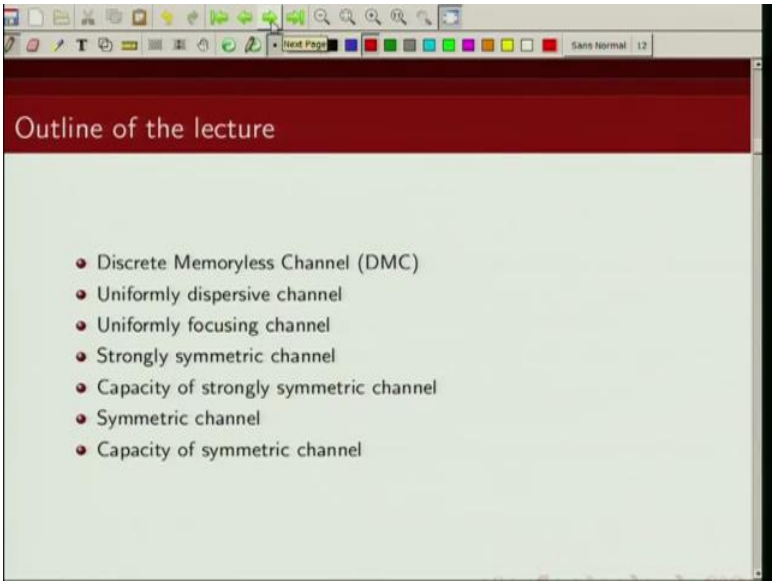
An Introduction to Information Theory
Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture – 09
Channel Capacity

Welcome to the course on an introduction to information theory. In today's lecture we are going to talk about, how to compute channel capacity of some very simple channel. If you recall, in the introduction we talked about what is channel capacity and channel in a celebrated noisy channel coding theorem. As mention that long as we transmit at rate below capacity, We can achieve reliable communication now we are going to formally state channel noisy channel coding theorem and its prove it and prove its converts in the next lecture.

In today's lecture we will be dealing with some very simple channels.

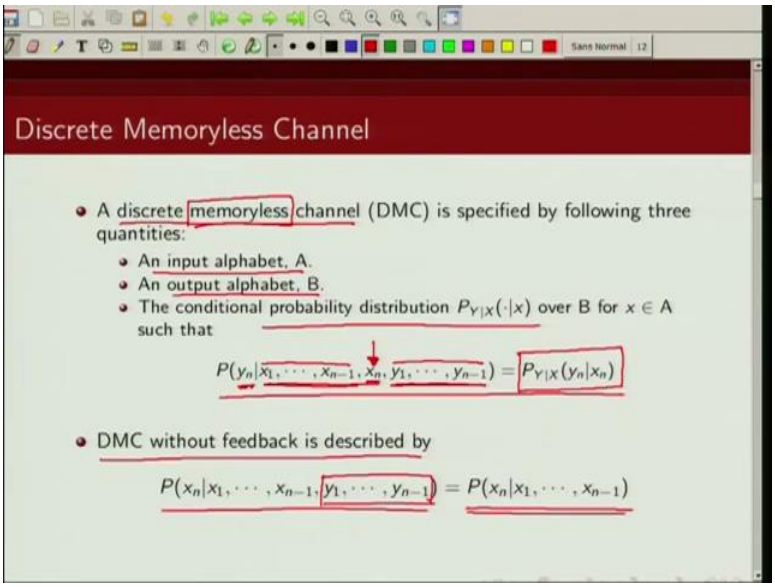
(Refer Slide Time: 01:06)



Ah we are looking at discrete memoryless channel, and we will try to compute capacity of those simple channels. So, we will start up with the description of a discrete memoryless channel, and then will describe 2 types of channel; one is call uniformly dispersive channel, and other is called uniformly focusing channel. And channels which satisfies both of this property are known as strongly symmetric channel and we will compute the capacity of a strongly symmetric channel and subsequently, we will show

that there is a class of channel which can be decomposed in to strongly symmetric channels and that is basically known as symmetric channel and we will compute the capacity of symmetric channels. So, this is a plan for this lecture.

(Refer Slide Time: 02:02)

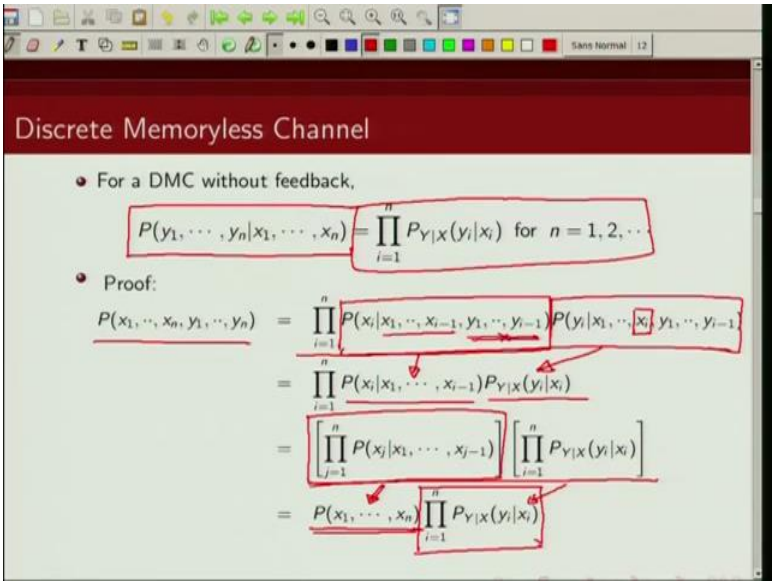


So, a discrete memoryless channel is specified by these quantities. One is input alphabet, if you denoting by a, output alphabet b and the conditional probability distribution over b for each x. So, we need to describe input and it is power output and we need to describe the conditional distribution of y given x, for a discrete memoryless channel. As you can see for a discrete memoryless channel, probability of y n given passed inputs which are given by x1 x2 xn minus 1 and the current input x n and the passed outputs y1, y2, yn minus 1 because it is a memoryless channel it does not depend on passed values. So, the probability of y n given x1, xn minus 1 x of n and y1, yn minus 1 it only depends on the present input because it is a memoryless channel, it does not depends on pass values of xs it does not depend of pass values of y's it only depends on the current value of x.

So, for a memoryless channel probability of y n given pass values of ys or pass values of xs it only depends on the current value of x. So, probability of y n given x1, x2, xn, y1 to y n minus 1 it only depends on y n given x of n. Now we can see a discrete memoryless channel without feedback as follows. So, probability of x1, x2 probability of x n given x1 x2 xn minus 1 and y 1 to y 1 minus 1, it only depends a probability of x n given x1 to

xn minus 1. So, it does not depend on the pass values of the output. So, this is our discrete memoryless channel without feedback.

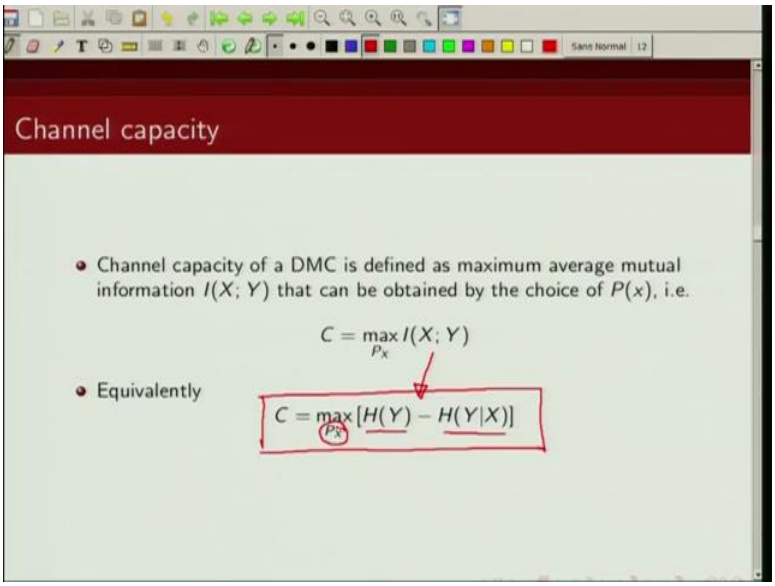
(Refer Slide Time: 04:34)



So, for a discrete memoryless channel without feedback, we can write probability of y1, y2, yn given x1, x2, xn as this product of these probabilities of y I given x i. So, this we can prove in this particular fashion we write the joint probability x1, x2, xn; y1, y2, yn. So, this can be written in this particular fashion. So, product of I going from 1 to n probability of x i given x1, x2 x i minus 1, y1, y2, yi minus 1 probability of i, I given x1, x2, xi; y1, y2, yi minus 1 now because it is a memoryless channel this term is equal to this. So, probability of y I given x1, x2, xi and y1, y2, yi minus 1 it only depends on present value of the input. So, this term can be written as in this particular fashion.

