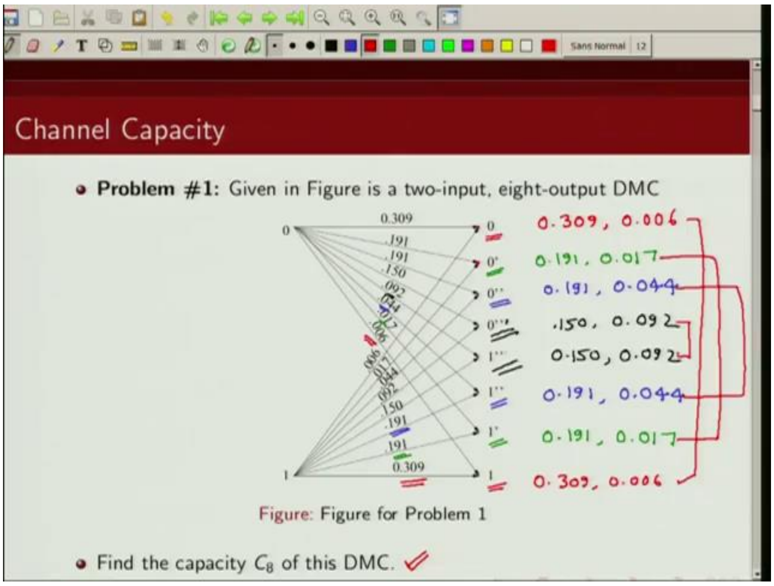
An Introduction to Information Theory
Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture - 12C
Problem solving session-III

Welcome to the course on an introduction to information theory. So, in this lecture we are going to solve from problems related to channel capacity computation.

(Refer Slide Time: 00:29)



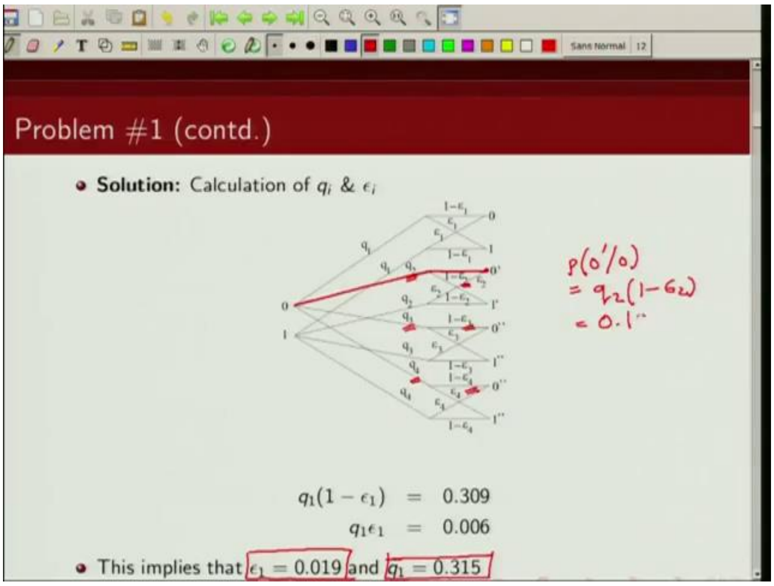
Since problem one, number one, we are going to compute capacity of this discrete memory less channel. So, you can see this channel has two inputs 0 and 1, and eight outputs which are given by 0 0 dash 0 double dash 0 triple dash 1 triple dash 1 double dash 1 dash and 1. So, these are the eight outputs of the channel, and these are the two inputs to the channel. The channel transition probabilities are given here. So, probability of 0 given 0 is 0.309, probability of 0 dash gains 0 is given by 0.191, and similarly you can see the transition probabilities are given here. The first question is, find the capacity of this channel.

So, first we will try to see whether this is a symmetric channel or not. So, clearly this it is not a strongly symmetric channel, because you can see this is not uniformly, like it is not uniformly focusing; for example if you consider this particular output the probabilities are 0.309 and 0.006, but if you consider this output, the transition probability of 0.191

and 0.017. So, clearly this is not a strongly symmetric channel, but we need to see whether this is a symmetric channel, whether we can decompose this channel into strongly symmetric channel. So, let us look at this output 0 and output 1. If you look at the probabilities this is in the decreasing order 0.39, and this is 0.006. What about this particular output. Again the probabilities, transient probabilities 0.309 and 0.0006309 and 0.006, which means zero and one have the same focusing. Let us look at 0 dash and one dash, this is 0.191 and what about this, is 0.017.

Similarly this one is 0.191 and this is 0.017. So, these two are also uniformly focusing. Similarly, look at 0 double dash and one double dash. This is 0.191, and 0.044. This is also 0.191, and this one is 0.044. And finally, this, these two also have same focusing, you can see this is 0.150, this is 0.092, and one triple dash is 0.150, and this is 0.092. So, you can see now, that we can decompose this output into four uniformly focusing set; one corresponding to these two, one corresponding to these two, then one corresponding to this, and one corresponding to this. So, to compute the capacity we will first try to decompose this into strongly symmetric channel.

