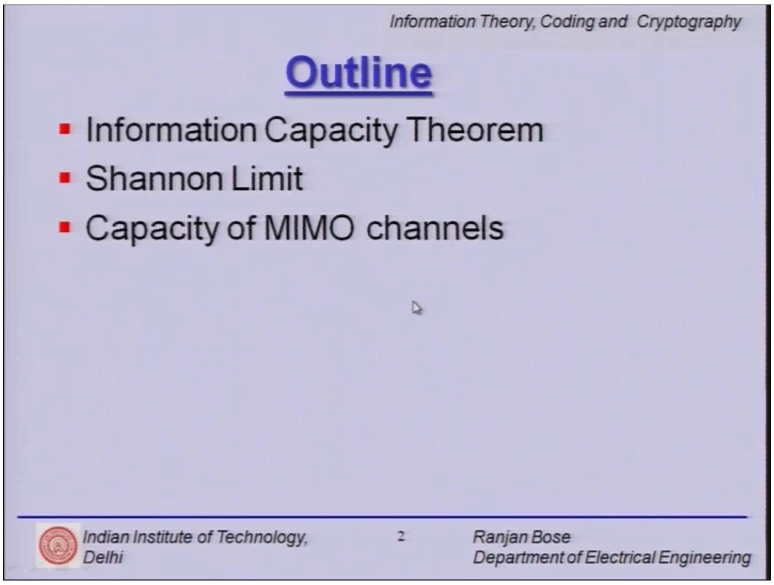
Information Theory, Coding and Cryptography
Dr. Ranjan Bose
Department of Electrical Engineering
Indian Institute of Technology, Delhi

Module - 12
Channel Capacity and Coding
Lecture - 12

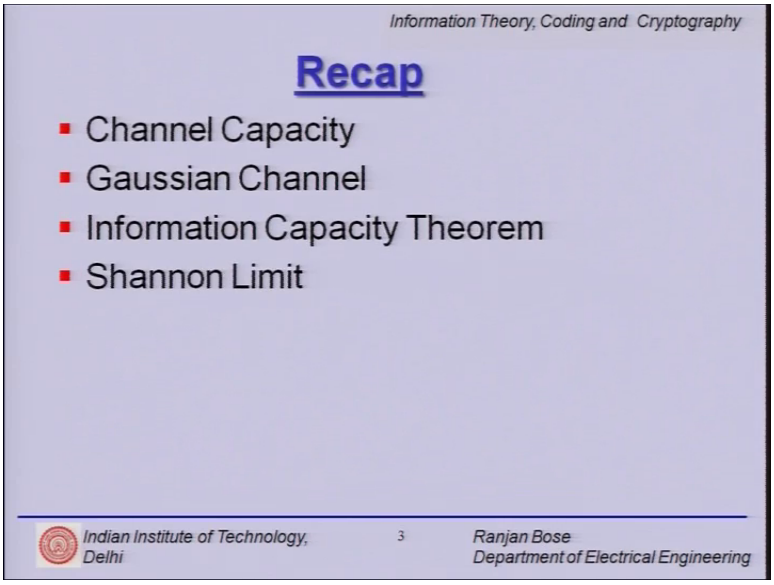
Hello and welcome to the next lecture on Channel Capacity and Coding.

(Refer Slide Time: 00:32)



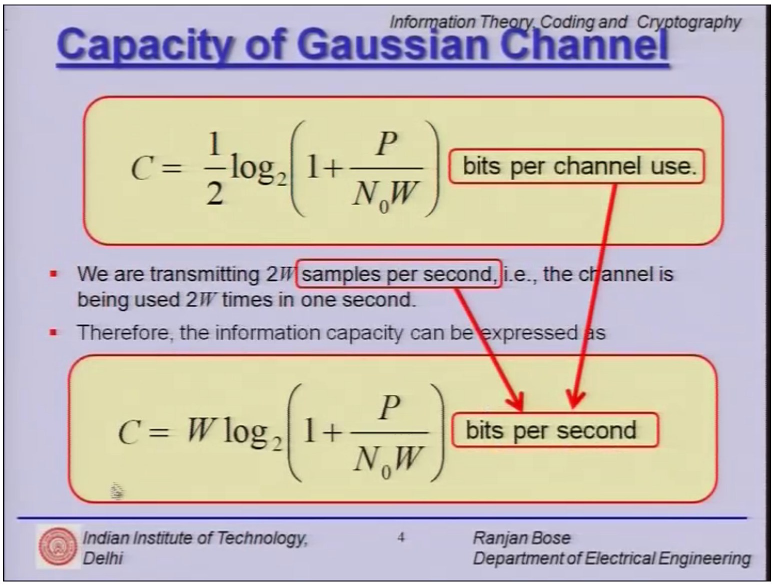
Let us start with a brief outline of today's talk because of it is importance we will revisit the information capacity theorem briefly and we will look at this Shannon limit what practical implications it has. And then we will look at this multiple input multiple output MIMO channels. They today form a part of most of the new wireless standards. So, they deserve some attention and therefore, the remaining part of our talk we will focus at the capacity of MIMO channels.

(Refer Slide Time: 01:11)



We start with a brief recap we have already learnt what is channel capacity? Then as a specific case and a practical case we looked at the Gaussian channel, then we derived the information capacity theorem and for the first time we realized how the bandwidth and the power is related. We looked at Shannon limit which we will revisit today.

(Refer Slide Time: 01:39)



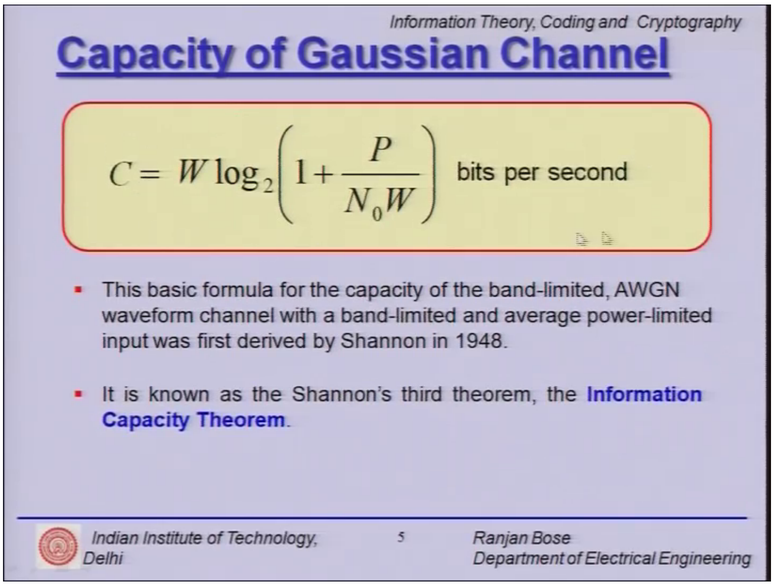
So, let us go back to the capacity of a Gaussian channel and if you remember from the previous class we derived that the capacity C is equal to half log to the base 2; 1 plus P N naught W bits per channel use. So, this was the ultimate step, where we said that the capacity is still being measured in bits per channel use is defined as follows. Since the log to the base 2 is being used hence the units are in bits.

