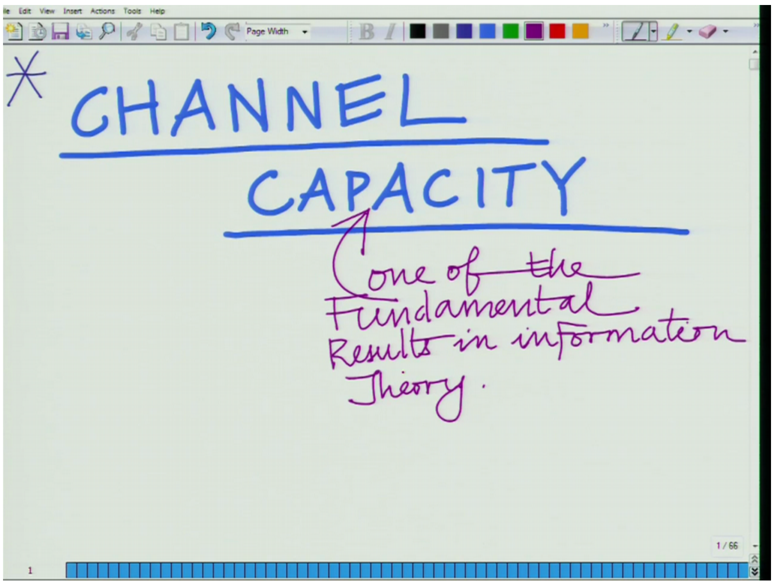
Principles of Communication Systems - Part II
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 35
Channel Capacity, Implications of Channel Capacity, Claude E. Shannon- Father of Information Theory, Example of Capacity of Binary Symmetric Channel

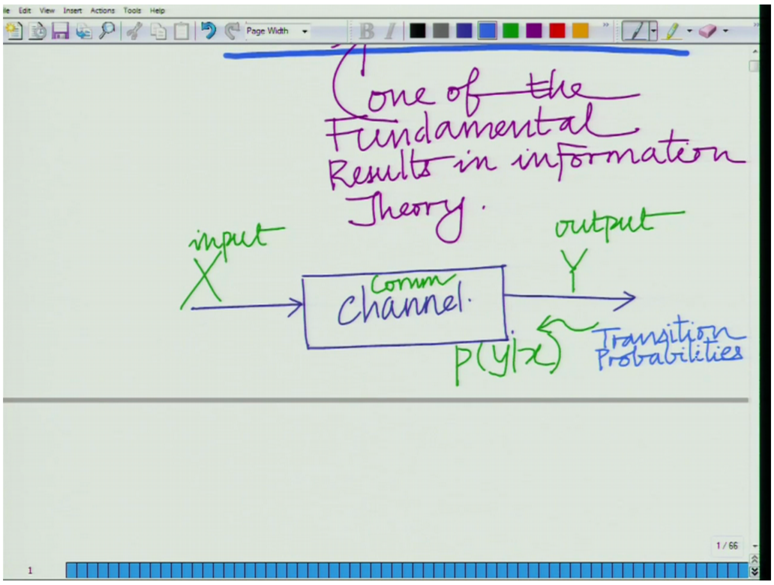
Welcome to another module in this massive open online course, alright. So, we are looking at the mutual information. We have looked at the properties of the mutual information, the definition the properties and the relevance of mutual information, alright and let us now look at one of the most important and landmark results in information theory which relates to channel capacity and which depends on the mutual information.

(Refer Slide Time: 00:36)



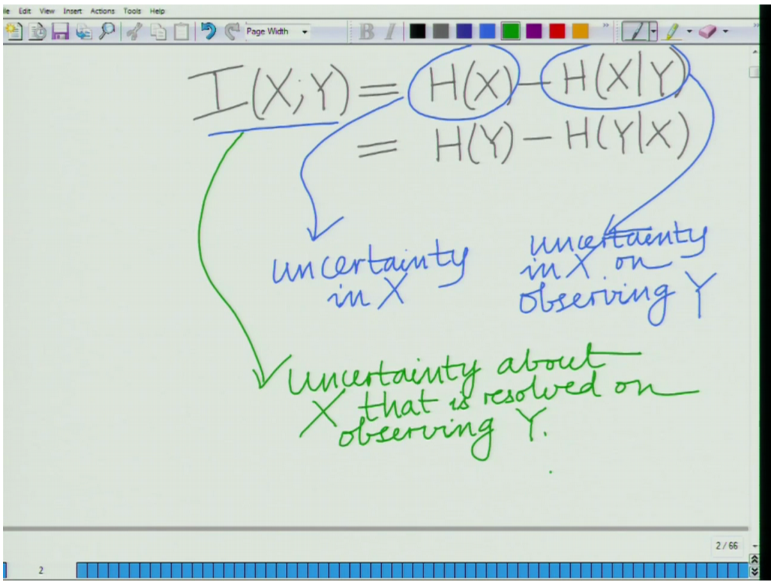
So, what I want to look at today is one of the fundamental results mission theory. As I already told you this is one of the fundamental, I can also say the landmark results, one of the fundamental results in information theory which is concerned itself, which can be described as follows that is if I have a channel and this can be any.

(Refer Slide Time: 01:28)



I have X which is the input stream and Y which is the output. This is your communication channel and described by the transition probabilities P y given x that is the probability that the output takes a certain value corresponding to a given input symbol.