(Refer Slide Time: 04:59)

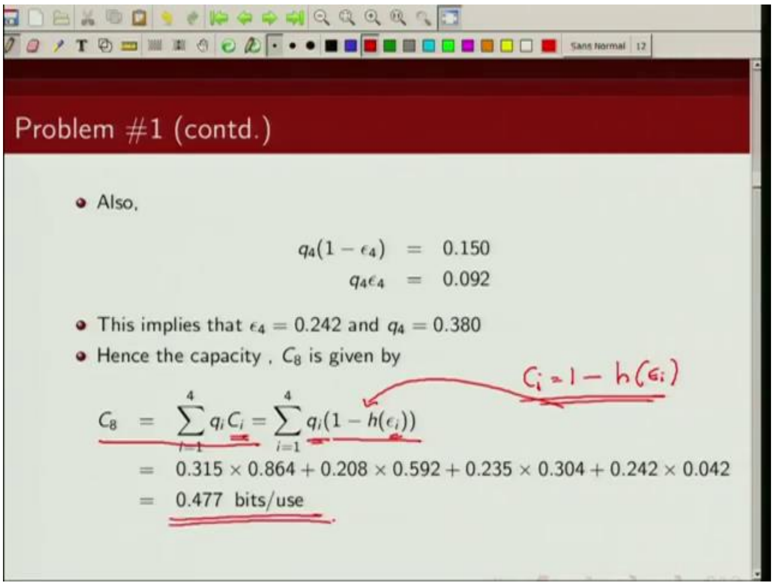


So, if you try to do that, as I said 0 and 1 they have same focusing 0 dash and 1 dash as same focusing 0 double dash and 1 double dash as the same focusing and 0 triple dash and 1 triple dash are the same focusing. Now, if I decompose the same channel into four such strongly symmetric channel, I can write the equivalent decompose channel in this particular fashion, where I am selecting this channel consisting of output 0 1 1 with selection probability q 1. I am selecting this channel which gives output 0 dash and 1 dash with selection probability q 2, and selecting this channel which gives output 0 double dash and 1 double dash, which selection probability q 3, and similarly I am choosing the fourth channel with selection probability q 4. And each of this individual channels are binary symmetric channels.

So, I am just writing this cross over probability as epsilon 1, epsilon 2, epsilon 3, and epsilon 4. Similarly these probabilities are 1 minus epsilon 1 1 minus epsilon 2 1 minus epsilon 3 and 1 minus epsilon 4. Now the first thing that I need to do is to find out the value of q 1 q 2 q 3 q 4 ,and epsilon 1 epsilon 2 epsilon 3 and epsilon 4. Now how do I find out these values? So, I am just going to map these transition probabilities. So, if you look at let say transition probability from. So, transition probability of 0 given 0, what is this in this diagram - it is given by q 1 times 1 minus epsilon 1; that is this, and if you go back to this diagram the transition probability of 0 given 0 is given by 0.309. So, what we can do is, we can write q 1 into 1 minus epsilon 1 is equal to 0.309.

Similarly, what is the transition probability of 1 given 0; that is given by from this diagram, this is q 1 times epsilon 1; that is this. And let us go back to original what is the transition probability of y 1 given 0; that is 0.006. So, q 1 epsilon 1 is 0.006. I have 2 equations, and two unknowns q 1 and epsilon 1. So, I can solve them to get q 1 and epsilon 1. So, I get epsilon 1 is to be equal to this, and q 1 to be this. I follow exactly the same procedure to get q 2 epsilon 2, q 3 epsilon 3, q 4 epsilon 4. I am just writing it here, we can do it once more. Let see look at transient probability of, what is the probability of o dash given o 0 0 dash given 0 that is given by q 2 times 1 minus epsilon 2; that is equal to q 2 times 1 minus epsilon 2, and this is nothing, but if you go look at the original figure probability o dash given o is 0.191. So, this is equal to 0.191, and that is what we have written here.

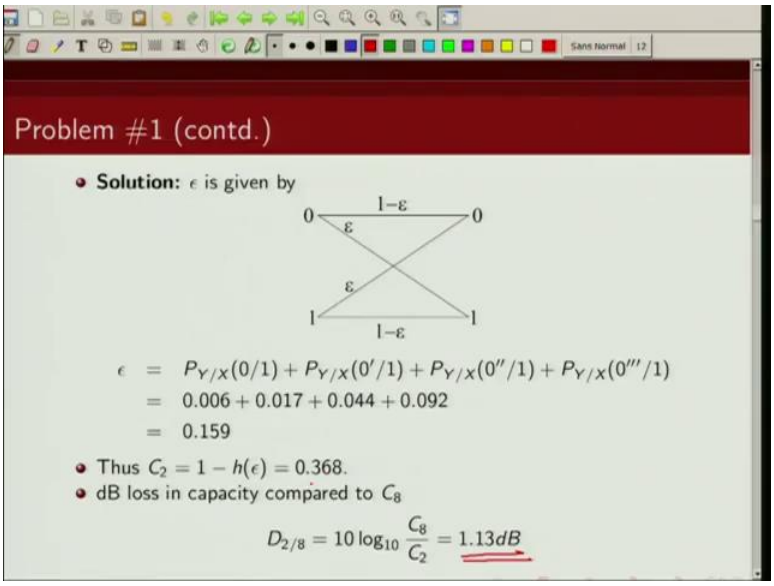
(Refer Slide Time: 09:34)



Similarly, we can write this and solve for epsilon 2 q 2, is exactly the same procedure that we get to find q 1 and epsilon 1. We can repeat this process to get q 3 epsilon 3. So, this is the value of epsilon 2 q 2, this is the value of 3 q 3 that we get, following the same procedure, and this is the value of epsilon 4 and q 4 that we get. So, now that we have this selection probabilities of each of these, so strongly symmetric channel, and each of these channels are binary symmetric channel which is the strongly symmetric channel.

So, we know the expression for capacity for each of these channels. So, then the overall channel we can capacity, we can write as summation of capacity of each of these strongly symmetric channel multiplied by their selection probability. So, if you sum that up, you get the capacity of a symmetric channel. So, and what is the capacity of each of these channel, these are a binary symmetric channel. So, it is capacity a binary symmetric channel is given by this. So, we plug that value in here, we know what epsilon is, we know this, we know what is q 1. So, we can calculate the overall capacity of this channel, and this comes out to be 0.477 bits.