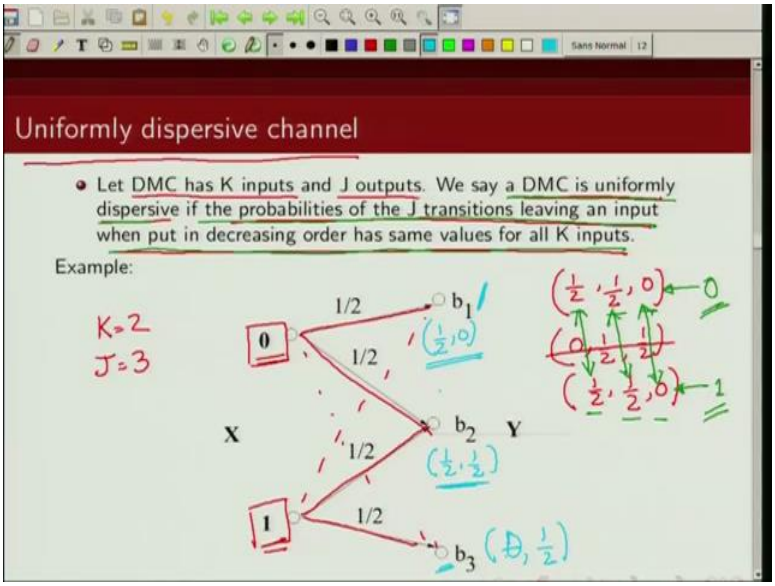
Since we are talking about discrete memoryless channel without feedback, this term can be written. So, probability of x i given x1, x2, xi minus 1 and y1, y2, yi minus 1 it does not depend on this it only depends on these values. So, I can then write this particular form. So, then I can write this in this particular fashion and what is this term this term is nothing, but probability of x1, x2 and xn multiplied by this term. Now if I divide both side by this on the left hand side I will get this and the right hand side I will get this. So, hence this proves that if I have a discrete memoryless channel without feedback probability of y1, y2, yn given x1, x2, xn is given by this expression probability of yi given xi and product of that all of these.

(Refer Slide Time: 07:08)



Now, we define channel capacity of a discrete memoryless channel as the maximum average mutual information that can be obtained over choice of p of x. So, capacity is defined as maximum mutual information between the input and the output of a channel and this maximization is taken over all possible input distribution. Now, mutual information we can also write from the definition of mutual information we can write this as uncertainty in y minus uncertainty in y given x. So, equivalently we can write the expression of capacity as maximizing the difference between h of y minus h of y given x maximizing over all possible input distribution p of x.

(Refer Slide Time: 08:09)



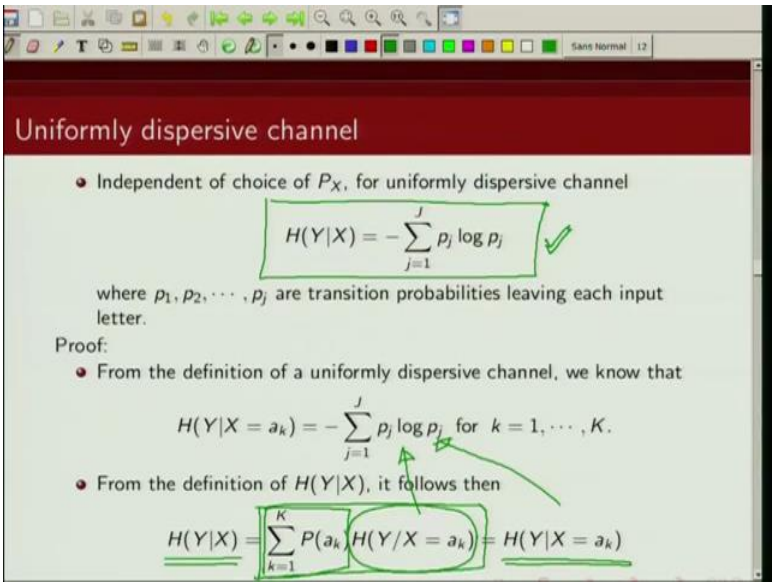
Now, let us consider some very simple channel. So, first channel that we will consider is what is known as uniformly dispersive channel. So, we are considering of discrete memoryless channel which as k inputs and j outputs. For example, in this particular case k is equal to 2 and j is equal to 3. Now we see a discrete memoryless channel is uniformly dispersive if the probability of the j transitions leaving an input when put in descending order has the same values for all k inputs. So, what does it mean? So, we look at each of this input we look at each of this k inputs and we write the transition probabilities leaving that particular input.

For example, take this input 0. So, what are the 3 transition probabilities this is half because this transition probability is half this transition of probability is half and this transition probability is 0. So, if I arrange this transition probability in the descending order this will be in order. Now look at this input this is 0 this transition probability is half this transition probability is half, now if I arrange this in descending order this would be half, half and 0.

Now, what does a definition says for a uniformly dispersive channel it says that a discrete memoryless channel uniformly dispersive if the probabilities of the j transitions leaving an input. When put in descending order has the same value for all the k inputs now when I put the j transitions corresponding to input 0 in descending order this is what I got half, half 0. Now when I did the same thing for input 1 I got half, half and 0, now note this is same.

So, when I arrange these transitions probability in descending order it is same whether, I consider 0 input or 1 input and that is what my definition says. So, this is an example of uniformly dispersive channel now for a uniformly dispersive channel.

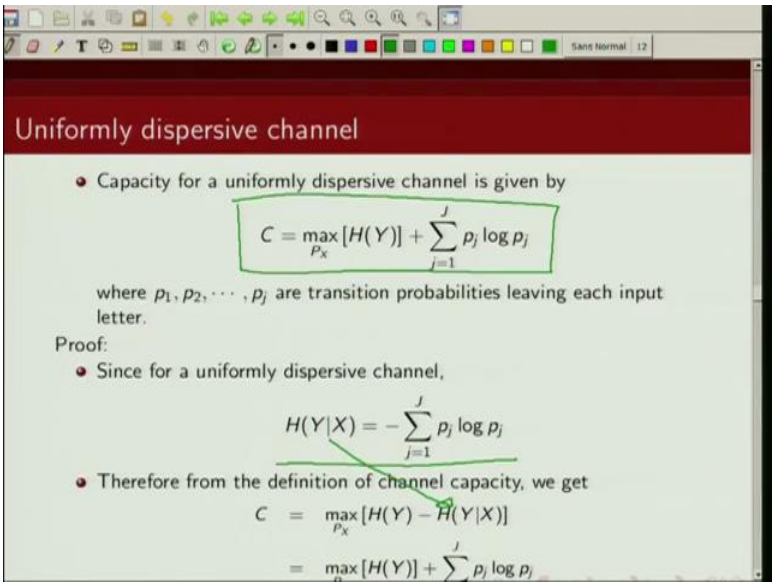
(Refer Slide Time: 11:21)



Independent of the choice of input distribution we can write the uncertainty in y given x as summation of minus p j log p j where this p js are transition probabilities leaving each input letter. So, let us prove this property. So, from the definition of conditional entropy of y given a particular instance of x a k this can be written as minus summation over all possible transitions p j log p j. Now we know that because this transition p1, p2, p3, pj. When they are arranged in the descending order it is same no matter what input you consider. So, from the definition of conditional entropy of y given x, we can write this as this. Now note this term; this term which is given by this expression this is same for all inputs for a uniformly dispersive channel if you take this out then you are left with summation of p a k sum over all k's, that summation is 1.