And then we made the observation that we can transmit 2 W samples per second. So, those of the number of times we can access the channel and therefore, we can express the capacity theorem in terms of bits per second why we have bits per use and use per second gives me bits per second. So, the final answer we have is capacity is equal to W which is the bandwidth we have a bandwidth limited channel log to the base 2 perfectly, 1 plus P over N naught W.

Now, this we observe is nothing but the SNR. So, the capacity is linked linearly with bandwidth and logarithmically with SNR. And this is the case we made for the use of CDMA systems in 3 G wireless standards wherein we have the logic that it is better to invest in excess bandwidth rather than power. So, we put money on W which gives me a linear increase in the capacity as opposed to SNR.

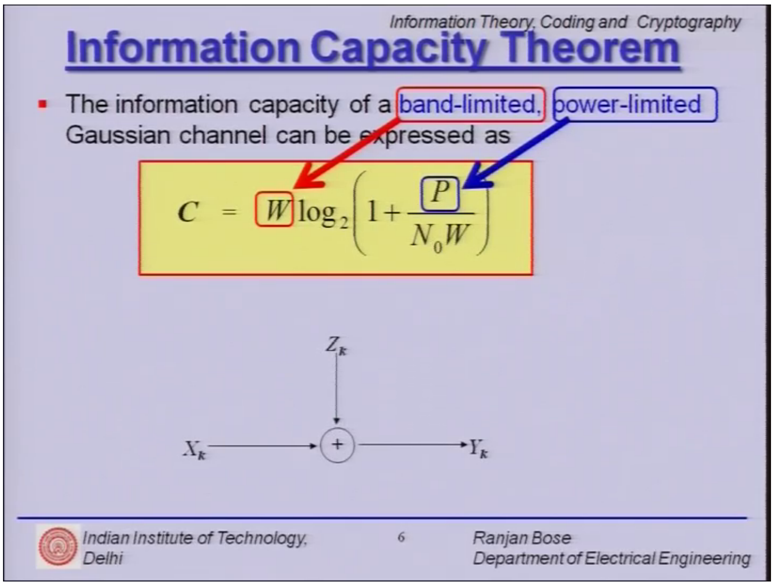
So, the CDMA systems have the signal almost noise like using a PN sequence. So, we reduce the power and we increase the bandwidth excess of what is required, so that I can get a higher capacity just a way to look at things practically.

(Refer Slide Time: 03:49)



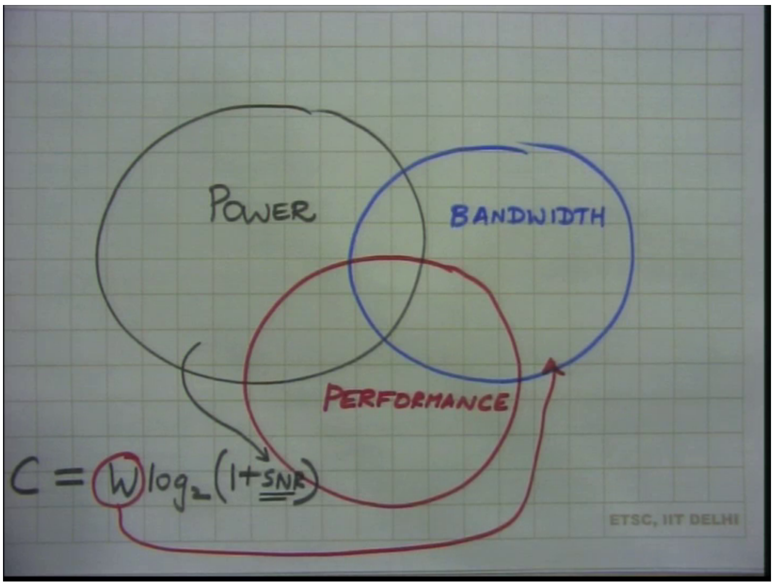
So, if you just observe this for a second this was derived in 1948 by Shannon and it is called the Shannon's third theorem and also referred to as the information capacity theorem it tells you about the C is the channel capacity. So, information capacity and channel capacity are being used interchangeably.

(Refer Slide Time: 04:13)



So, we just quickly revisit the model that we had. So, X k was the input, Z k or N k in some literature is the noise and Y k is the output and this system is band limited and power limited ok. So, this is the band limited portion and P is the power limited portion. So, it is very explicit that the C is now for the first time linking and permitting us to trade off between power and bandwidth alright.

(Refer Slide Time: 05:00)

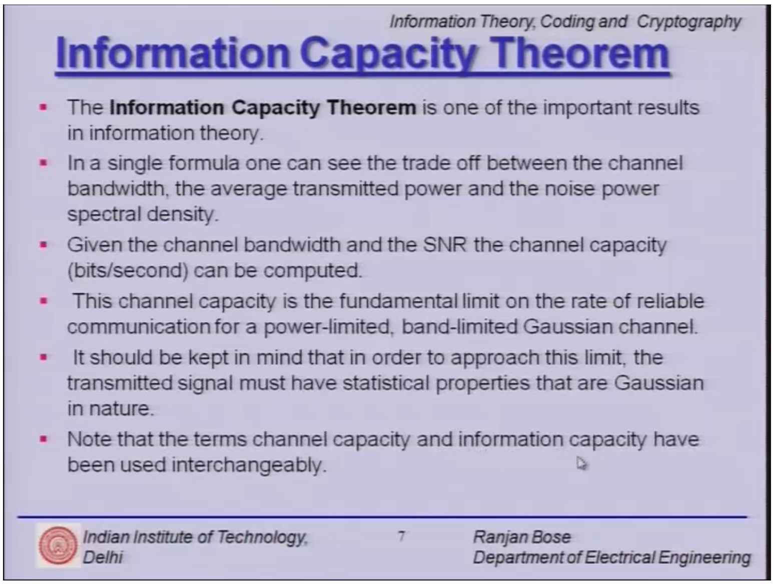


So, if you look at a Venn diagram we have this circle for power, and for bandwidth and finally, we have performance. Now what it tells us is that I can trade off one for the other. So, if you give me more power then maybe I would say I do not need more bandwidth even a reduced bandwidth to give you the same performance.