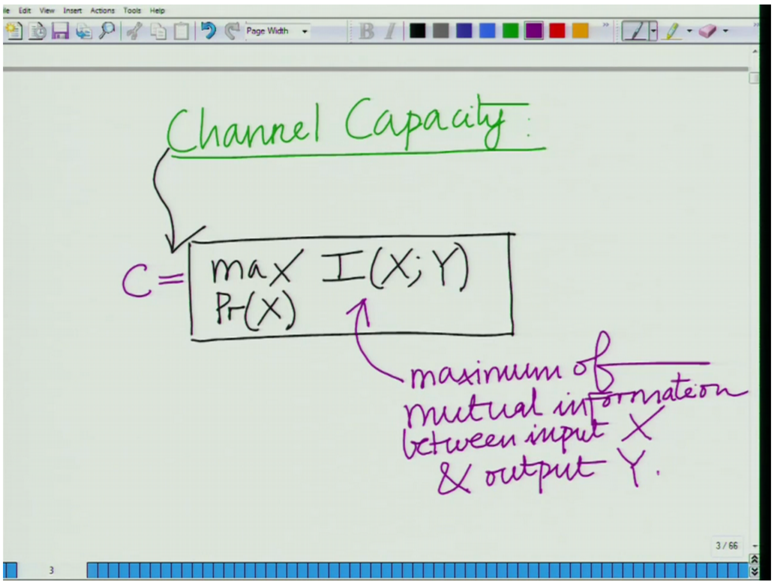
(Refer Slide Time: 02:27)



So, the channel is described by a set of transition probabilities correct and we already have seen what is the mutual information. The mutual information remember we said this as a fundamental role to play in defining the rate at which information can be transmitted across the channel. This is H X minus H X given Y also equal to H Y minus H Y given X. So, if you look at this, this is basically the uncertainty in X, uncertainty if you look at H X. Let me just repeat that logic again. If you just look at simply H X, this is the uncertainty in X. H X is simply the entropy or the uncertainty in X. That is the input symbol stream X, the input X; H X given Y is uncertainty remaining in X.

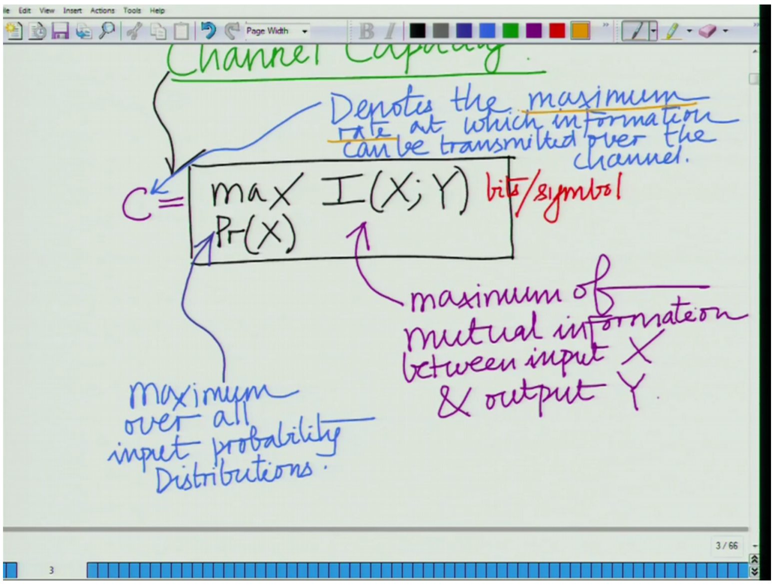
So, the entropy in X conditioned on Y that is on observing Y. So, this H X given Y uncertainty in X on uncertainty in X on observing Y, which means we look at this i of x y which is H X minus H X given Y. This is uncertainty in X minus uncertainty remaining in X. So, this is basically uncertainty resolved uncertainty about X about X, that is resolved or uncertainty about X that is resolved on observing y and we said that for a good channel. This should be high for a good channel. This quantity should be high that is we would like to make this mutual information between X and Y, mutual information of X and Y is nothing, but the information that is uncertainty in X that is resolved by observing or uncertainty of about Y that is resolved on observing X. That is you can also say that is how much information is basically conveyed or regarding X by Y, how much information about Y is conveyed regarding X, ok.

(Refer Slide Time: 05:43)



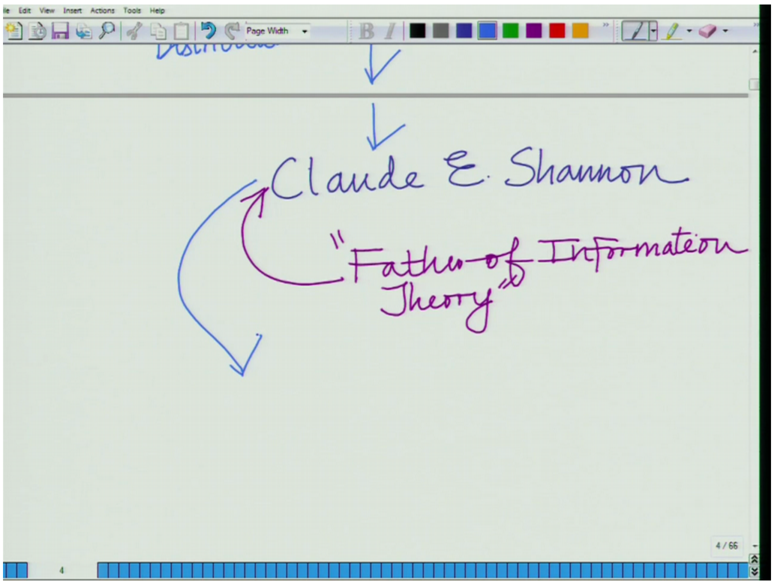
Therefore, we want this quantity to be high and in fact, the fundamental result for the capacity of the channel, the fundamental result for the channel capacity, this states that the maximum rate of information transfer across the channel is given by the maximum of the mutual information, where the maximum is taken over all the input probability. In fact, this is one of the most fundamental, the channel capacity is the maximum of the mutual information. You can call this as C equals maximum of the mutual information between input and output between input X and output Y and this maximum, remember this maximum is taken over all maximum input that is probability.