So, then we can we have proved that uncertainty in y given x is given by uncertainty of y given x is sum a k and this is nothing, but given by this. So, we have proved that no matter what is the choice of input distribution the uncertainty in y given x for a uniformly dispersive channel is given by this expression.

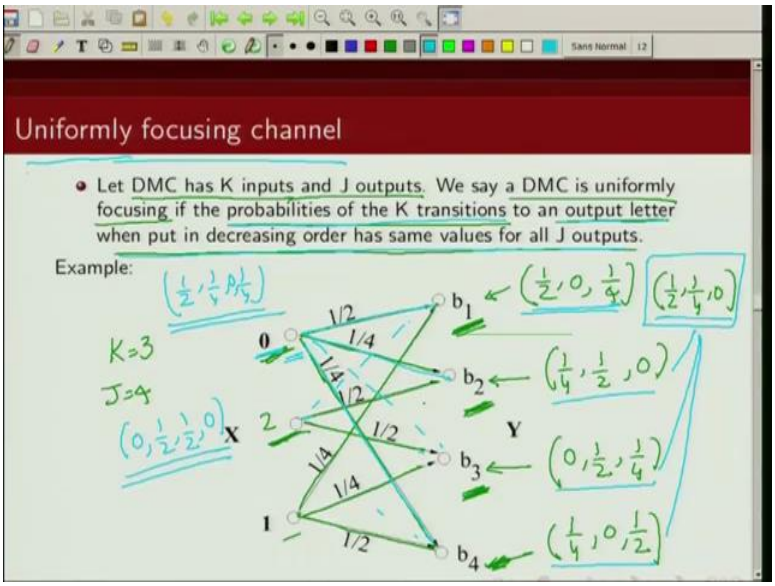
(Refer Slide Time: 13:50)



Next let us. So, then we can write down the expression for capacity for a uniformly dispersive channel, now what was capacity. So, this was maximizing over all input distribution h of y minus h of y given x of this quantity is no matter what is the choice of p of x this quantity is fixed for a uniformly dispersive channel. So, minus of this is essentially given by this expression. So, in other words if you want to find out the capacity of all uniformly dispersive channel this is given by maximizing h of y over all input distribution p of x plus this quantity and we know and why the discrete random variable it is maximize the entropy is maximize when it is uniform. So, we need; capacity will be maximized for the input distribution p of x which will induce a uniform distribution on y.

So, this follows a straight away we just prove that uncertainty in y given x is given by this expression, now if we plug this value here in the expression for capacity we get our desired expression. So, this is the expression for capacity for a uniformly dispersive channel.

(Refer Slide Time: 15:41)



Now, let us now talk about another type of channel which is known as uniformly focusing channel. So, what is a uniformly focusing channel? So, let us consider a discrete memoryless channel which has k inputs and j outputs. So, in this particular example k is 3 because of 3 inputs this 1 this 1 and this 1 then call this 0, 1, 2 and whatever you want to call it and there are 4 outputs j is 4 b1, b2, b3, b4.

Now we see a discrete memoryless channel is uniformly focusing if the probabilities of k transitions to on output letter when put in decreasing order has the same value for all j outputs. So, what we need to do is we need look at the output letters we look at b1, b2, b3 and b4 and look at the probabilities of k transition to this output letter. So, let look at b 1 the transition from here this has probability half b1 transition from this node 2 is 0 and transition from this has probability 1 by 4.

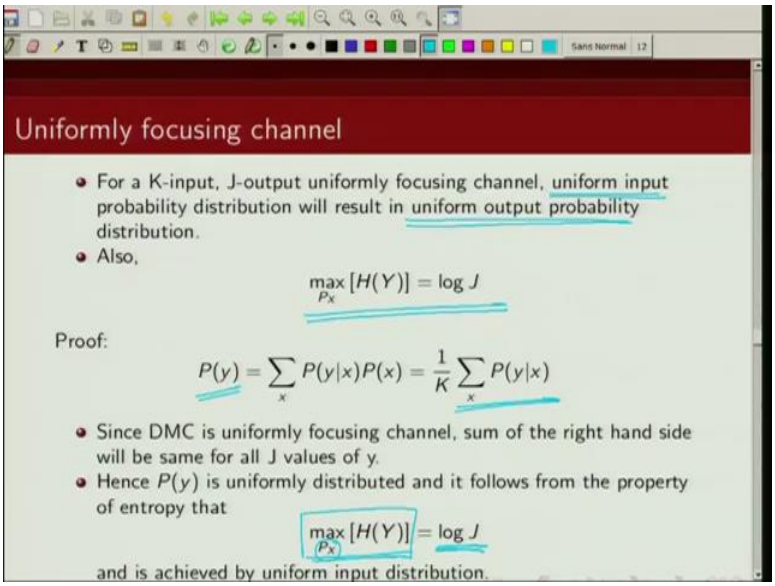
So, this is what is particular output look at this particular output transition from this particular input 0 has probability 1 by 4 transition from here has probability half and transition from here this is 0 this is no transition similarly b3. There is no transition from this input 0 from this input transition probability is half and from this input the transition probability is 1 by 4 and for this particular output transition probability from 0 is this 1 one by 4 transition probability from this particular input is 0 and transition probability from this particular input is half.

So, if I arrange all of them in the descending order for whether I consider this output this output this output or this output I will get this. So, node that. So, probabilities of all this k transitions to any output letter when put in decreasing order if they have the same values for all j output then it is a uniformly focusing channel and in this case you can see if I put this in descending order I get this if I put this in descending order I get this if I put this in descending order I get this probability similarly if I put this in descending order I get this. So, no matter what my output is whether it is b1, b2, b3 or b4? I get this same probability order or half 1 by 4 0. So, this is an example of a uniformly focusing channel.

Now, let us go back and look at this example that we did is this a uniformly focusing channel just look at this. So, if you look at b1 the transition from here is half from this particular node there is no transition what about this node the transition probabilities are half and half and transition for this particular node transition probabilities are 0 and half. So, you can see this is not an example of uniformly focusing channel why because, when I arrange the probabilities in the descending order for these particular output b1 and b3 it is half and 0. Whereas, for this particular output it is half and half, this is a uniformly dispersive channel, but it is not a uniformly focusing channel similarly look at this example is this uniformly dispersive channel.

So, for the uniformly dispersive channel what we need to look at we need to look at the transition probability leaving that particular input for this particular input is 0 these transition probabilities are half 1 by 4, 1 by 4 for this transition probabilities are. So, just a minute this is this is half this is 1 by 4 this guy is 0 and this is 1 by 4 for this is 0, 0 this is half this is half and this is 0. So, clearly if I arrange this probability in the descending order we arrange this probability in descending order they are not same. So, this is an example of a uniformly focusing channel, but it is not a uniformly dispersive channel. So, I hope the difference between a uniformly focusing channel and a uniformly dispersive channel is clear.

(Refer Slide Time: 21:45)



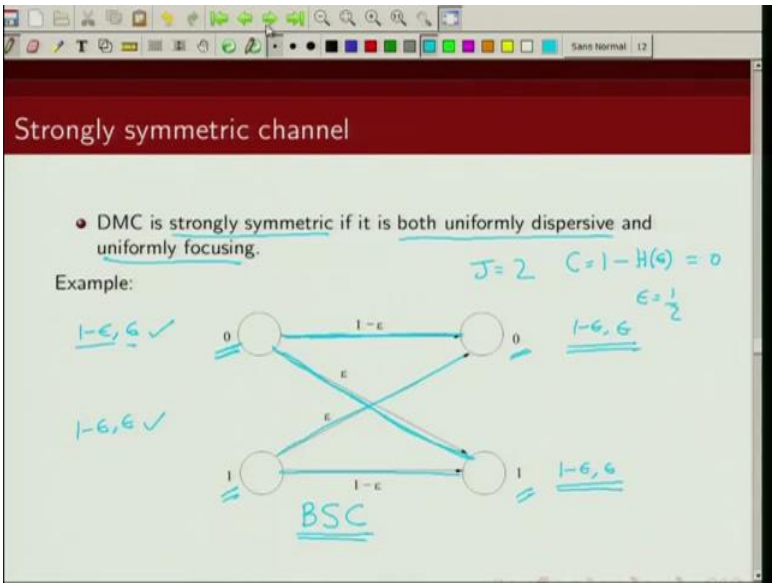
Now, what can we say about a uniformly focusing channel for a k input j output uniformly focusing channel uniform input probability distribution will result in uniform output distribution. So, let us prove this and also maximum value of h of y is given by log of j where j is a number of possible output values. So, p of y can be written in this particular way right? Now if we consider uniform input distribution since I have k inputs. So, p of x will be 1 by k. So, if I plug this value of p of x here then p of y become this now what is this term these are the transitions terminating at output y coming from all possible inputs now for a uniformly focusing channel this term is same no matter, what is the value of I in this example if you see whether my output is b1 or b2 or b3 or b4 summation of probability of y given x is same whether I consider y to be b1, b2, b3 or b4. So, for a uniformly focusing channel this term is a constant.