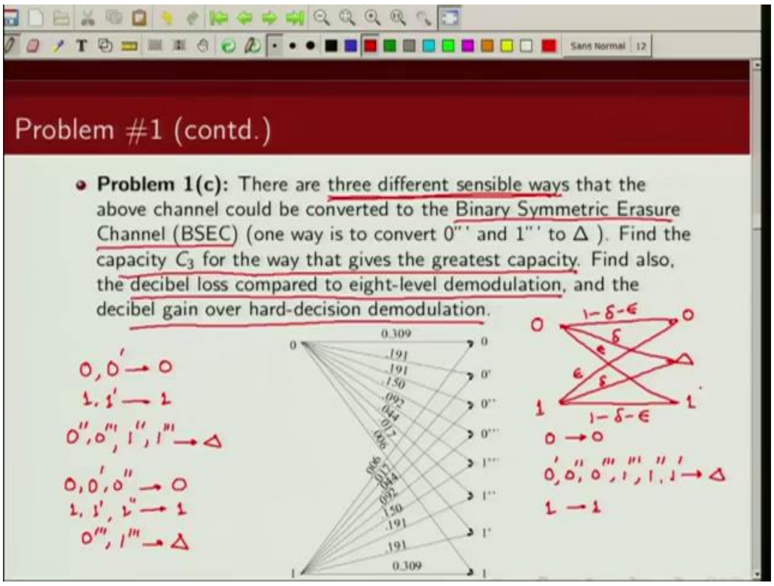
(Refer Slide Time: 11:26)



The next question what we have saying is, if you use the hard decision demodulation, then we would combine these four outputs in those into one symbol, call it zero. And we would combine these output symbols into the output symbol, let us call it one. Now I am asking so that would then result in per binary symmetric channel. So, this question ask, what is the capacity of the resulting binary symmetric channel, if you combine all these inputs 0 0 dash 0 double dash 0 triple dash into 1 output 0, and do the same thing with these 1 1 dash 1 double dash and 1 triple dash. So, that resulting channel would something like this. So, you have two inputs 0 1 1 and we will also have two output 0 1 1, and what are these transient probability. What is the probability of getting zero given input zero that would be probability of, equal to probability of zero getting zero plus probability of 0 double dash given 0 plus probability of 0 triple dash given 0, so that would be this probability.

And similarly we can compute the other probability as well. So, the resulting channel that we have is given by this, and this cross over probability as I said is given by probability of 0 given 1, probability of 0 dash given 1, probability of 0 double dash given 1, and probability of 0 triple dash given 1. So, if you add them up, we get epsilon to be this. Now this is a binary symmetric channel. So, it is capacity will be given by 1 minus (Refer Time: 13:39) to be function of epsilon. So, this is nothing, but this many bits. Now, if you compare this capacity to the capacity that we computed for their original channel. The original channel capacity was 0.477 bits per use; whereas, capacity of this channel is 0.368. So, as you can expect if you are doing the hard demodulation of the output we are losing information. So, our channel capacity decreases. So, the loss in capacity compare to the previous case is around 1.13 db, this is 10 log of 0.477 divided by 0.368; that is 1.13 db.

(Refer Slide Time: 14:49)

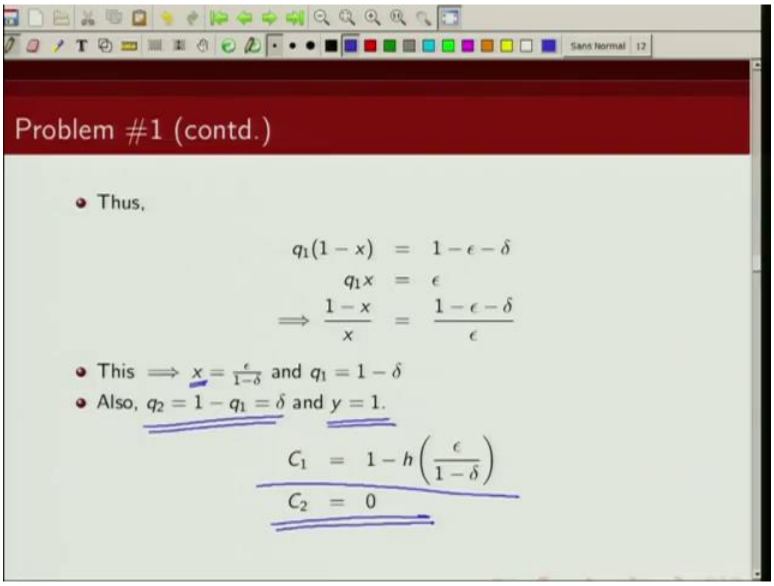


Now, the third part of this question is as follows. There are three different sensible ways of converting this channel into a binary symmetric erasure channel. Now what is a binary symmetric erasure channel? So, in a binary symmetric erasure channel you have two inputs 0 and 1, and you have three outputs zero erased output and 1. So, one second, let me draw the model for binary symmetric erasure channel. So, let say this is a delta probability of bits getting erase, let say this is an epsilon probability of bits b in error. So, this will be 1 minus delta minus epsilon 1 minus delta minus epsilon. So, this is a binary symmetric erasure channel.

Now, I am mentioning there are three different sensible ways to convert this original channel into a binary symmetric erasure channel. Now what are those three different ways; one way is I club 0 dash 0 double dash 0 triple dash 1 1 dash 1 double dash and 1. I convert all of them into eight bits in a map 0 to 0 and 1 to 1; that is one way of doing it. Second way of doing it I map 0 and 0 dash 2 0 1 and 1 dash to 1. and I map 0 double dash 0 triple dash 1 double dash and 1 triple dash, I map it to the erase bit, or I can do 0 0 dash and 0 double dash, and map it 2 0, 1 1 dash and 1 double dash I map it to 1, and I map 0 triple dash and 1 double dash I map it to the any aspect.