(Refer Slide Time: 07:10)



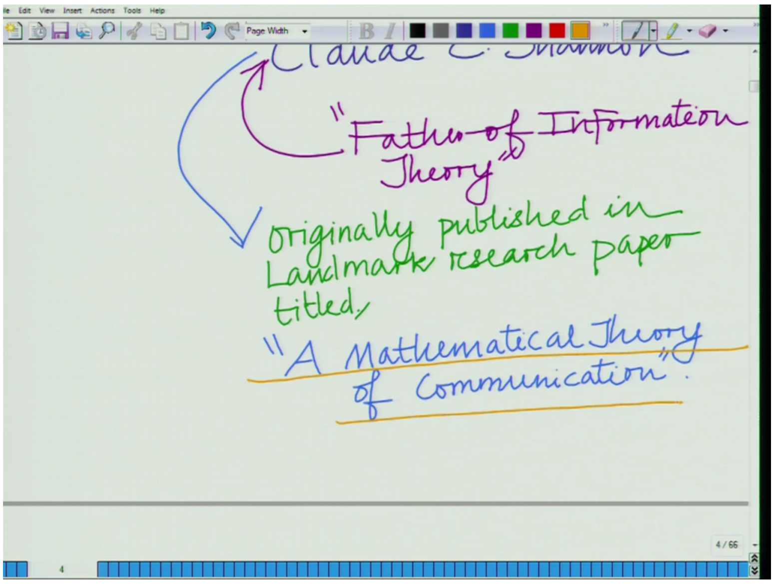
That is what is a input probability distribution for which the mutual information is maximum. That is the capacity of the channel. The units are bits per channel use or bits per symbol, bits per symbol transmitted and it is important to understand what this channel capacity denotes. This channel capacity denotes or the C, the channel capacity C denotes the maximum rate at which information can be transmitted. This is the maximum rate information can be transmitted, maximum rate at which information can be transmitted over the channel. This is in bits per channel. Use that maximum number of bits per each symbol, that is the maximum number of bits that can be packed over each symbol that when we transmitted over a channel, needless to say this is a landmark result because this characterizes in a fundamental way. What is the maximum? That is information rate. What is the maximum bits per second? For instance, we have several communication system. So, 3G, 4G, 5G, right once would like to know what is the maximum rate at which information can be transmitted over the channel in each of these communication system. This framework which is the channel capacity gives us a very central result, a landmark result to characterize the maximum rate.

(Refer Slide Time: 10:11)



Most of you must have heard this result is due to none other than well the very celebrated figure in information theory. So, this result is due to the well known sign test Claude E Shannon. So, this is the Shannon's result on the channel capacity and Shannon is also must be known to many of you is known as the Father of Information Theory. He was a person who developed, originally developed this framework of information theory which right now has grown to include and do not have applications in so many diverse areas. So, one of them is of course in wireless communications, but it can be applied in many other. So, he is known as the father of information theory and this result was published.

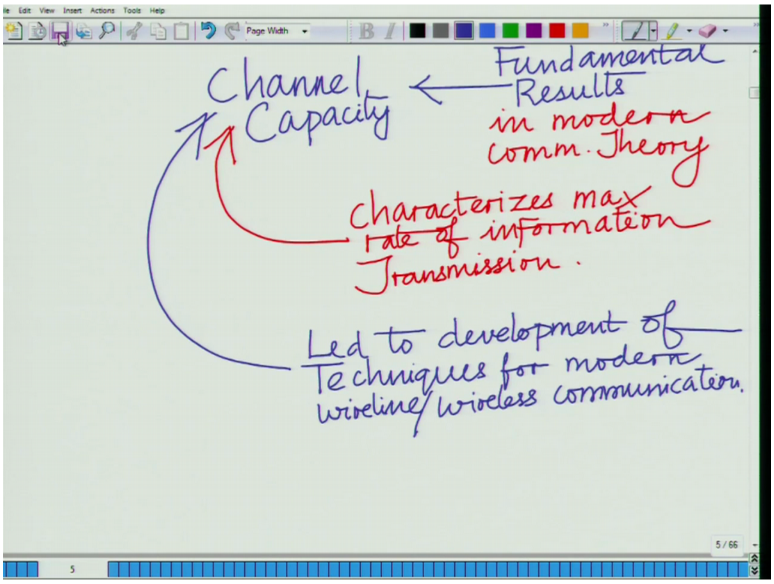
(Refer Slide Time: 11:21)



The channel capacitor result was published in the Landmark article title originally given or originally published. This result was originally published in the Landmark article or basically let see the landmark paper, landmark research paper titled and the title is The Mathematical Theory which was later changed to The Mathematical Theory of Communication, ok.