On the other hand if you say look power my battery is running out my comparator is saying so many hours of talk time why cannot, so I said ok. We reduce the power. But then I would need extra bandwidth to give the same performance, or you can say look I do not have enough power and bandwidth costs money say I want to reduce both of them then I will have to pay in terms of performance. How is performance measured? Well bit error rate, bandwidth in hertz, power in milli watts. So, this kind of gives us the trade off perspective.

Now what is present in the information capacity theorem? If you see this C no pun intended is W log to the base 2 1 plus SNR. So, this SNR part is the power the W part is a bandwidth, but performance is implicit. So, there is no where we are talking about the performance here, but clearly if we have a large bandwidth we can use stronger and stronger error control codes. And we can achieve a lower and lower residual error rate and hence improve our performance. So, that is how things are built in.

(Refer Slide Time: 07:35)

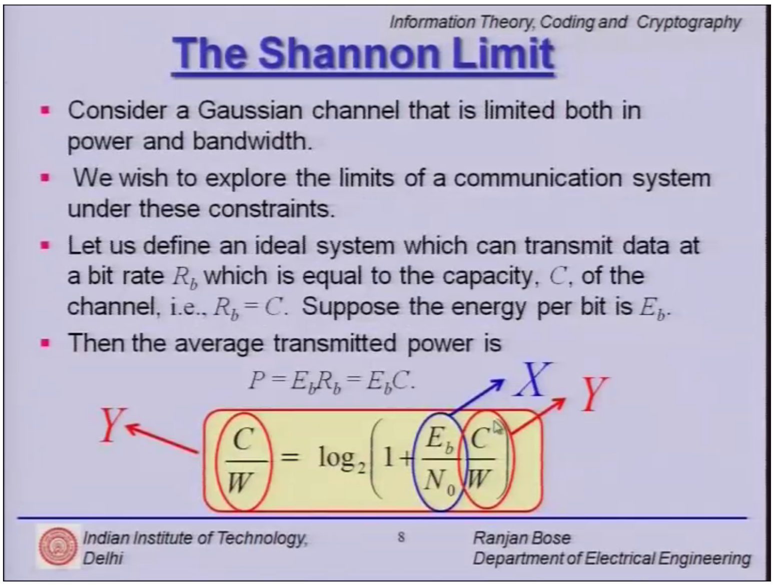


So, we go back to our slide and we say that information capacity theorem is indeed one of the most important results in information theory. We will extend it today to MIMO systems and in this one single formula we have the tradeoff between bandwidth transmit power and power spectral density of the noise.

Given the channel bandwidth and SNR the channel capacity can be computed, and this channel capacity is the fundamental limit. Because we have made the assumption of a Gaussian channel where X is Gaussian, N is Gaussian, and Y is Gaussian in real life well noise can be taken as Gaussian, but typically the signal may not be Gaussian. And therefore, you away from this, this is the best we can do. So, it is really a fundamental theoretical limit.

So, as I mentioned that in order to achieve capacity for a Gaussian channel the signal transmitted should have statistical properties which are Gaussian in nature. And so far we have used channel capacity and information capacity interchangeably.

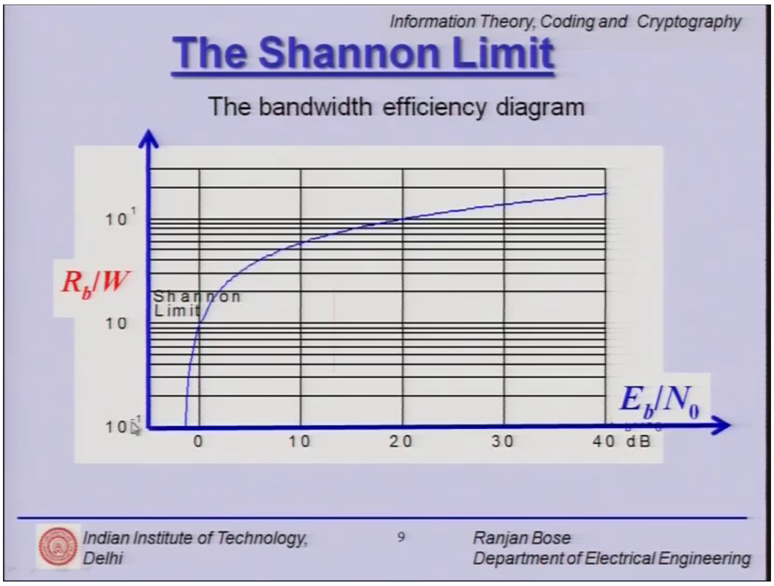
(Refer Slide Time: 08:48)



Now, we have a very interesting outcome of this information capacity theorem which is the Shannon limit and if you rewrite the capacity theorem with C is equal to the W, the W goes down here log to the base 2 1 plus P. Now, what do I write P as? So, I said well we define the data rate R b and for a change it is going at capacity is bits per second, R b is bits per second, so we have R b equal to C.

So, we replace power as E b into R b E b stands for energy per bit R b stands for bits per second. So, this is like joules per second which will give me watts. So, power is nothing but E b into C say I replace E b into C in the numerator N naught W was in the denominator, so have an alternate form. Why do I do it? Well C over W is kind of a normalized capacity with respect to bandwidth I call it Y, make another observation C over W figures again I call it Y and E b over N naught figures once. And this is E b is energy per bit over this N naught. So, this is the kind of measure for SNR.