So, this is a constant then probability of y is a constant. So, that is a uniform distribution. So, for a uniformly focusing channel if your input is uniform then, it will induce a uniform output distribution and since maximum value of h of y is then we know that h of y lie between 0 and log of j. So, an input distribution will result in output distribution which is also uniform and maximum value of h of y for a uniformly focusing channel is given by this.

Now, what about channels, if you noticed for a uniformly focusing channel we are able to get maximum value of this and for a uniformly dispersive channel we were able to get

the expression for h of y given x. So, what if a channel is both uniformly dispersive and uniformly focusing then we have the exact expression for its capacity and that is what we are showing.

(Refer Slide Time: 25:19)

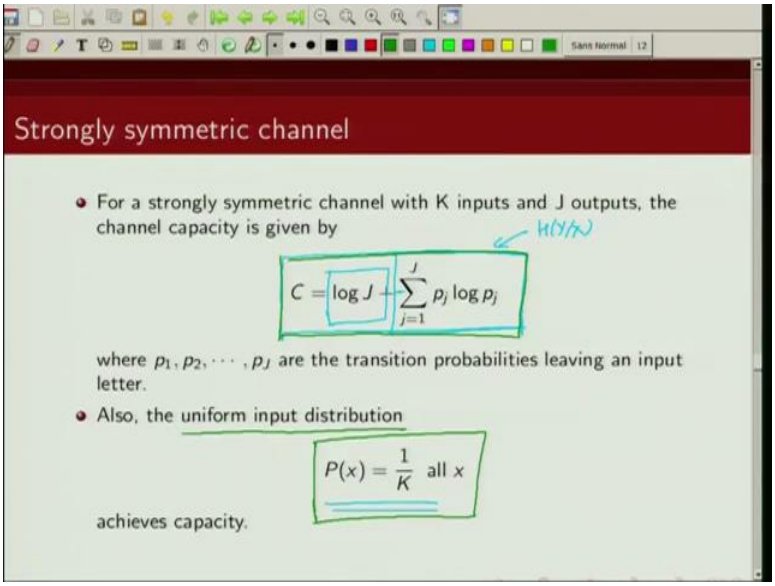


So, a channel which is both uniformly dispersive and uniformly focusing is known as strongly symmetric channel. So, a discrete memoryless channel which is both uniformly dispersive and uniformly focusing is known as strongly symmetric channel. So, I have given example let us just checked. So, for it to be uniformly dispersive we will have to look at transition leaving an input. So, what are these transitions probability this is 1 minus epsilon and epsilon similarly here transient probabilities are 1 minus epsilon and epsilon is to become a smaller number if I arrange them in descending order no matter whether I consider input 0 or 1 these probabilities are order of these probabilities are same. So, this is an example of a uniformly dispersive channel.

Now, let us look at the output 0 and one. So, what are the transition probabilities of bits terminate of the transition from the input this is 1 minus epsilon this is epsilon similarly here is 1 minus epsilon and epsilon. So, this is an example. So, these probabilities again when arrange in descending order they are same no matter what is my output. So, this is an example of a uniformly focusing channel. So, this is a uniformly dispersive channel as well as uniformly focusing channel. So, this is a strongly symmetric channel. In fact, this channel is known as what we call binary symmetric channel.

So, binary channel inputs of binary output of binary and its symmetric this is a binary symmetric channel and it is typically use to model if we use additive white Gaussian noise channel followed by hard d modulation, we can model this channel that channel by a binary symmetric channel. So, then since this channel is strongly symmetric can we find out its capacity the answer is yes what was the capacity of dispersive channel capacity of dispersive channel is this right maximizing h of y over p of x. Now for uniformly focusing channel we know what this quantity is and for a uniformly dispersive channel we know what this quantity is right since a strongly symmetric channel is both uniformly focusing and uniformly dispersive, we know exactly what is the expression of a channel capacity for a strongly symmetric channel, this will be log of j plus summation of p j log p j. So, the expression for capacity for this strongly symmetric channel is given by this expression all right.

(Refer Slide Time: 28:56)

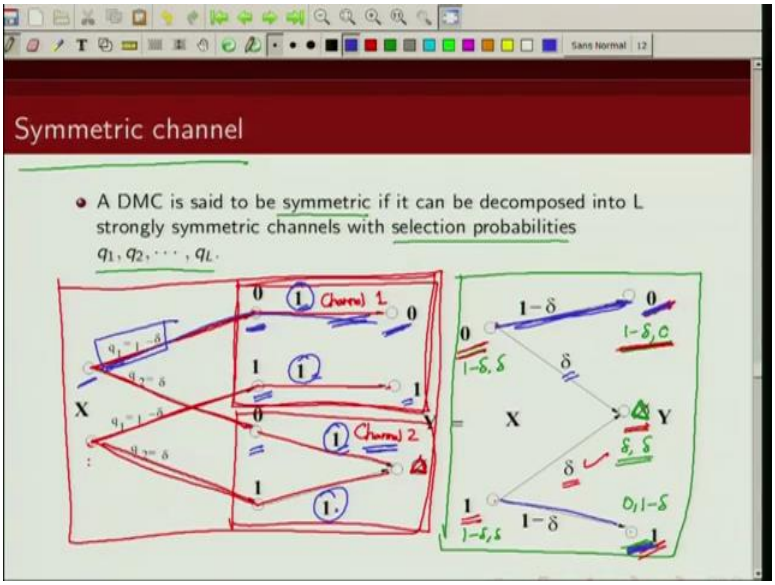


So, if you have a strongly symmetric channel with k inputs j outputs then, channel capacity is given by this expression we know from the uniformly focusing property that the distribution that maximizes h of y is uniform distribution and the maximum value of a h of y corresponding to this uniform distribution is given by log of j, where j is a number possible outcomes outputs and for a uniformly dispersive channel, we know irrespective of the input distribution h of y given x is given by minus of this quantity. So, we know exactly the expression for capacity of a strongly symmetric channel. So, then can you tell me what is the expression for capacity of this binary symmetric channel. So,

here j is two. So, log of 2 is one. So, of the capacity is 1 minus and these probabilities are 1 minus epsilon epsilon. So, this can be written as binary entropy function of epsilon. So, this is the expression for capacity for a binary symmetric channel.

And you can see for example, if the epsilon is half; that means this 50 percent probability of making a mistake. So, you can see in that case capacity is 0 which is iteratively correct because this is completely randomly flipping the bits to 0's and will be the equal probability and the another point to be noted is the input distribution that causes basically input distribution that achieves capacity is uniform distribution because we know for a uniformly dispersive channel irrespective of what my input distribution is h of y given x is given by minus of this quantity and this quantity h of y is maximize for uniform input distribution. So, the distribution that achieves capacity for a strongly symmetric channel is uniform input distribution. So, if p of x is 1 by k for all x then that will be achieve capacity for a strongly symmetric channel.

(Refer Slide Time: 31:57)



Now, we are going to talk about a class of channels which can be decomposed into strongly symmetric channel. So, this we call as symmetric channel. So, a discrete memoryless channel is said to be symmetric if it can be decomposed into l strongly symmetric channel with selection probabilities given by q1, q2, q3, ql. So, look at this particular channel look at this particular channel is this strongly symmetric channel this is not a uniformly focusing channel. You can see if I consider this output my

probabilities are 1 minus delta 0, 0 and if I consider this particular output my probabilities are delta, delta and if I consider this output my probabilities transition probabilities are 0 and 1 minus delta.