So, it was originally published in this landmark research paper and lenus to say the channel capacity is a central result you know all of one of the central results in all of communication theory because it helps characterizes the very fundamental way, what is the maximum rate at which information can be transmitted in bits, per bits, per channel use and naturally we are going to see later that that can also be extended to per second across the channel.

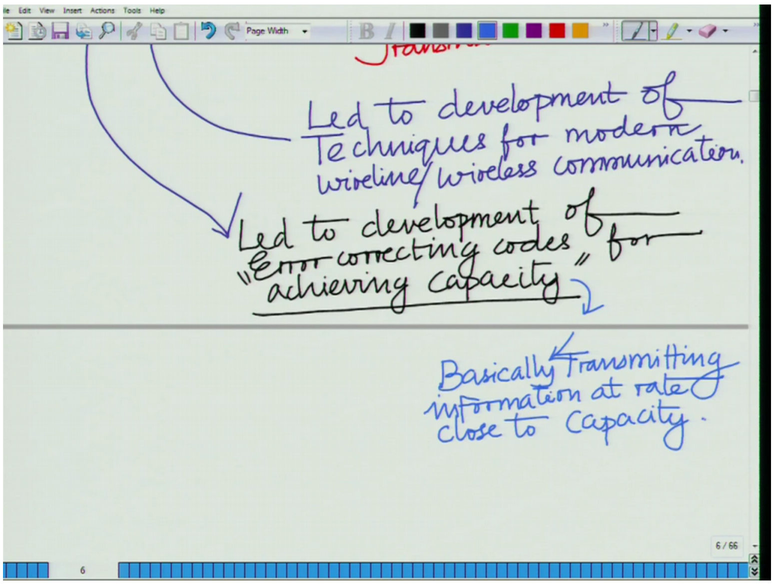
(Refer Slide Time: 13:13)



So, it is a fundamental result. So, this channel capacity I cannot over emphasize. I cannot emphasize this enough. This is a fundamental as far as fundamental results, there cannot be many more, anymore results which is more fundamental than this. So, this is the channel capacity. So, in fact one of the fundamental results in modern communication theory, alright characterizes the maximum rate of information transmission correct. It has led to the development of several modern communication systems or several model techniques, does led to development of techniques for modern wireline, both wirelines slash wireless communication and more importantly this is also, all the channels result if you look at it, it says that they exist that is it is possible to transmit information at this rate over across the channels, but this is not explicit about what is the technique that can be used to transmit information at this rate over the channel.

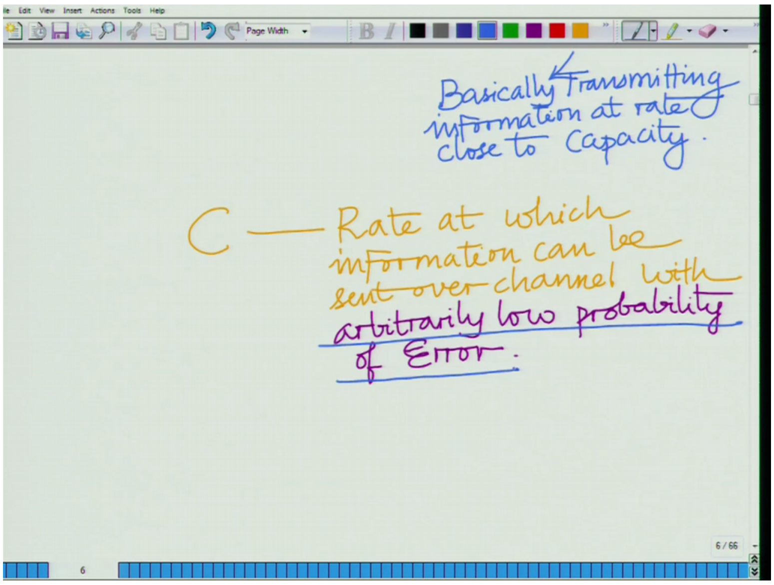
So, that calls for the design of techniques which make possible robust codes which make it possible to transmit information at this rate given by the maximum mutual information of the channel capacity across the channel. So, that has to lead to the development of several efficient techniques or error correcting codes to enable one to transmit push closer and closer to the channel capacity or transmit information at this maximum rate, ok.

(Refer Slide Time: 16:04)