So, there are three ways in which I can convert this particular channel into a binary symmetric erasure channel. Now this question asks, find the capacity of the channel, and find the capacity for the way that gives the greatest capacity. So, out of this three different ways in which we can convert this channel into a binary symmetric erasure channel, which one will give us maximum capacity; that is the first part of this question, and second part of the questions says find also the decibel loss compare to the eight level demodulation, and the decibel gain over hard demodulation. So, is asking us to compute the losing capacity, compare to this original channel, and gain in the capacity compare to the binary symmetric channel that we did in the problem 1 b.

(Refer Slide Time: 18:28)



So, they said this is our binary symmetric erasure channel. This is erase bit denoting by delta, and these are the other received bit zero and ones. With delta probability the bits are getting erased bit probability epsilon the bits are getting in error. So, the probability of receiving the bits correctly is 1 minus epsilon minus delta. Now this binary symmetric erasure channel can be decomposed into two strongly symmetric channel; the one which as input 0 and 1 the other which has this erase bit. You can see if you look at this output 0, the transient probabilities here are 1 minus epsilon minus delta, and epsilon similarly if you look at this input these transient probabilities is a 1 minus epsilon minus delta, and this is epsilon. So, 0 and 1 have the same focusing. So, they can be clubbed together.

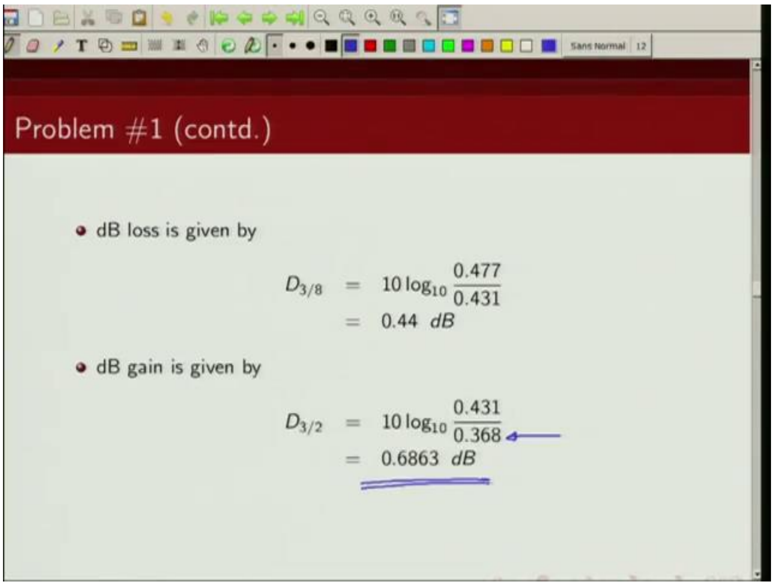
However it is focusing it is difference from to focusing of delta which is delta and delta. So, this is not a strongly symmetric channel, but we can decompose it into two strongly symmetric channels; one which has output zero, and one the other which will have output delta, and that is what we did here. We decomposed into two channels; one which is output zero, and one other which as this erase bit.

Now following the same procedure then we need to compute what is the selection probability of selecting this channel which has in output zero and one, which we denote by probability q 1. And similarly we need to find the selection probability of selecting this channel whose output is this erase bit delta, and that is denoted by q 2 on this corresponding probabilities are similarly denoted by x and 1 minus x, and this is denoted by y. Now we will again check the transient probability of each of this output given input, and compare it with the original binary symmetric erasure channel, to get these probabilities.

So, we can see q 1 times 1 minus x is 1 minus epsilon minus delta. So, if you look at probability of 0 given 0, this is given by q 1 times 1 minus x this is q 1 times 1 minus x and let us go back to the original probability of 0 given 0; that is this, that is 1 minus epsilon minus delta. So, that is why we wrote q 1 times 1 minus x is 1 minus epsilon minus delta. Similarly we can find out what is the probability of 1 given 0, this is nothing, but q 1 times x, so this is q 1 times x now let us look at. Now let us look at it is original channel probability of 1 given 0, this is given by epsilon. So, the epsilon is equal to this and that is what we wrote here. So, we have two equations; this is equation one, this is equation two, and we have two unknowns; one is q 1, other is x. We solve for it and we get the value of x to be this, and q 1 to be this. Similarly we can find out what is q 2 and y.

We can find out that. So, if you do that y comes out to be 1 and q 2 comes out to be 1 minus q 1 which is delta. Now the first set of channel this, this channel which I denoting by c 1 this is the binary symmetric channel. So, it is capacity is given by one this, and capacity of this channel is zero. So, we plug that in. this is the capacity of channel one, where this cross over probability x is given by epsilon divided by 1 minus delta, and the capacity of the second strongly symmetric channel is zero. The overall capacity is given by q 1 times c 1 plus q 2 time c 2, which comes out to be this.