So, you can see if I consider this particular state which I call era state my probabilities are different from if I consider this particular output. So, this is not a uniformly focusing channel even though this were uniformly dispersive channel because the probabilities if you look from the input size this transition probabilities are 1 minus delta and delta and this is 1 minus delta delta this is uniformly dispersive channel, but it is not a uniformly focusing channel. So, this is not a strongly symmetric channel, but it can be decomposed into a strongly symmetric channel now look here.

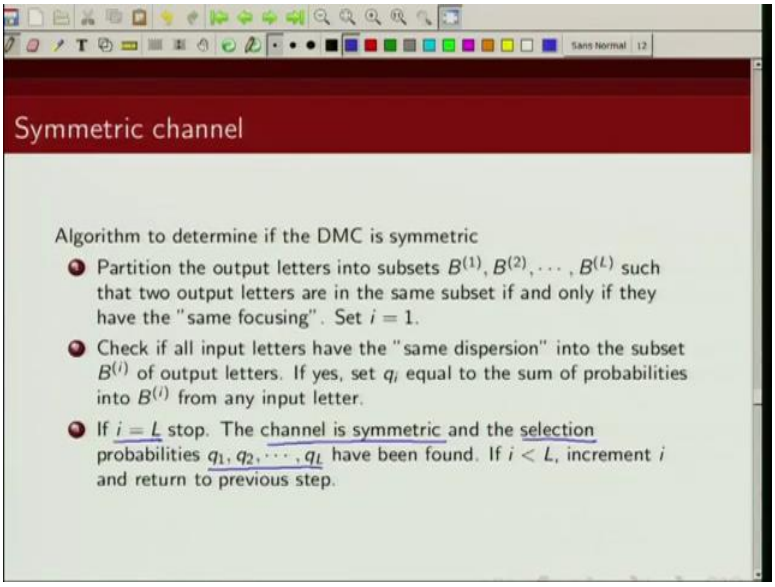
So, if you noticed these inputs 0 and 1 they have the same focusing because this probabilities, when arranged in the decreasing of the probabilities 1 minus delta 0 this is 1 minus delta 0. So, I can decompose this into 2 channel 1 is this channel corresponding to this output 0 and 1 and other 1 is corresponding to this particular input. So, I am decomposing this channel into 2 channels 1 is this particular channel and other is this particular channel and with this probability q1 I am going to select this channel let us call this channel 1 and with probability q2 I am going to select this channel 2 now know what that is these 2 channels are exactly equivalent why.

So, look at transition probability for this input to this output this is 1 minus delta this is 1 minus delta into 1 now look at transition probability from 0 to this particular output that is delta. So, this is delta into 1 similarly when you consider this particular input the transition probability to this particular output is delta. So, from 1 you can go to this state this delta into 1 this is exactly same and 1 to 1. So, 1 to 1 this is 1 minus delta into 1. So, these 2 channels are exactly equivalent another point to be noted this channel 1 and channel 2 are strongly symmetric channel you can verify that if this is uniformly dispersive because a probability in descending order is 1 and 0 for whether I consider this input or this input this is also uniformly focusing if I consider the output 0 or 1 again the probability arranged in decreasing order will be 1 and 0. So, channel 1 is strongly symmetric.

Similarly, channel 2 this is; obviously, uniformly focusing because only 1 output and this is uniformly dispersive also because probabilities are 1 and 0. When arranged in

decreasing order. So, channel 1 and channel 2 individually is strongly symmetric channel and you can think of it as we are selecting channel 1 with probability 1 minus delta and channel 2 with probability delta. Now why are we doing this why are we trying to decompose this symmetric channel into strongly symmetric channel that is because we know exactly the capacity of a strongly symmetric channel. So, if we can write a symmetric channel in terms of strongly symmetric channel perhaps we can get the capacity of a symmetric channel as well and that is where we are headed.

(Refer Slide Time: 37:58)



So, let us go over an algorithm to partition to see whether we can partition any channel in to strongly symmetric channel. So, what are the conditions that are needed to be satisfied for first to be able to partition a channel into strongly symmetric channel? So, we are describing the algorithm to determine whether, a discrete memoryless channel is symmetric channel or not enhance symmetric channel can be partitioned in to strongly symmetric channel.

So, the first step is partition the output letter into subsets b1, b2, bl such that 2 output letters are in the same subset if and only if they have the same focusing now what do I mean by same focusing. So, if you look at those output letters and if you look at the transitions from the input to those letters when they are arranged in the decreasing order it should be same. For example, if you look at this particular example you see that 0 and

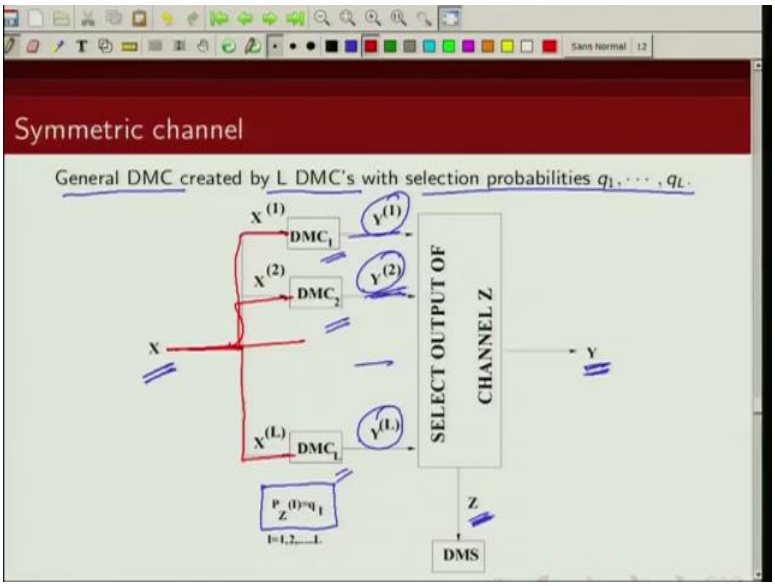
1 these outputs have the same focusing because the probabilities are 1 minus delta and 0. And that is why we club them together.

So, what we are saying is you partition your output letters into subsets b1, b2, bl such that 2 output letters are in the same subset if and only if they have the same focusing now next thing we going to do is we need to check if all input letters have the same dispersion into the subset b I of output letters. So, we need to check. So, once we partitioned the output in to l sets where within each sets they have the same focusing now we need to see whether all the inputs have same dispersion in to that subsets and if that is the case then we set qi to be equal to sum of probabilities of these the transition probability into that particular set.

For example, if you look go back and see this example you see here we are reaching here probability 1 minus delta 1 minus delta and that is. So, the dispersion is common. So, we whether we are from 0 to 1 we are selecting this channel probability 1 minus delta whereas, this particular channel we are selecting with probability delta. So, if all input letters have the same dispersion into that particular subset of output letters then, we should be able to partition our channel into strongly symmetric channel and we are going to select each of these channels with probability qi which is equal to sum of probabilities from the input to that particular subset bi and once we know these selection probabilities q is then rest this probabilities, we can easily find out because we know the transition probability for example, from this input to this output should be this into this which is this quantity. So, we can find out this transition probability this transition probability this transition probability. So, we can find out all these transition probability by mapping the transition from each input of this channel to this channel.

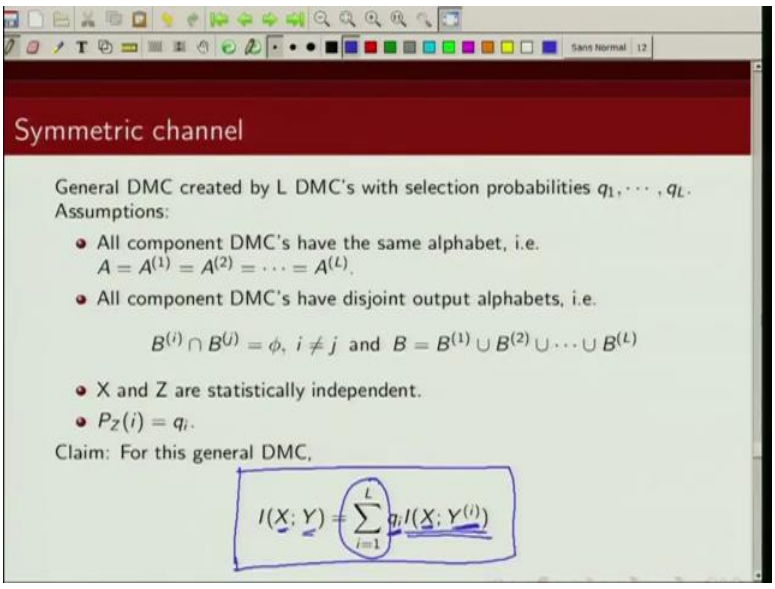
And this we continue until we are able to do it for all output letters. So, if I equal to l we will stop and will say the channel symmetric with selection probabilities given by q1, q2, q3, ql and if I is not l will continue until we partition all the output letters into these subsets.