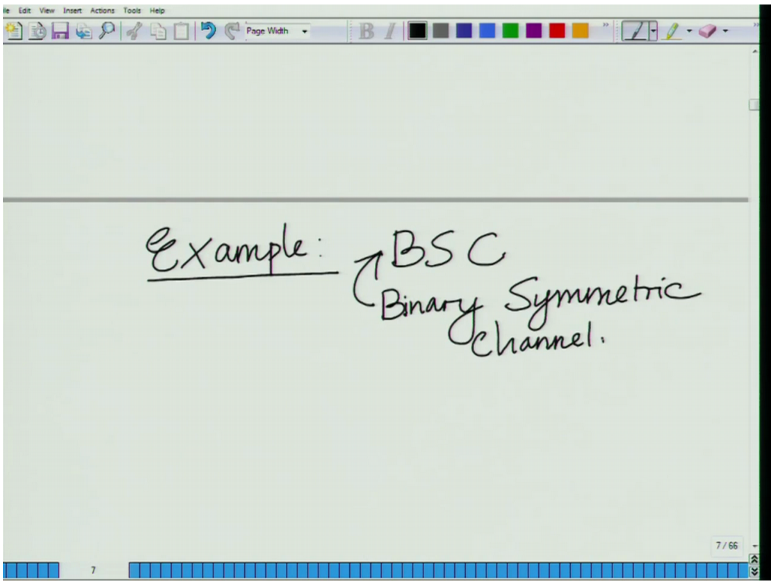
So, this channel capacity has also led to because performance of every error correcting code can bench mark against that of the channel capacity. It has led for achieving capacity. When we say achieving capacity, we mean that basically transmission or basically transmitting information at rate and another interesting aspect of the channel capacity is also that at this rate C, it guarantees that if you look at this channel capacity C guarantees information can be said at an arbitrarily low probability of it.

(Refer Slide Time: 17:32)



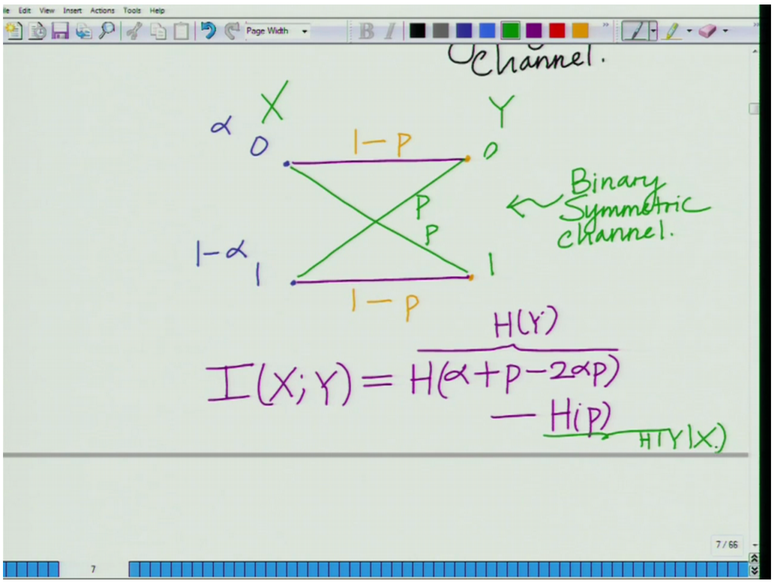
This is the rate at which information can be sent over channel with arbitrarily low probability of error and this is the keyword arbitrarily low probability of error implying that I can keep devising better and better schemes as the probability of error is driven close to 0, not exactly 0, but it can be driven as close to 0 as one wishes by devising better and better schemes in particular considering larger and larger block lengths or which the symbols can be encoded. So, that is one of the keys legasis of share this information theory and Shannon's paper as a mathematical theory of communication which comes up with the framework of information theory and a fundamental characterizes the maximum rate at which information can be transmitted over any given channel which is also known as a discrete memory list channel. Although we have not gone into the details of that, right and it is basically given by the maximum of the mutual information, where the maximum is respect to all the possible inputs source distributions, that is the input symbol distributions, input alphabet probability distributions, ok.

(Refer Slide Time: 19:33)



To just take an example for channel capacity, let us go back to our Binary Symmetric Channel and quickly look at what we had yesterday.

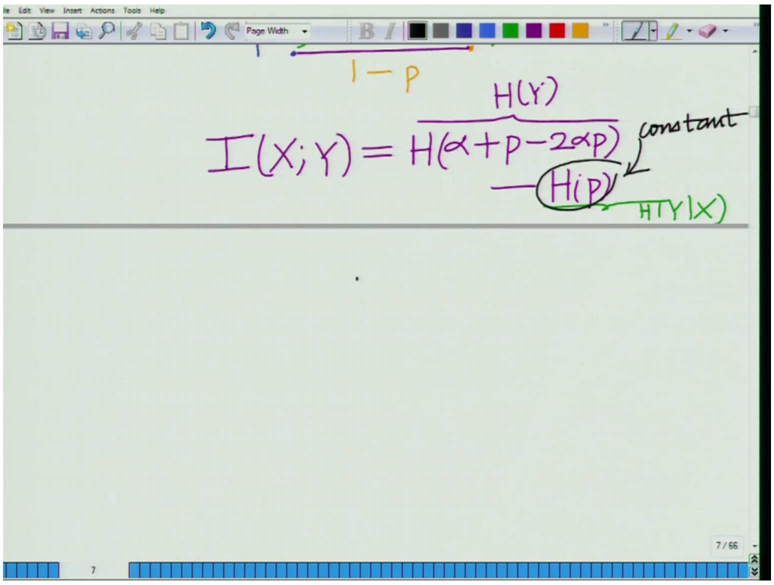
(Refer Slide Time: 20:03)