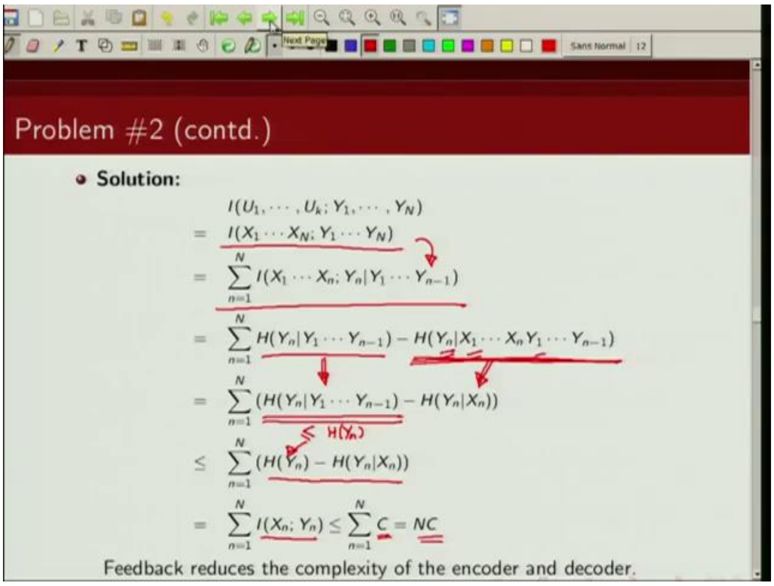
(Refer Slide Time: 24:04)



Now, as I said there are multiple three different ways, in which I could combine the outputs. If I combined 0 triple dash and 1 dash to be delta, my epsilon in this case comes out to be 0.067, and delta comes out to be 0.242. So, capacity in this case comes out to be this, and similarly for the two other cases I can find out the capacity. So, you can see somewhere out of the three ways of combining, this case should be the best to the highest capacity. So, what I do is, I map 0 and 0 dash and 0 double dash to 0, I map 1 1 dash and 1 double dash, I map it to 1, and I map 0 in triple dash and 1 triple dash I map it to delta, see if I do that I get the maximum capacity which is 0.31 bits.

Now, if we compared this with the capacity of the original channel which was 0.477 bits. So, this is a 0.44 db loss; however, if you compare it to the capacity of the binary symmetric channel in problem 1 b. This is a gain of capacity of around 0.69 db. So, let us look at this problem, which shows a channel with feedback. So, you have a source u bit 7 coded into x, and this is said to a discrete memory less channel whose output are y, but now this is a feedback from the output to the input.

(Refer Slide Time: 26:03)



So, the channel output is fed back found is fed back to the input side. So, now, this selection of x depends also on what is been fed back. So, subsequent selection of x depends not only on u, but it also depends on y, which is the symbols which have fed back from the channel. Now we want to show that, the channel capacity of this channel with feedback, for a discrete memory less channel with feedback.

Feedback does not increase channel capacity. So, a capacity of a discrete memory less channel is not increased by results of presence of feedback from the receiver to the transmitter. Now, the question is, a feedback does not increase the capacity of a discrete memory less channel, why should we do feedback from the output to the input. So, let us prove this. now this want to make comment that, if the channel is, channel has memory then feedback can increase the capacity of the channel; however, if the channel is discrete memory less then feedback does not increase the capacity of the channel. So, we want to prove that, presence of feedback does not increase the capacity of this channel.

So, let us do that. So, the mutual information between u 1 u 2 u 3 u k and y 1 y 2 y 3 y n can be written using definition of mutual information like this. So, this is joint entropy of y 1 y 2 y 3 y n minus conditional entropy of y 1 y 2 y n give u 1 u 2 u 3 u n, which using the chain rule can be written in this particular function. So, this is h of y 1 y 2 y n minus h of y 1 y 2 y n given u 1 u 2 u k, and this using chain rule can be written like this. Now, what I did was, I further condition this on x 1 x 2 x 3 x n. Now, since I am further conditioning on x 1 x 2 x 3 x n, we know that, this will be greater than equal to this particular term, is equal to happening if x would not. So, we know that uncertainty in let say a given b is more than equal to uncertainty a given b c right. So, that is why this term was larger than this term. So, since we are subtracting a smaller term, I have written here greater than equal to.

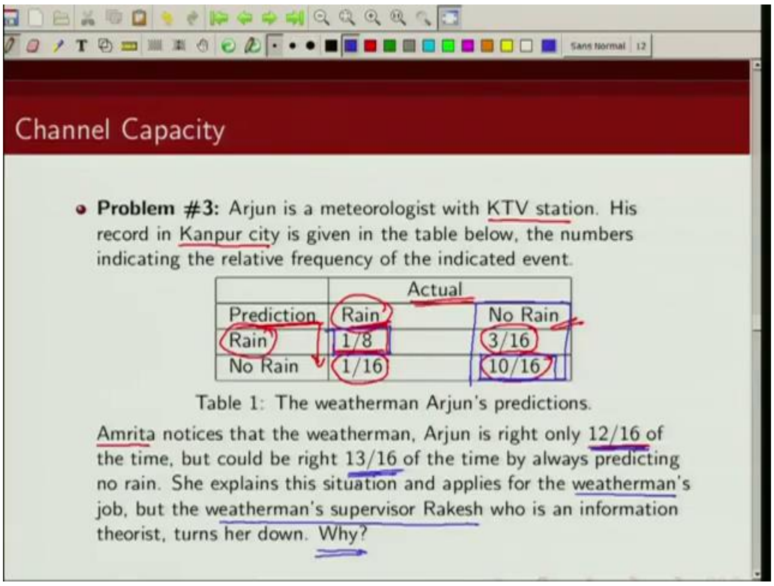
Now, what is the uncertainty in y n given y 1 y 2 y n minus 1 and u 1 u 2 u 3 u k and x 1 x 2 x 3 is x n. given x 1 x 2 x n y n is independent of u 1 u 2 u k. So, I can write this as this. So, this then becomes joint entropy of y 1 y 2 y 3 y n. you can see here given x n, y does not depends on u. So, that is why you could write this like this. Now using chain rule I can also write this joint entropy as uncertainty in y n given y 1 y 2 y n minus 1, where n goes from 1 to n minus this is come from here. Now from the definition of mutual information, this can be written as mutual information between x 1 x 2 x n and y 1 y 2 y n. So, what I have written so far is mutual information between u 1 u 2 u k and y 1 y 2 y 3 y n is less than equal to mutual information between x 1 x 2 x 3 xn and y 1 y 2 y 3 yn.