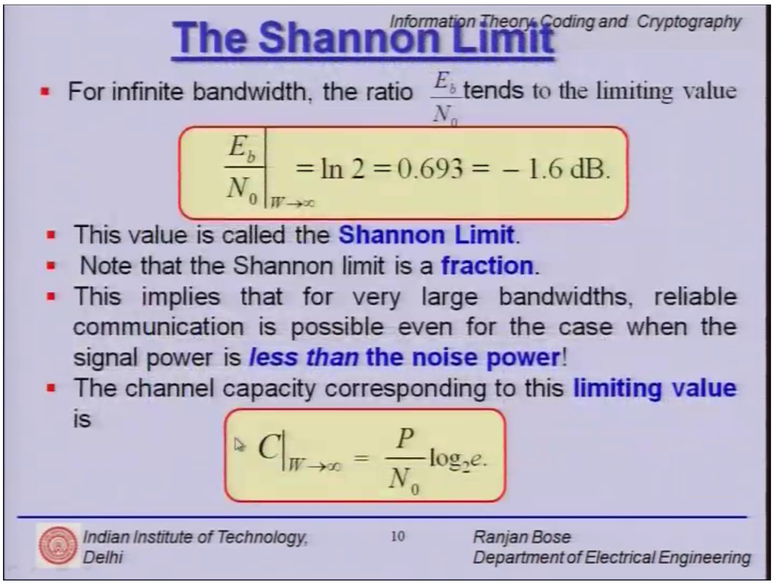
(Refer Slide Time: 10:23)



So, let us put them in the axis X axis E b over N naught and Y axis this. So, I have this y is equal to log to the base 2 1 plus X into Y. So, if you kind of plot it on the X axis we have in blue E b over N naught on the Y axis in red we have R b over W. Please note, R b and C has been used interchangeably they are equal then the plot gives me this curve which is called the capacity boundary, the bandwidth efficiency diagram.

Now, what is interesting to note in the previous figure is that this line if you look at here along the x axis for large bandwidth. So, bandwidth W is in the denominator R b is a system designed a parameter.

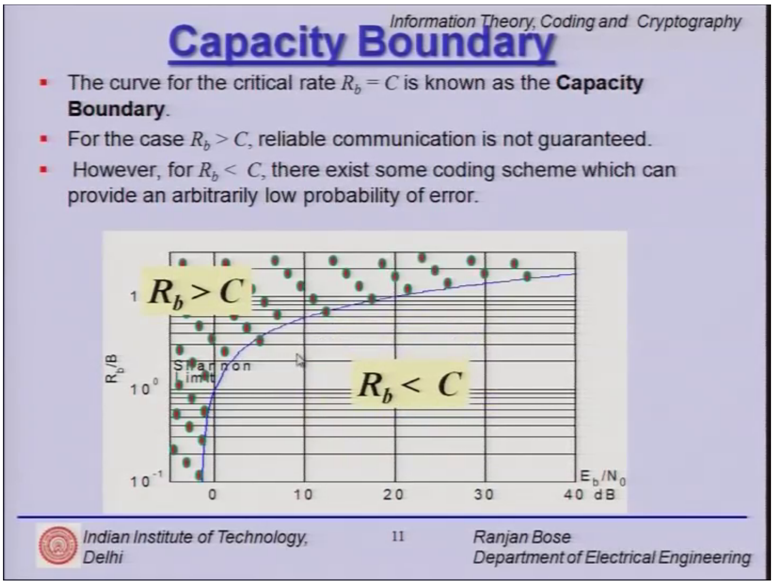
(Refer Slide Time: 11:23)



So, as I increase W, I go down the y axis as I go down the y axis this guy saturates. And for W tending to infinity E b over N naught is actually ln 2 which is a fraction which means that it is possible to have reliable communication over unreliable channel right. If you have large enough bandwidth and what is strange as is that that the signal power can be actually less than noise power and still you can have reliable communication.

So, this counterintuitive result this boundary is the Shannon limit which is a fraction. So, this is not obvious that even if my signal power is less than noise power I can still do reliable communication. On the other hand, if you have the expression for capacity as W tends to infinity, then you are limited by this power of the signal this SNR. So, this kind of gives you a limiting factor about the capacity ok. So, regardless of how much bandwidth I give you I cannot keep on increasing the capacity of my Gaussian channel.

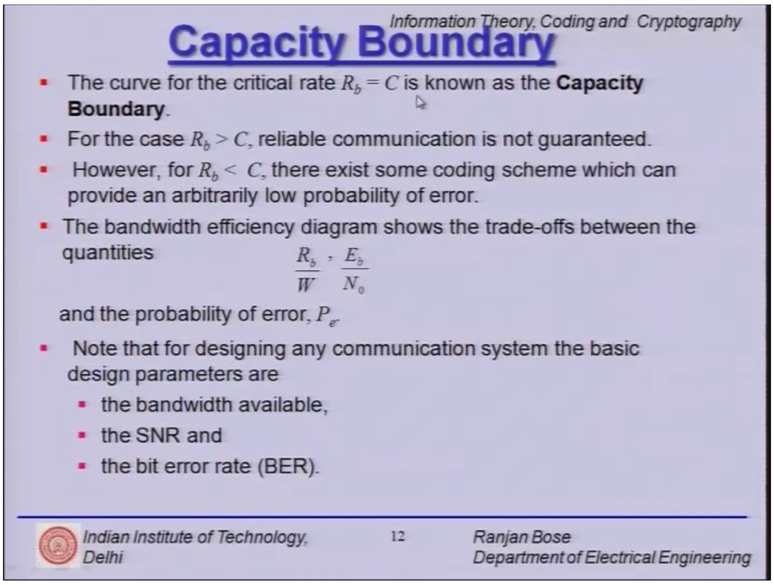
(Refer Slide Time: 12:42)



If you look at the capacity boundary above this line R b exceeds capacity and we know that if this rate is greater than this capacity reliable communication is not possible. Whereas so this is the region where we have R b greater than C and below this blue line is R b less than C where we again have reliable communication possible ok, so all of this region.

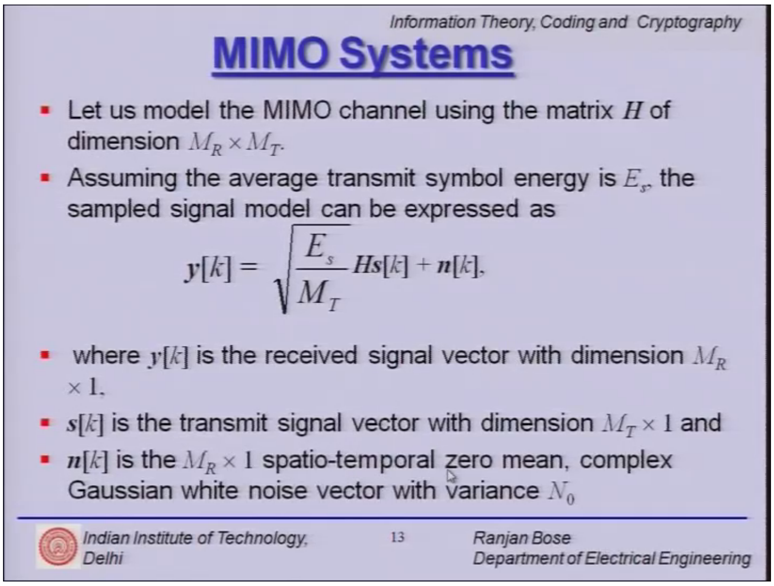
So, any point on this diagram any point is an operating point it gives me a particular SNR and a normalized data rate and I can design a system around it and that system should give me as low probability of error as I desire. So, in this diagram if you could plot a third axis you can possibly put the probability of error, but this is not clear from this figure where the probability of error the reliability component comes into picture. All it says is that I can have as reliable communication as I want, reliable means bit error rate 10 power minus 10 I will give it to you I do not have a recipe for that, but I have an existence proof for it.