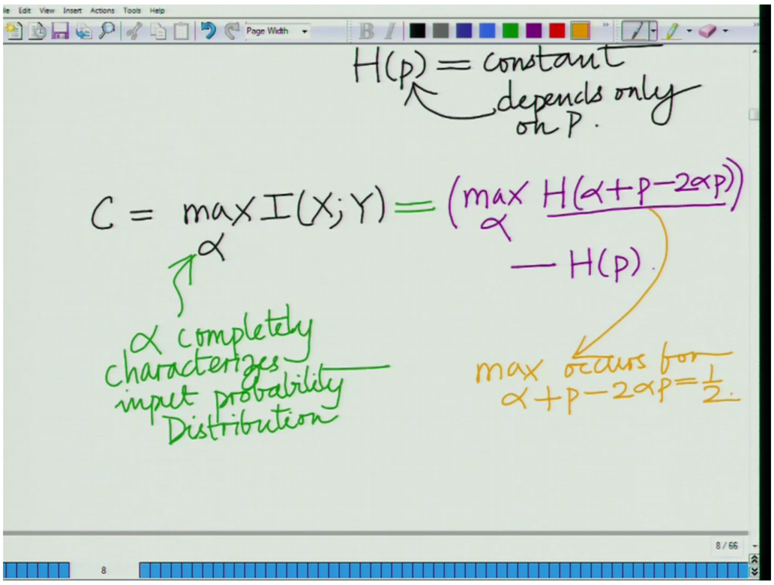
Let go back to our BSC that is this is your Binary Symmetric Channel, where you can have 0 and 1 with probability alpha input symbol 0 and 1 and correspondingly, you can have output symbol 0 and 1. The direct probabilities are 1 minus p. The cross over or the flip probability are basically P and P you can receive this as 0, you can receive this as 1, this is x, this is y. This is symmetric because the flip probability is for 0 and 1 are the same direct probabilities for 0 and 1 are the same. So, this is your binary symmetric channel model and we have seen the mutual information, right. In the previous module, we have calculated the mutual information of this channel, mutual information between the input and output for the given probability, the flip probability P and the input probabilities alpha 1 minus alpha and we have seen that the mutual information equals h of alpha plus P minus 2 alpha p. This is the h of alpha plus P minus 2 alpha p. Remember this is your h of y minus h of p, this is your h of y given x.

Now, if you look at this h of p, realize that this is a constant for a channel h of P equals constant because for a given channel.

(Refer Slide Time: 22:02)



(Refer Slide Time: 22:09)

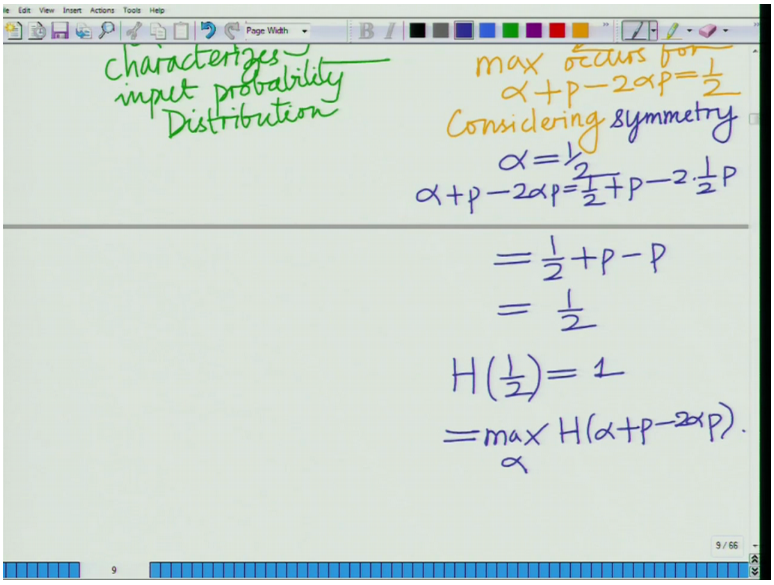


That is probabilities P and this depends only on P. So, when we are trying to maximize the mutual information with respect to the source probabilities and remember the source probabilities are characterizes by alpha, so we can write this capacity as maximizing I X Y mutual information with respect to alpha because it completely characterizes the source probability. The source given alpha I can derive the source probabilities are alpha and 1 minus alpha.

So, maximizing the mutual information with respect to the source or the input probability distribution is same as maximizing it with respect to alpha correct because alpha is a parameter which completely characterizes the input probability distribution. So, maximize because alpha completely this is equal to therefore, max alpha H alpha plus P minus twice alpha P minus H of p, H of P is a constant. So, I can just look at maximization of H alpha plus P minus 2 alpha p.

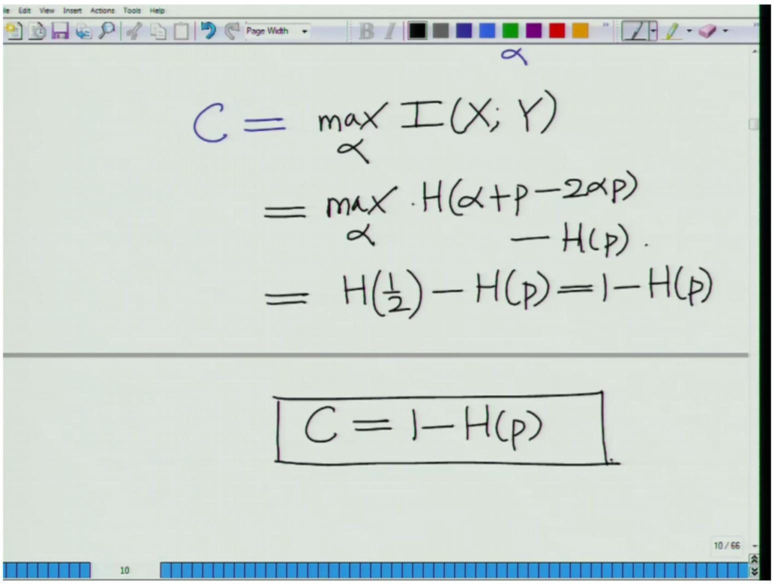
Now, we know that this maximum occurs for alpha plus P minus 2 alpha P equals half.