Now, proceeding further I can write this mutual information between x 1 x 2 x 3 x n and y 1 y 2 y 3 y n, using chain rule for mutual information I can write this in this particular form. So, I have applied to chain rule for mutual information. So, I can write this mutual information in this particular form. Now from the definition of mutual information I can write it in terms of entropy. So, this can be written as uncertainty in y n given y 1 y 2 y n minus n 1 minus uncertainty in y n given y 1 y 2 y n minus 1 and x 1 x 2 x n. Now, what is the uncertainty in y n? Given - x 1 x 2 x 3 y 1 y 2 y n minus 1.

Now, note that we are talking about a discrete memory less channel. So, what is the uncertainty in y n given x n x n minus 1 and y 1 y 2 y n minus 1, it only depends on xn. So, then we can write this as, uncertainty in y n given x n. So, this term comes as it is, and because it is a discrete memory less channel uncertainty in y n given x n and all other parameter, because it is just depend on x n.

Now, we can further simply this term, this is entropy of y n given y 1 y 2 yn minus 1. now we know that conditioning cannot increase entropy. So, each of these individual terms will be less than uncertainty in y n. So, invoking that property we can write this as less than equal to h of y n, and then we do it like this we write it like this. Now what is uncertainty in y n minus uncertainty in y n given x n. this from the definition of mutual information is nothing, but mutual information between x n and y n and we know what is the maximum mutual information in x n and y n over all possible input distribution that is given by capacity c. So, this can then be written as equal to n times c. So, what we have shown is, mutual information between ui's and yi's is less than equal to n time c. So, feedback from the output to the input does not increase capacity for a discrete memory less channel. However it does reduces the complexity of the encoder and decoder, and that is why we do commonly use feedback in our communication systems. Let us look at the first problem.

(Refer Slide Time: 35:29)

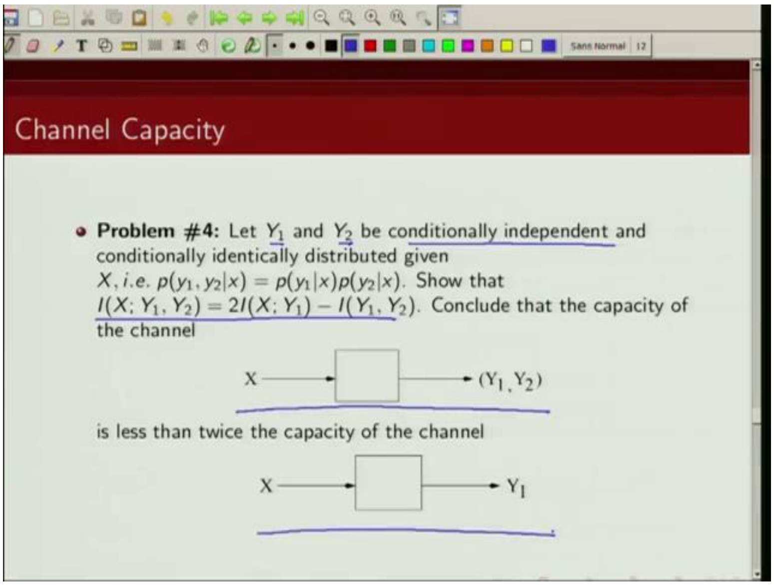


So, Arjun is a meteorologist with a TV station let us call it K TV his records for predicting the weather in Kanpur city is given below. So, this is what the actual happening. So, this is whether the rain happens or whether the rain does not happen, and on this column, I have the prediction of Arjun. So, when actually rain happens, and it predicts rain that probability is 1 by 8. And when there is no rain, and it predicts that rain is going to happen that probability is 3 by 16. Similarly, one day actually rain and it predicts that there is no rain that probability is 1 by 16. And probability that there is no rain, and you actually predict is correctly there is no rain is given by 10 by 16. So, these are related frequency of prediction.

Now a student Amrita notice that this weatherman Arjun is right only 12 by 16 times, he is right when he predict correctly the rain happens, which is this probability and it is correct when there is no rain and you predict you correct there is no rain. So, Arjun is right only 12 by 16 times. So, it is 3 by 4 times is correct, and 1 by 4 times it is wrong.

Now she notices that if you always predict no rain, what is the probability. You always predict if you always predict. So, he could be right 13 by 16 times by always predicting there is no rain. So, if you predicts always there is no rain. now Amrita explains this situation to an information theorist, he is supervisor she explains this situation, and applies for the job of weatherman forecast weather forecast reporting; however, her supervisor Rakesh who is an information theorist, he rejects her application. So, you have to tell why Amrita's application was rejected, why her prediction, her way of splitting the weather which is right 13 by 16 times, was rejected by her supervisor through the information theorist. So, the answer to this question lies. So, we are going to look at actual weather and prediction as input output relationship, and we are going to look at the capacity of such channel. And he will show that Amrita's channel gets less mutual information compare to the channel of Arjun, and Rakesh who is an information theorist that is why rejects the application of amrita.

(Refer Slide Time: 39:06)



So, let us define a random variable x representing the actual weather. So, when x is equal to 0; that is there is no rain, and when x is 1, it means rain has occurred. He also defines another random variable, we call it y, and this is the prediction of weatherman. So, when y is equal to 0; that means, the weather man is predicting there is no rain, and then y is 1 the weather man is predicting that there is a rain. Now, if we view this weatherman has a channel that conveys some information about an underline process.

In this case the event is whether there is rain or there is no rain, then we can define transition probabilities. So, what is the probability that a weatherman will say there is no rain, given there is no rain. What is a probability that weatherman will predict, there is rain when there is actually no rain. What is a probability that the weatherman will predict more rain, when there is actually rain? And what is the probability of rain when there is actually rain. So, we can use these four transition probability to describe this weather men's channel. Similarly we could also define a prior probability of having rain or no rain.