(Refer Slide Time: 14:16)



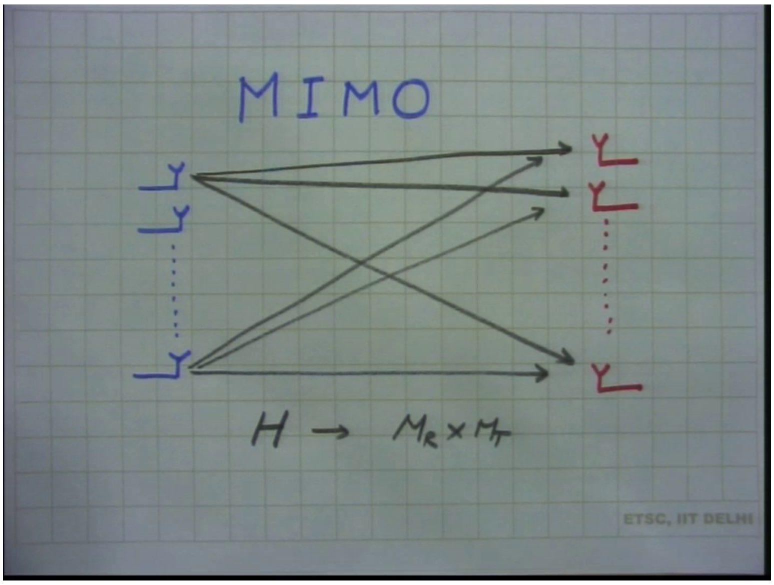
So, R b equal to C is the capacity boundary and please note that for designing any communication system that basic design parameters are the bandwidth available SNR and the performance measure BER. So, this we have now understood in terms of the slides we have seen ok. So, BER is also designated as a probability of error.

(Refer Slide Time: 14:42)



Now, let us look at MIMO systems. So, let us quickly revisit what is this MIMO system? So, we go back and refresh your memories.

(Refer Slide Time: 14:56)



So, in wireless communication I can have several transmit antennas and I can have several receive antennas. Well, in theory we can always do in practice the number of antenna antennas on a handset for example, would be limited right. So, I can have the linkage between transmit and the receive antennas.

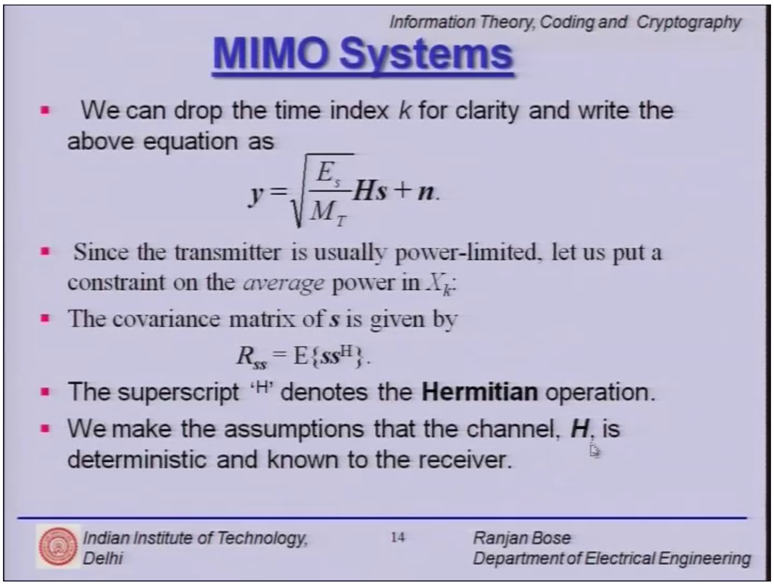
So, as you can see that I have a channel matrix H which is of the dimension M R into M T where M R could represent the number of antennas on this side, or empty could represent the number of antennas on this side. Now, clearly we can have a much higher capacity just looking at it intuitively because there is so many data pipes that I can possibly envisage in this MIMO systems ok.

So, let us look at the capacity of MIMO systems. Now we can have two scenarios one is we know how good or bad the channel is that is we know the channel characteristics the channel state information is available at the transmitter or whipping a blind game and we have no clue how the channel is which is good or bad. Why is it important? Well, look maybe some channels are good some are not good it is wireless after all and it makes sense to allocate power in a manner which should maximize the mutual information transfer and hence, the capacity of this MIMO system. So, that will be the general game plan for today to figure out how to gain understands the capacity of this MIMO system.

So, coming back to the slide if you see that if we assume that the average transmit symbol energy is E s E sub s then the sample signal model can be represented as y k k-th sample E s over MT. So, M T is the number of transmit antennas, H represents the channel matrix which is the represents our channel and sk is the transmits symbol n k is the noise. So, this is just the sample signal model.

So, what is yk? Y k is M R into 1 which is the number of received antennas, sk is the transmit signal, M T is the number of transmit signals. So, T stands for transmit, R stands for receive. So, I can have vector, matrix, vector, vector. So, nk I can have an assumption that is a spatio temporal zero mean complex Gaussian white noise with a given variance N naught ok. So, let us put this as our system model.