(Refer Slide Time: 24:19)



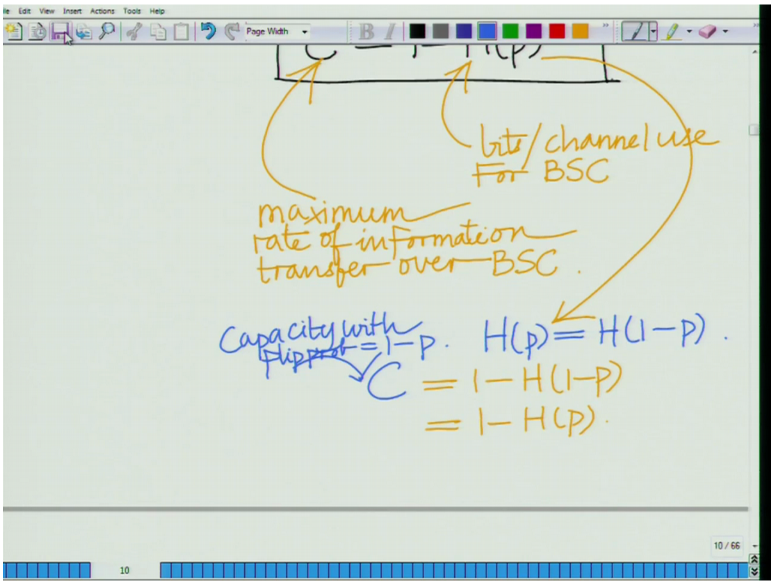
Considering the symmetry, if we set remember everything is symmetric, so it should be obvious what is the value for its maximized. Considering symmetry if we said alpha equals half, then alpha plus P minus 2 alpha P equals half plus P minus 2 into half into P which is half plus P minus P which is equal to half. So, we obtain H of half, this is equal to 1 which is the maximum value of which is indeed maximum value of H of alpha plus P minus 2 alpha because remember we have said that the entropy of a binary source that is H of 0, it is 0. H of 1 is 0, at 0 is 0, at 1 it is 0 and achieve its maximum at corresponding to when the binary source alphabet like each of the alphabet in the binary source alphabet as probability half inch, alright and we have shown by choosing alpha equal to half, we are able to achieve right H of half. So, we are able to achieve H of half which is basically equal to 1 which is the maximum.

(Refer Slide Time: 25:55)



So, therefore, now we get C which is equal to maximum over alpha, the mutual information is equal to maximum over alpha H of alpha plus P minus 2 alpha P minus H of P and the maximum of this is basically occurs when alpha is equal to half which is equal to H of half minus H of p, but H of half is 1. So, this is 1 minus H of p.

(Refer Slide Time: 26:51)

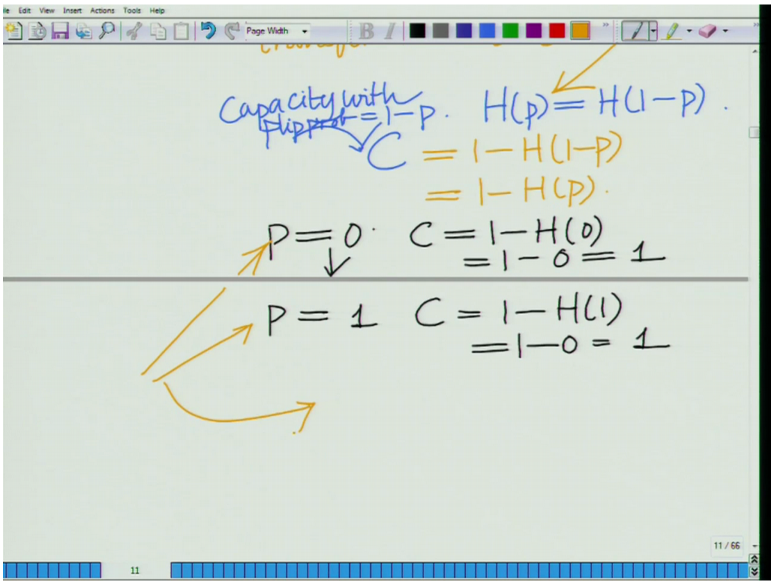


So, C equals 1 minus H of P bits per channel, use C equals H of 1 minus H of P bits channel use for binary symmetric channel, ok. So, this is C is the maximum rate of information transfer. Again needless to say we have said this several times, this is the maximum rate you have not formally proved this because that is a very involve proof. So, we are taking this as given this is the maximum rate of that is the maximization of, mutual information is a very involved proof. This is a maximum rate of information transfer over BSC that is equal to 1 minus. So, we have not seen a formal proof of that statement that is the capacity is the maximum rate at which information can be transmit over the channel which follows from channels theory.