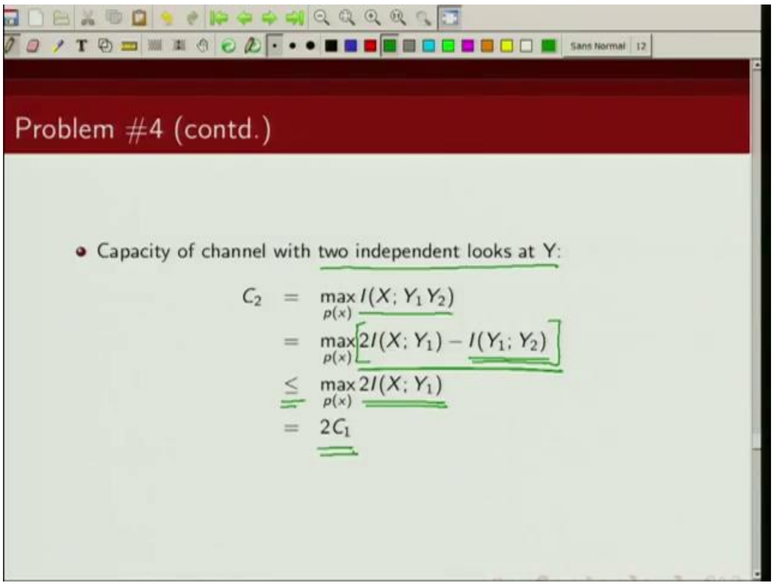
So, you can compute probability of x being 0, which is nothing, but if you sum of this joint probability of x being 0 and y, is a more all y's that comes out to be 13 by 16. Similarly probability of x being 1 comes out to be 3 by 16. Next this transition probability for the Arjun channel, we compute. So, probability of y given x equal to 0, can be given as, can be written as probability of x equal to 0 y equal to 0 divided by probability of x equal to 0, and that is given by 10 by 13. similarly probability of y equal to 1 given x equal to 1, can be written as probability of x equal to 0 and y equal to 1 divided by probability of x is equal to 0; that is probability 3 by thirteen. And similarly we compute the probability of y being 0 given x equal to 1 that comes out to be 1 by 3, and probability of y being 1, given x 6 1 comes out to be 2 by 3.

So, this is Arjun's channel; one denotes there is rain, zero denotes there is no rain. This actual value, this separated value, and these are the transition probabilities. These are a prior b prior probability on x. So, we can find out what is the probability of y being 0 that is 11 by 16, and probability of y being 1, that is 5 by 16. If you use this compute the mutual information, Arjun's channel is going to be worse roughly 0.09 bits of information. Now, let us look at Amrita's channel. Now what is Amrita's channel do Amritas channel always predicts no rain. So, no matter whether actually there is rain or no rain, Amrita's channel always predicts there is rain. Remember y denotes reflected value, x denotes actual weather. So, this is Amrita's channel.

So, similarly we could compute probability of y being 0 given x equal to 0 that is 1, this probability of y being 1 given x equal to 0 comes out to be 0, probability of y being 0 given x is 1 comes out to be 1, and probability of y being 1 given x equal to 0 comes out to be 0. If you compute the probability of y equal to 0, in case of amrita's channel that is always one, because she is always predicting that there is no rain and probability of y equal to 1 is 0. So, if you compute the capacity of Amritas channel that comes out to be zero.

And this is the reason the weatherman's supervisor Rakesh, who was information theorist rejected Amrita's application, because her channel conveys zero information. Whereas, if you go back and look at Arjun's channel, it gives 0.09 bits of information, and that is why Arjuns channel in the information theoretic cells is better than Amrita's channel. Now let us look at next problem. So, we have y 1 and y 2, and they are given (Refer Time: 44:39) and they are conditionally independent, and conditionally identically distributed. So, what do we mean by identically distribution given x, what do we mean probability of y 1 and y 2, given x is given by probability of y 1 given x into probability of y 2 given x. show that mutual information between x and y 1 y 2 is 2 types mutual information of x and y 1 minus y 1 y 2, and use this relation to conclude that capacity of this channel is less than twice of capacity of this channel.

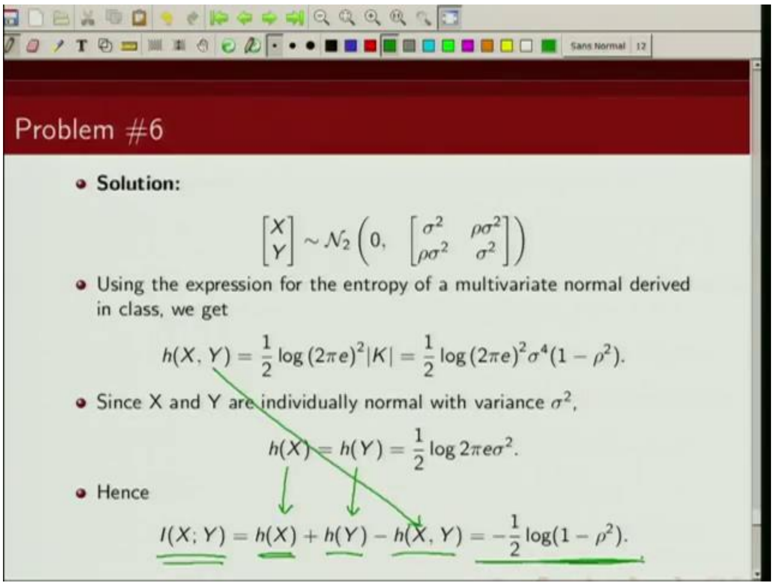
(Refer Slide Time: 45:29)



So, mutual information between x 1 y 1 y 2, using definition of mutual information I can write this as, entropy of y 1 y 2 minus entropy of y 1 y 2 given x. now this joint entropy between y 1 y 2 can be written as uncertainty in y 1 plus uncertainty in y 2 minus mutual information between y 1 y 2. And this particular term, because given y 1 y 1 y 2 are conditional independent. So, uncertainty in y 1 y 2 given x can be written as uncertainty in y 1 given x plus uncertainty in y 2 given x. So, this term can be split into these two terms all right. Now we can take this term and this term. This is nothing, but mutual information between x and y 1, and take this term and this term, this can be as mutual information between x and y 2, and this we write as it is. So, mutual information this comes out as mutual information between x and y 1 plus mutual information between x and y 2 minus mutual information between y 1 and y 2.