(Refer Slide Time: 18:45)

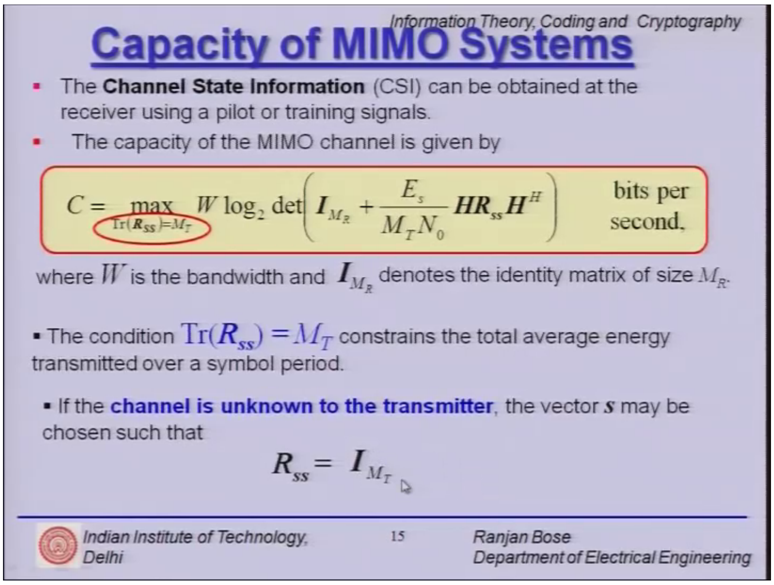


If we now drop this time index k for brevity we can write the same equation as y it is a vector. It is in volt phase E s over M T T's represents M T is the number of transmit antennas, H is the channel matrix, s is a vector, n again volt phase is a vector. Now, it is fair to say that the transmitter is power limited power cost money even if I have the money I should not transmit more than I should because my signal is somebody else is interference in a wireless situation.

So, it is good to be green today people are crazy about green communication which is essentially to use only as much power required as usual. So, since the transmitter is usually power limited and we have so many other constraints let us put a constraint on the average power in X k the transmit signal, please note the term average.

So, average power can be defined as the covariance matrix of s which is the transmitted signal which is given by R ss is expected value of SS Hermitian ok. The superscript H denotes the Hermitian operation we make the assumption that the channel H is deterministic and known to the receiver. So, the first condition is that the channel is deterministic well in real life it is not, but for the sake of discussion. And it is known to the receiver how do I know this channel well I can conduct experiments, I can send pilots, I can get some feedback, I can possibly have an estimate of H.

(Refer Slide Time: 20:45)



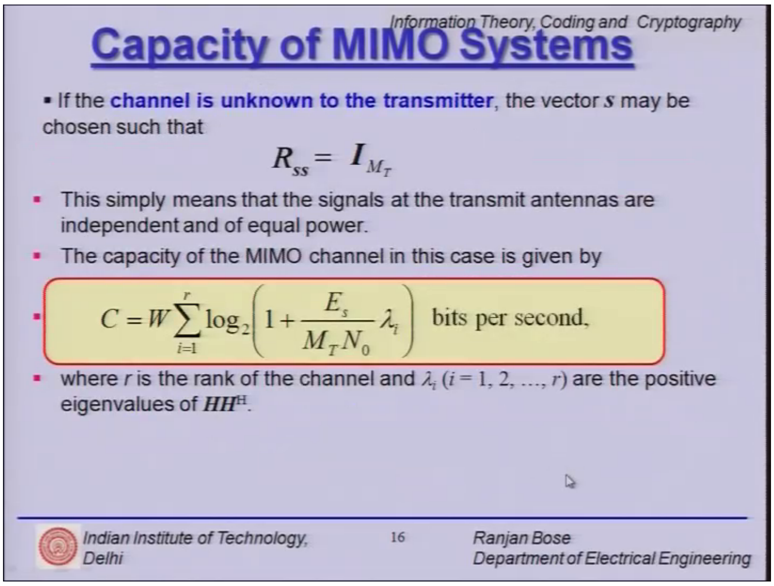
So, the channel state information the information about the channel gain matrix H can be obtained at the receiver using a pilot, or training signals. Then the capacity of this MIMO channel is given as follows C is equal to maximize over trace R ss equal to M T we will talk about it W log to the base 2 determinant of I MR is the identity matrix plus.

So, this is the same structure, but this time instead of 1 plus SNR we have now M T cross MR. So, this is M R plus E s over M T N naught again this is a notion of the SNR part H R ss, H Hermitian bits per second, W is the bandwidth, IMI denotes the identity matrix of size M R. Please note this constraint the condition trace R ss is equal to M T constraints the total average energy transmitted over a symbol period. So, this maximization is under the constraint that we do not have infinite power at the transmitter side, so we restrict that.

Now in the case when the channels are known to the transmitter what do I do? Well I treat all my individual data pipes identically I do not treat them differentially I say well I would rather put all my power equally in the all the transmit antennas. So, that is what the covariance matrix tells you it is an identity matrix of size M T it has 1 along the diagonals, and 0 elsewhere how many M T?

So, each one normalized power is 1 this is the best I should do this is the best I can do because the channel is unknown to the transmitter it is unfair to put more power in one of the antenna elements as opposed to other. Only when I have some idea about the channel if one of the channels is poorer, I should focus and put more power in the good channel as opposed to a not so good channel. But if the channel is unknown to the transmitter this is my best bet.

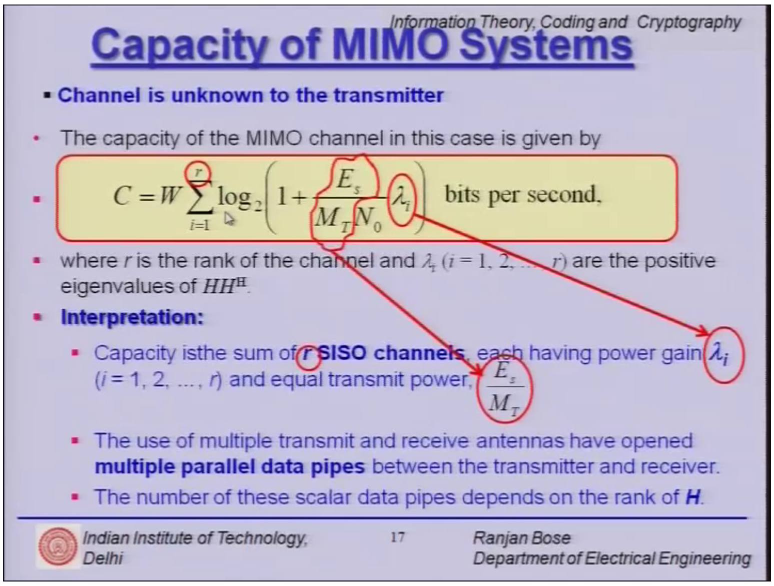
(Refer Slide Time: 23:26)



So, continuing with the channel is unknown to the transmitter the vector s may be chosen such that R ss are the identity matrix. So, this simply means that the signals at the transmit antennas are independent and of equal power that it what it means any cross terms would have shown up, so it would not have been an identity matrix.