(Refer Slide Time: 42:55)



So, this is a block diagram of how things will look like. So, you can think of our general discrete memoryless channel created by l discrete memoryless channel. Where with selection probabilities of q1, q2, q3, I have my input x and you can think of a l such discrete memoryless channel I am choosing each of these channels with probability q1, q2, q3, ql. So, the output you can think of is I am choosing either this channel, this channel, this channel or this is my selection probability. So, x is my input here y is the output of discrete memoryless channel and z is deciding which channel is being chosen.

(Refer Slide Time: 43:47)

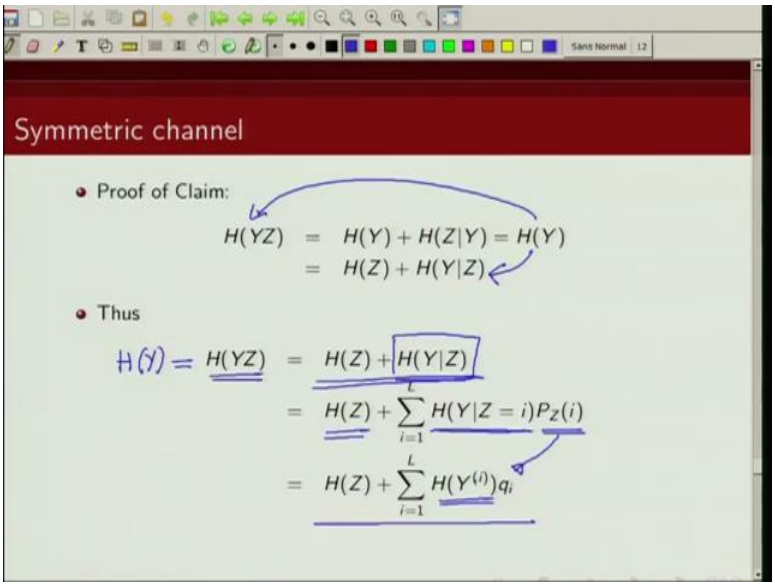


So, we are considering a general discrete memoryless channel which is created by this l discrete memoryless channel with selection probability is given by q1, q2, q3; q l and we make these following assumptions.

All of this, one discrete memoryless channel they take as a input they take as inputs same alphabets. So, input alphabet for channel 1, channel 2, and channel 3. Channel l they are all same second assumptions all discrete memoryless channels have disjoint output alphabets. So, there is when we partition them into l discrete memoryless channel the output letters are disjointed. So, there is some output letter appearing in 1 particular set it would not appear in any other set. So, the intersection b I and b j is 0 and of course, the union of all these l disjoint set should be my overall output letters and this selection is which is denoted by this a random variable z this is independent of x and the selection probability is given by q i.

Now the claim I am making it is mutual information between the input x and output y can be written in terms of mutual information of these individual channels and this is the expression relating the mutual information between the input of this channel and the output of the channel. So, from this expression basically as you can see we can write the mutual information between input x and output y as mutual information between input x and output of this individual l channels multiplied by this selection probability and sum over all such l channels. So, we are now going to prove this result.

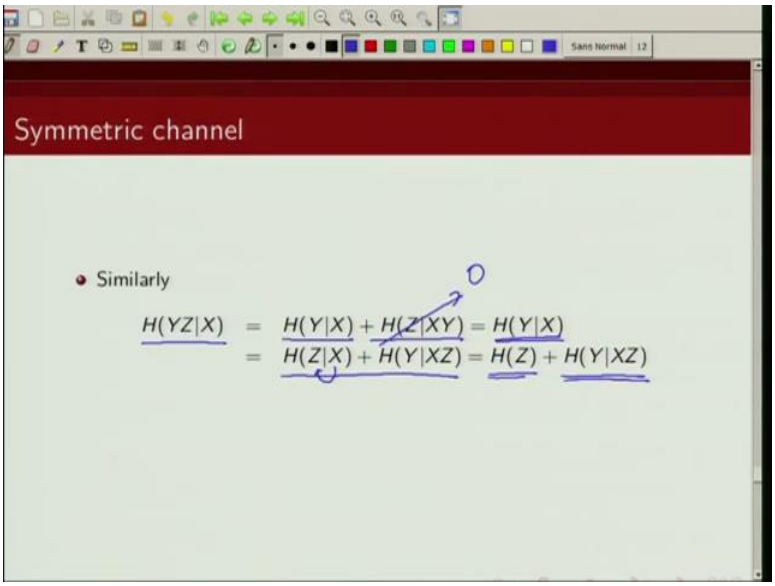
(Refer Slide Time: 46:21)



So, joint entropy of y and z using chain rule can be written like this uncertainty in y plus uncertainty in z given y. Now if I tell you what my y? If you will go back to the diagram here y is basically v1, y1, y2, y3, yl depending upon which channel is selected. So, if I tell you what my y is you know precisely what channel has been selected. So, z is known. So, there is no uncertainty in z given y. So, this quantity is 0 because if you tell me what y is I know because my output symbols letters has been partition into disjoint sets if you tell me what y is I know exactly which channel has been selected. So, there is no uncertainty in z given y hence this can be written as uncertainty in y now this is same as uncertainty in z plus uncertainty in y given z again I am applying chain rule in a different way. So, from this 2 I can write h of y is equal to h of y z this is from here I get and this is same as this quantity.

Now, let us expand this particular term. So, I can write this as h of z plus uncertainty in y given z is I multiplied by probability of z. So, this is from the definition of conditional entropy now this term is given by q I and this uncertainty. So, given that I have selected I h channel I know the output is yi. So, I can write uncertainty in y given z is I as uncertainty in yi. So, then I can write this expression as h of z plus summation of h of yi into qi and sum over all l such channels.

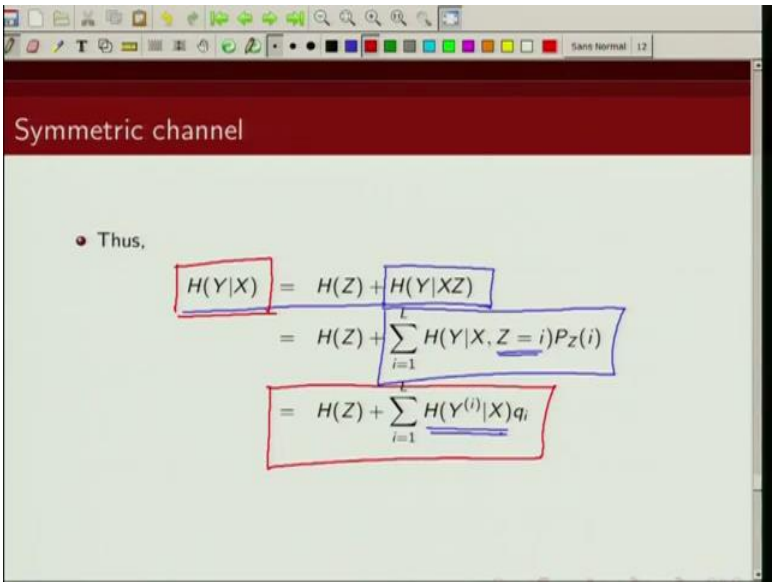
(Refer Slide Time: 49:08)



Similarly, I can write uncertainty in y z given x as using chain rule uncertainty in y given x plus uncertainty in z given x and y and as I said given y there is no uncertainty in z. So,