Now let us look at capacity of channel with single look. So, that capacity is given by mutual information between x and y 1, where we maximize mutual information over all input distribution. Now, let us look at this capacity of this channel with two independent look. So, we have just now showed that mutual information between x and y 1 y 2 can be written like this. Now mutual information is greater than equal to 0. So, then this whole term is less than equal to this, this is less than equal to two times the capacity of a single look channel. So, this proves our result that capacity effect channel with two independent looks, is not twice of capacity of a channel with single look.

(Refer Slide Time: 48:24)



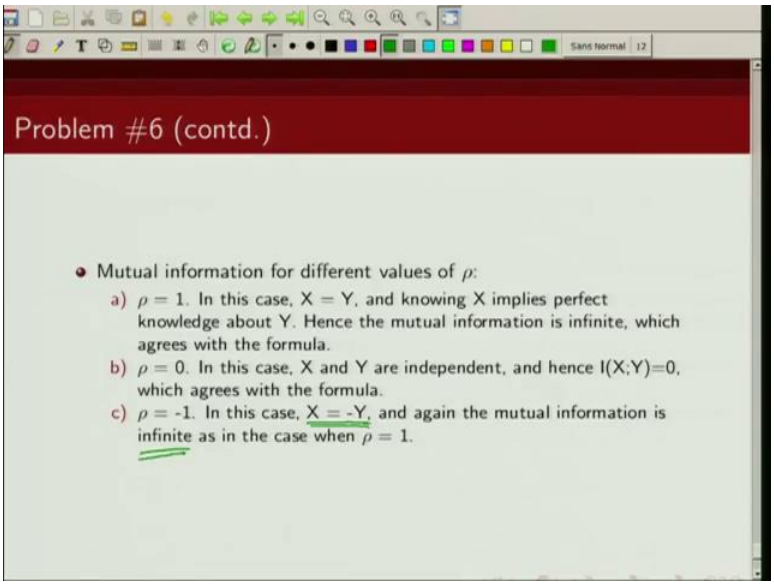
Next consider a pair of parallel Gaussian channel. So, we have two parallel Gaussian channel; first channel the output y 1 is given by x 1 plus z 1. Now the second Gaussian channel x 2 is y 2 is given by x 2 plus z 2, and z 1 z 2 is Gaussian distributed with this mean, covariance matrix is this. There is a power constraint which is given by expected value of x 1 square plus x 2 square is less than 2 times p. And let us assume that the first channel is bad compare to the second channel. So, sigma 1 square which is the noise variance corresponding to the first channel; this is higher compare to the noise variance of the second channel.

So, the question is at what power the channel stops behaving like a single channel with noise variance sigma square 2, and begins like a pair of channels. So, if you recall from our discussion on water filling algorithm. So, if you have pair of Gaussian channel, we will starts first flouring, we first start giving allocating power to the channel which is less noisy until. So, it will continue to behave like a single channel, until it reaches a point, when the sum of power and noise variance becomes equal to the noise variance of the second channel. So, once that happens, and then power will start getting distributed in both the channels. So, this is very straight forward.

So, we will put all signal power into the channel with less noise this follows from the optimal power allocation strategy, until the total power of noise and signal in that channel equals the noise power in the other channel. And subsequently what is going to happen is we are going to split the power between those two channels. So, the combine channels begin to behave like a pair of parallel channels, when the signal power is equal to a difference of the noise powers. So, when this happens it start behaving like a pair of Gaussian channel. Now, let us look at what is the mutual information if Gaussian generated random and multivariate Gaussian random variable, with distribution given by this.

So, let us compute mutual information for various values of rho correlation. So, as we are given that, this is a multivariate Gaussian distribution random variable. So, we know what is this entropy; entropy of a multivariate Gaussian random variable is given by half log of. So, here any tools of the 2 bit 2 pi e square and the determinant of this covariance matrix. The covariance matrix determines is sigma 4 into 1 minus sigma square. So, you joint entropy in this case is given by this expression. Now, let us look at what is the differential entropy of these individual random variables x and y. this is given by half log of 2 pi e sigma squares. So, the mutual information between x and y can be written as mutual information between, can be written as differential entropy of x plus differential entropy of y minus joint differential entropy of x and y. this follow from the definition of mutual information. So, we know the values of this. We also know the value of this, we clubbed this in what we get is mutual information is given by minus half log of 1 minus rho square.

(Refer Slide Time: 53:30)



So, let us compute mutual information for different values of rho. So, when rho is 1. So, when rho is 1, this is log of 0. So, that would be, mutual information will be infinite, and this agrees with the fact, because x is y x m y have complaint information of each other. When rho is 0 this corresponds to the case, when x and y are independent. So, in that case we expect the mutual information to be 0. And similarly when rho is minus 1; that is a negatively correlated x is minus y, and then also we expect the mutual information to be infinite.

So, with this we are going to conclude our discussion on problem solving session. In the next class we are going to talk about rate distortion theory.

Thank you.