But right now it says that it is independent of the equal power and in that case you can derive from the previous general formula the capacity of the MIMO channel is simply given by C equal to W summation i is equal to one through r, we will talk about r being the rank of the channel, 1 plus again this is kind of the SNR expression E s over M T N naught lambda i, where lambda i, i equal to 1 through r are the positive eigenvalues of this HH Hermitian.

(Refer Slide Time: 24:30)



So, let us just look at it a little bit more carefully because this is a very strong physical interpretation what are we trying to do? We have a MIMO channel, so empty transmitters, M R receivers. What are we trying to do? Find the capacity of this very big MIMO system and we have this expression for the capacity in terms of bits per second. We would like to have a physical intuitive understanding of this.

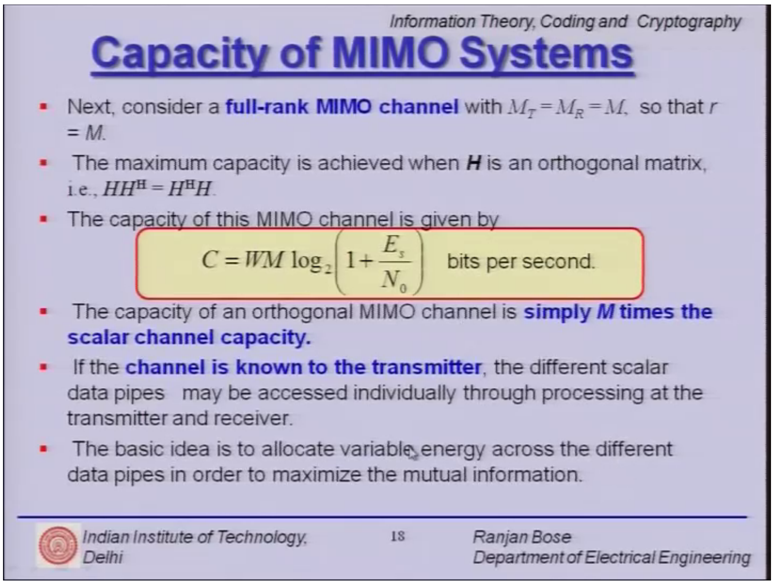
So, let us begin with our interpretation first look at this sum, so if we open up this sum and look at this W. So, it is W log to the base 2, 1 plus E s over M T N naught lambda 1 plus W log 1 plus E s M T N naught lambda 2 and so on so forth up to r. But what is an expression W log 2 1 plus SNR it is nothing but the capacity of a single input single output channel which have derived for this Gaussian case.

So, it only tells me that this combined capacity of a MIMO channel where the channel is unknown to the transmitter is nothing but the sum of r SISO channel each having power gain of lambda i and equal transmit power E s over M T that we had established earlier itself. So, let us look at this expression E s over M T is the transmit power right lambda i is the power gain alright, and r I am a identical SISO channels and what is this r it is the rank of the channel.

So, the way to interpret it is that this MIMO channel is effectively multiple parallel data pipes which are each SISO, how many are in number very interesting. So, use of multiple antennas and receive antennas have effectively opened multiple parallel data pipes between the transmitter and receiver and I am really excited because it meant it will really improve my capacity.

And the number of this scalar data pipes depends on the rank of H. What it means is? If it is a full rank then I have a much higher capacity ok. So, the more number of independent channels I can carve out of my H it depends on the real wireless channel the higher my capacity is.

(Refer Slide Time: 27:53)



So, let us look at a full rank MIMO channel and for the sake of discussion let us have equal number of transmit and receive antennas. So, full rank which means M T is equal to M R is equal to M, so the rank is indeed M ok. So, the maximum capacity is achieved where H is an orthogonal matrix, and this capacity of the MIMO channel is given by C is equal to W M log to the base 2, 1 plus E s over N naught bits per second this is almost intuitive ok.

So, I have got M scalar data pipes and so capacity is simply thus some of these M scalar capacities and which is nothing but M times W log to the base 2 1 plus E s over N naught this is the SNR. So, the capacity of an orthogonal MIMO channel why orthogonal?

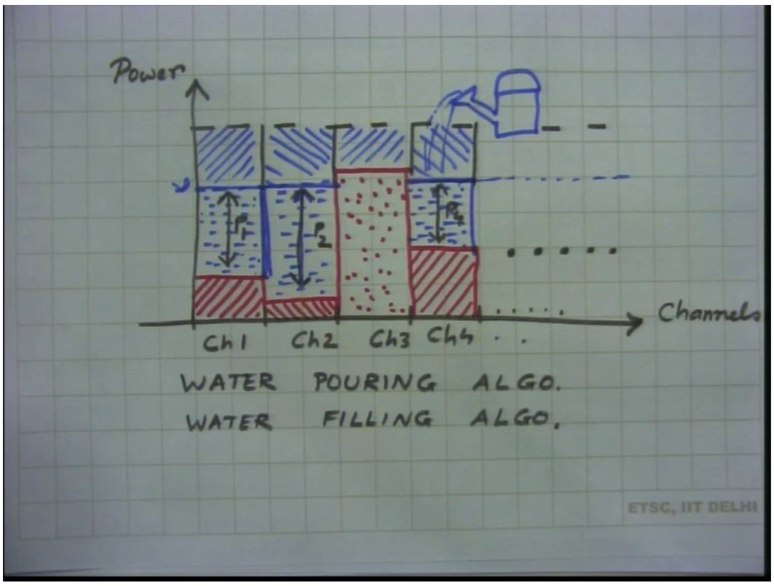
Because we are now started to talk about a MIMO channel in terms of it is channel matrix H.

And now we are only focused on the characteristics the mathematical characteristics of this matrix H that this matrix H is pretty easy to write because in general we have M T transmit antennas, M R receive antennas. Each antenna at the transmitter is connected to an antenna at the receiver and there is a channel gain from one to the other. So, it is pretty easy to construct this matrix H.

Now, this H matrix may have several interesting properties based on these mathematical properties we are making comments about the capacity. So, here we have a full rank matrix and H is an orthogonal matrix under this condition of orthogonality and this leads us to the best possible capacity of a MIMO system where C is given as this formula. What is it? It is simply M times the scalar channel capacity. Here we have assume the number of transmit antennas is equal to the received antennas is equal to M.

Now comes the question what if the channel is known to the transmitter? The different scalar data pipes will be accessed individually through processing at the transmitter and receiver alright. So, the basic idea is to allocate variable energy across different data pipes in order to maximize the mutual information.