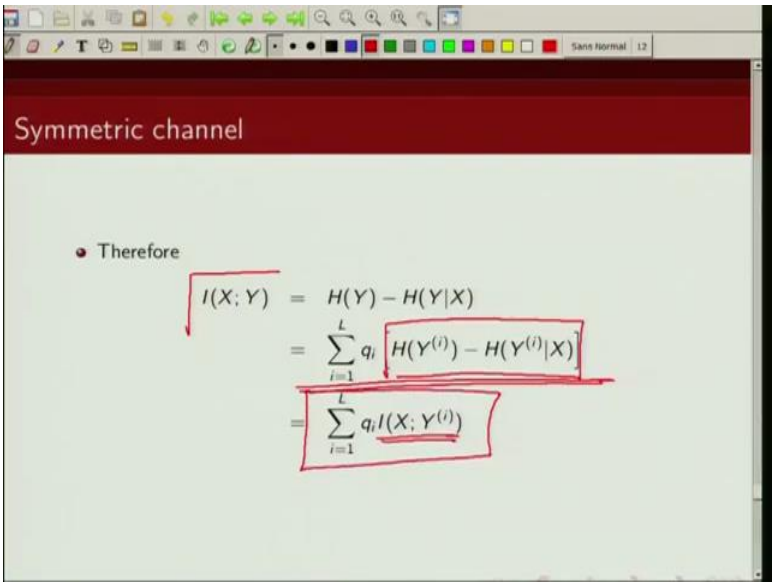
this term is 0. So, this can be written as uncertainty in y given x now I can apply chain rule again in a different way and I can write this expression as uncertainty in z given x plus uncertainty in y given x z. Now x and z independent, so knowing x does not give me information about z. So, this will be h of z plus h of y given x z.

(Refer Slide Time: 50:04)



Next; from the previous expression then I can write h of y given x as h of z plus h of y given x z now I can expand this. So, I can write this in this particular fashion. So, this is uncertainty in y given x and given z is I multiplied by probability of z i. Now if I tell you what z is. So, you know exactly what channel is. So, this can be written as uncertainty in yi given x multiplied by qi. So, I have gotten in expression of this which is given by this expression and earlier I got the expression of h of y which was given by this expression right. So, mutual information is maximizing x of y minus h of y given x and maximizing over all input distribution. So, if I subtract h of y given x from this h of z is common in both of them. So, that gets cancelled out.

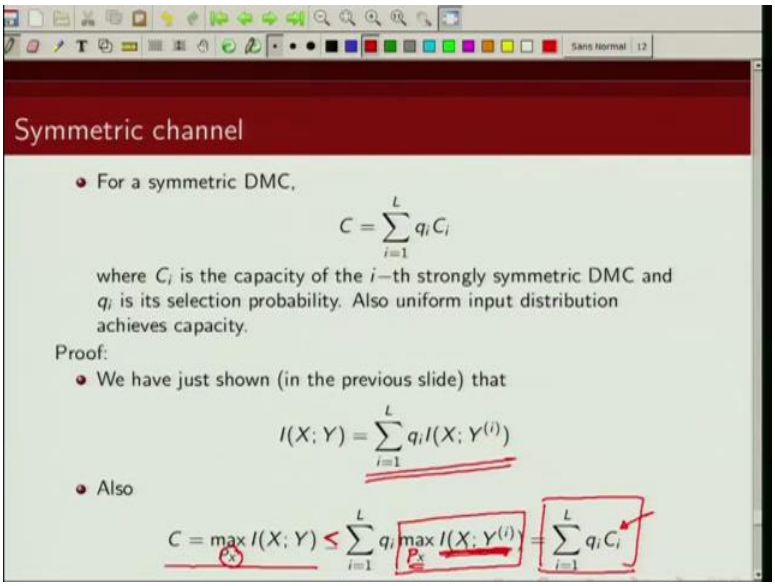
(Refer Slide Time: 51:34)



What is left is this particular term, now what is this h of y I minus h of y I given x this is mutual information between x and y i. So, then I have proved that when I decompose my channel in to these l discrete memoryless channel following the assumptions, which I made earlier I can write the mutual information between the input and output in terms of mutual information of the individual discrete memoryless channel all.

And we are going to exploit this to get the capacity of a symmetric channel. So, if we are able to decompose our symmetric channel into strongly symmetric channel and since it is a same input which goes to all of these strongly symmetric channel and we know that it is the uniform input distribution that maximizes that achieves capacity. So, we should be able to get the expression for capacity of a symmetric channel.

(Refer Slide Time: 53:00)



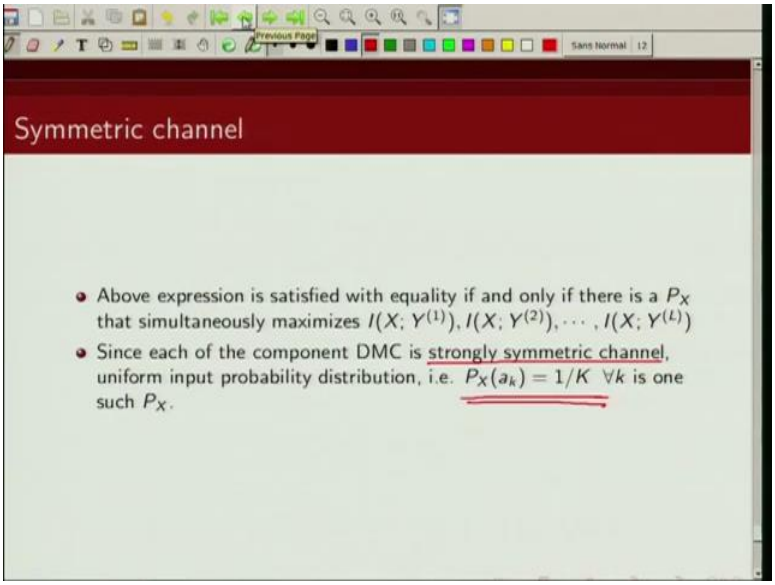
If we are able to decompose it in to that is after decomposing it in to strongly symmetric channel. So, for a symmetric discrete memoryless channel the capacity is given by this expression which is product of capacity of these individual discrete memoryless channel multiplied by their selection probability and sum over all l such discrete memoryless channel and it is the uniform input distribution that will achieve capacity the proof is very straight forward once we have established that the mutual information between x and y can be written as summation of mutual information of this individual discrete memoryless channel multiplied by this selection probability.

Now, capacity as you know can be given by mutual information between x and y and maximizing over all input distribution now this is less than equal to. So, I need to find basically p x which will maximize this. Can I write it less than equal to this now remember in case of strongly symmetric channel it is the uniform distribution input distribution that achieves capacity and since I have the same input going to all of these l channels. So, the input distribution is same. So, the distribution that achieves capacity for ith channel the same distribution achieves capacity for that is the jth channel. So, that is why for a strongly symmetric channel I can write capacity to be equal to product of capacity of the individual discrete memoryless channel multiplied by the probability of selecting that particular channel.

So, again I repeat the input distribution that maximizes this for an ith channel is same whether, I consider ith channel or jth channel because it is the uniform input distribution that will achieve capacity and since my input. If you go back to the diagram that I showed you it is the same input which is going to it is the same input x is going here is going here it is going to all these discrete memoryless channel the same input is going and this and since uniform distribution will achieve capacity for each of this individual channel and it is the same input. So, it is a same input that will achieve capacity for this channel.

So, hence we have proved that capacity of a strongly symmetric channel is given by this expression.