(Refer Slide Time: 30:52)

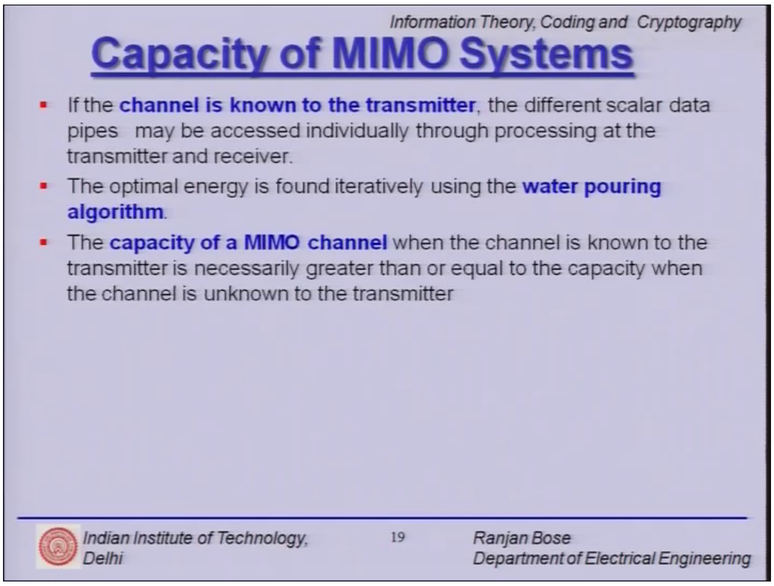


So, let us again have a intuitive understanding first about what we have to do? What is our problem? So, these are my channels let us demarcate them channel 1, 2, 3, 4. And on this axis we have to allocate power, so this is channel 1, channel 2, channel 3 and so and so forth. The problem is the channels may not be the same by now all of you appreciate the fact that we have multiple channels between the transmitter and receiver. The aim is to maximize the mutual information and we have said that the channel is known to the transmitter.

Suppose the channel 1 is affected by some level of noise. So, this is a power axis, so I have denoted a certain amount of noise power. And the channel 2 is a luckier channel and we have less noise. And channel 3 is my bad luck it has much higher noise, and again channel 4 has moderately high noise.

Now the question is we can do this measurement of how good the SNR in each channel is and thereby understand the level of noise each of the channels for all the possible channels. Question now is how much power should be allocated to each of the channels to maximize the mutual information that is the problem to be solved.

(Refer Slide Time: 33:10)



So, we come back to our slide we are trying to look at the problem when the channel is known to the transmitter and the job of the transmitter is to figure out how to optimally allocate power to the different antennas. This is a very practical problem. Now the different scalar data pipes may be accessed individually through processing at the transmitter and receiver ok.

So, the optimal energy is found by iteratively applying an algorithm called the water pouring algorithm we will explain why it is called. And the capacity of a MIMO channel when the channel is known to the transmitter is necessarily greater than or equal to the capacity when the channel is unknown to the transmitter. So, even though it is intuitive it is it can be mathematically shown that if you know the channel at the transmitter you can squeeze more out of it as opposed to doing a blind guesswork. And the water pouring algorithm if you go back to our original drawing we had this choice and suppose we have a total given transmit power which is limited by some quantity. So, it tells me channel 1 should be allocated this much of power. So, it is like I have to set the level of a water which is consistent for all channels.

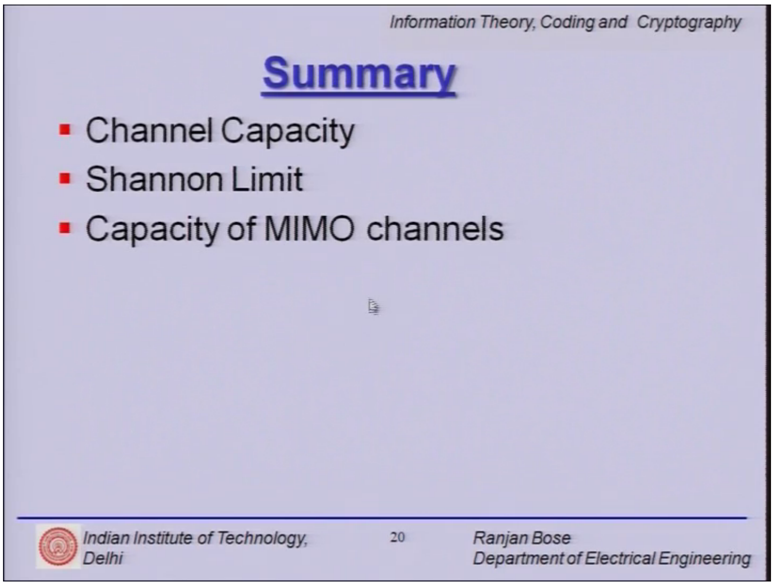
Now, since channel 1 had already this much of noise the amount of power that I will allocate to the channel we will be only limited, but is this is a good channel. So, I will be allocating more power to this channel, this is the power I have allocated to this channel. If I want to make it more explicit whereas, because channel 1 had more noise it got a lesser share of power for this one, so this is P 1 and this is P 2.

So, if you go to the third channel I am shocked to see that this has gone above the water level because who sets this water level. Originally I had a total amount of power to be allocated that decides if we increase my total available power I can raise this level. But right now with a given level I must do justice to everybody the channel 3 is really noisy it does not get any power. So, I will not use channel 3, I go back to channel 4 it is below this water level.

So, I have my can right and I pour water and it must fill in and reach this level. So, the power for this third fourth channel, channel 4 will be P 4 and so on and so forth. So, I can do this and therefore, this is called the water pouring algorithm ok. There is a mathematical proof for doing this, but essentially this is what I am trying to do.

Now, if you have the luxury of a larger quantity of power suppose I set my threshold here, then I will see that even channel 3 which was left out in the previous case also got the power. So, if I choose to change my algorithm and say no I will allocate more power now then I can redistribute additional power in this way to get maximal mutual information and thereby the capacity. So, this is called the water pouring algorithm, or sometimes it is called the water filling.

(Refer Slide Time: 38:33)



So, we come back and summarize where we have reached today we started off with the channel capacity theorem. And we looked at the first fall out which is the Shannon limit from this information capacity theorem, and then we spend quite a bit of time on the capacity of MIMO channels. We looked at different cases whether the channel state information is known to the transmitter, or it is not known. And how we can intuitively visualize scalar data pipes depending upon the mathematical properties of this matrix H, which links the M T transmit antennas to M R received antennas.

With that, we come to the end of this lecture.