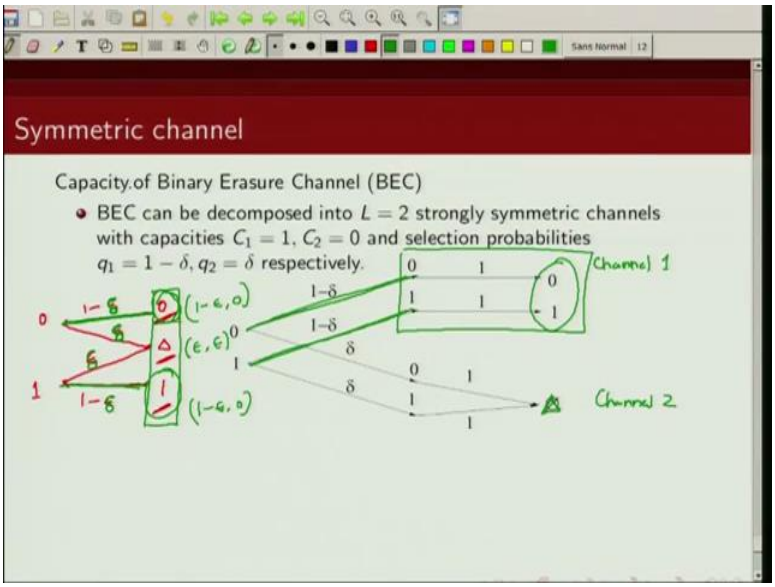
(Refer Slide Time: 56:24)



So, the above expression as I said is satisfied with equality if and only if there exist an input distribution that simultaneously maximizes the mutual information of each of this individual channel and we know it is a uniform distribution that will that will maximize this why because each of these individual discrete memoryless channel are strongly symmetric channel. So, this the uniform input distribution that will simultaneously maximizes the mutual information of each of these discrete memoryless channel. Hence the capacity is given by summation of product of the selection probability multiplied by the probability of the individual capacity and sum over all such l discrete memoryless channel.

So, we will illustrate how we can decompose a symmetric channel into strongly symmetric channel and then, we will use the expression of strongly symmetric channel to compute the capacity of a symmetric channel.

(Refer Slide Time: 57:49)



So, taken example of a binary erasure channel, in binary erasure channel you have 2 inputs 0 and 1 and you have 3 outputs 0 1 and another output state which is era state with probability 1 minus epsilon you receive the bits correctly and with probability epsilon you receive the bits incorrectly now, packet data networks can be modeled by this very simple binary erasure model you either receive the packets correctly or if you do not receive the packet correctly then, you want to discard them.

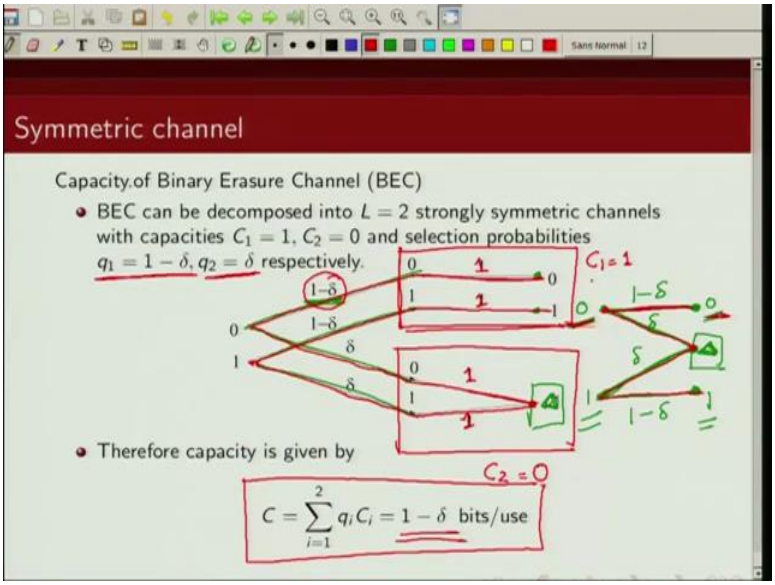
Now this binary erasure channel can be as you can see here if you look at the focusing of this particular input 0 and 1 they are same and focusing of this delta is same. So, first thing that we will do is we will decompose it in such a way that this particular output symbol and this particular output symbol they belong to the same channel like here because, they have the same focusing you can check this is 1 minus epsilon 0 this is 1 minus epsilon 0 and this has a different focusing which is epsilon. So, this will be a different channel.

So, this is my 1 strongly symmetric channel 1 and this is my channel 2, now what is this selection probability this is. So, from the inputs 0 and 1 I reach this set consisting of output symbol 0 and 1 with probability 1 minus epsilon. So, let us make it delta because I

am in the text I am following 1 minus delta. So, let us just change this notation to 1 minus delta. The same thing just changing notation because I just notice that I using this notation delta, I am just changing this notation this I am writing as 1 minus delta and this as delta.

So, then I can select this channel you can see I can select this channel with probability 1 minus delta. So, this is this selection probability is 1 minus delta this selection probability is 1 minus delta. If you recall and go back to the yes, the first step is we are partitioning them into subset b1, b2, bl such that no 2 letters are in the same subset to such that 2 output letters are in the same subset. If they have the same focusing see in our example 0 and 1 output letters they have the same focusing and delta has a different focusing. So, that is kept in a different subset. So, b1 consist of output symbols 0 and 1 and b2 consist of delta the next is check if all inputs have the same dispersion then, how do you check the same dispersion from the particular output symbol you see the probability of the transitions from the input side. Now if you look here in this particular example.

(Refer Slide Time: 61:40)



If you look at this particular example you reach 0 from 0 with probability 1 minus delta. Similarly you reach 1 from 1 the probability 1 minus delta and from 1 you cannot reach 0 and from 0 you cannot reach one. So, they have the same focusing and that is given by 1 minus delta.

So, the selection probability here in this case is given by is 1 minus delta as you consider the second channel here the second channel is similarly selected with probability delta again lets draw redraw the diagram for binary erasure channel 1 minus delta, delta, delta 1 minus delta. Now look at how you are selecting this channel which consist of this output symbol delta from input 1 this is probabilities delta from input 0 this probabilities again delta. So, this particular channel is selected with same probability. So, whether you consider this input or this input both have the same dispersion and similarly for this particular channel which consist of output symbol 0 and 1. You can see I can reach 0 with probability 1 minus delta and from here 0 probabilities similarly I can reach 1 with 0 probability here and 1 minus delta. So, they have the same dispersion; that means, I can find out what is selection probability. So, I will select this channel consisting of 0 and 1 with probability 1 minus delta and I will select this channel which has this era state with probability delta.

Now, what I need to do is I need to look at this probability from input to the output in this case this is 1 minus delta now if you look here. So, this probability should be 1 minus delta I already know my selection probability is 1 minus delta. So, then this probability should be 1 similarly from 0 to delta this is probability this era state this probability delta. So, 0 to this era state this is probability delta. So, this has to be 1. So, that the overall probability is delta and likewise I can calculate from 1 to the lets see it again this probability delta. So, this delta into this should be delta. So, this probability is 1, now similarly from 1 to this probability 1 minus delta. So, since the selection probability is 1 minus delta this has to be probability 1.

So, now my selection probability to you first set of channels q1 is 1 minus delta and selection probability for the second set of channel is delta. Now what is the capacity of this channel this is the strongly symmetric channel and this is 1 minus this capacity of this channel is 1 we can easily see its capacity is 1 and what is the capacity of this particular channel this capacity of this particular channel is 0, we already know the expression for capacity of strongly symmetric channel. So, just plug in the value j in the case is 2. So, log of 2 is 1 and the transition probabilities are 1 and 0. So, the contribution due to h of y given x will be 0. So, capacity of this will be 1 and similarly capacity here j is one. So, log of 1 is 0 and these probabilities are one. So, overall capacity of this particular strongly symmetric channel is 0. So, the overall probability will be selection

probability of this channel which is 1 minus delta times 1 plus delta into 0. So, the capacity of a binary erasure channel is 1 minus delta 0. So, capacity is basically given by this.

So, this example illustrates. Note that this binary erasure channel is not a strongly symmetric channel, but it is the symmetric channel because we are able to decompose this channel into 2 strongly symmetric channels and since from the inputs we are able to reach these channels with same dispersion. We can find out what is the selection probability of these channels and subsequently we use this expression for capacity of a symmetric channel in terms of capacity of strongly symmetric channel and the selection probability to get the expression for capacity of binary erasure channel which is 1 minus 1 minus delta. If delta is 1 capacity is 0 if delta is 0, we can see that capacity will be 1.

So, with this we conclude our lecture on channel capacity in the next lecture we will talk about channels noisy channel coding theorem, we will prove channels noisy channel coding theorem and then we will prove the converts of noisy channel coding theorem.

Thank you.