Of course, Shannon has given detailed proof for this statement or for this result, however in the interest because the scope of this course is limited, we will not go into details of the proof and we rather take it as given that this as the result which governs the maximum rate at which information can be transmitted and implying this result, we are found that basically the maximum rate at which information can be transmitted across the binary symmetric channel with flip probability P is 1 minus H and needless to say of course, H of P and you can say several things from here. First of all, H of P equals H of 1 minus p. We have already seen that H of P equals H of 1 minus P which implies the capacity of a binary symmetric channel with flip probability 1 minus p. That will be 1 minus H of 1 minus P equals 1 minus H of p.

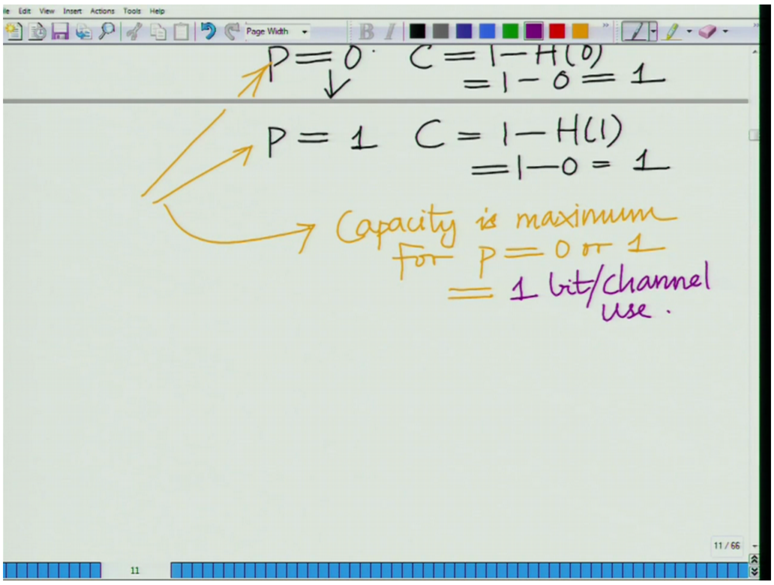
So, this is capacity with flip probability capacity with flip probability equal to 1 and actually because its symmetric channel, so if flip probability 1 minus P you can see, you can reason that similar to having a binary symmetric channel with flip probability p. So, both these channels have the same, I have an identical capacity which is 1 minus H of P or 1 minus H of 1 minus P which is basically same as 1 minus H of p.

(Refer Slide Time: 29:36)



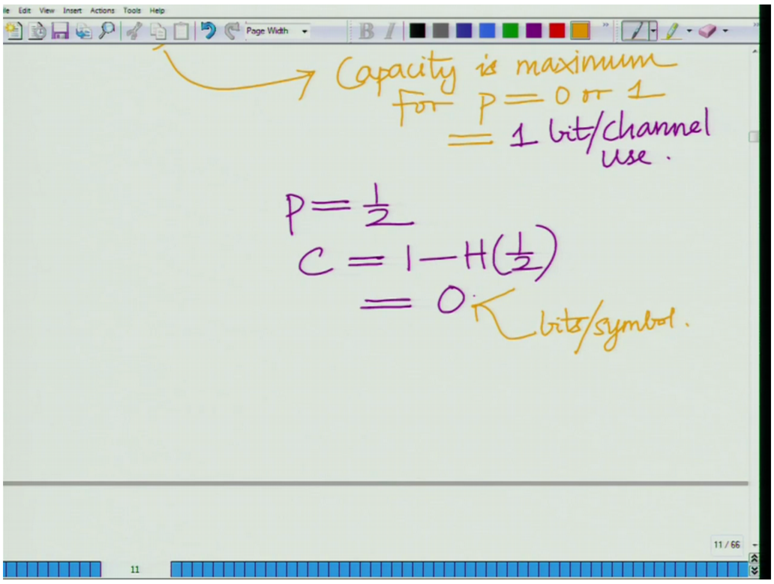
Before we part, before we complete this Module 2, Special cases one capacity is maximized for P equal to 0 in which case C equals 1 minus H of 0 equals 1 minus 0 equals 1 1 bit per channel use and also by symmetry P equal to 1, C equal to 1 minus H of 1 equals 1 minus 0 equals 1, ok.

(Refer Slide Time: 30:13)



So, for both capacity is maximum equal to 1 bit per channel use which is something very interesting. If you think about it because P equal to 0, it is fine. There is no error in the channel is 0 is always output. The output corresponding to 0 is always 0 and 1 is always received as 1, but the case corresponding to P equal to 1 is very interesting because 0 is always received as 1. With probability 1, it is flip to 1 and with probability 1, it is flipped to 0. So, the channel is always flipping a 0 to 1 and 1 to 0 and you can see you get nearly reason out in this case is also identical to the previous case. So, there is a certain symmetry about although it is counter intuitive, there is a certain symmetry about V equal to 0 and P equal to 1 scenarios and if both these scenarios, the channel capacity is exactly 1 bit per channel, use a one bit per symbol, however the minimum capacity occurs.

(Refer Slide Time: 31:29)



Naturally for P equal to half C equals 1 minus H of half equals 0, that is information rate is 0 0 bits per symbol that is you cannot transmit information. So, what the channel is doing is with probability half flipping 0 to 1 with probability half, it is transmitting 0 as a 0. Similarly, for 1 with probability half, it is flipping 1 to a 0 with probability half. 1 is received as 1 and this turns out to be the worst possible channel in which the capacity is 1 minus H P, that is 1 minus H of half which is 1 minus 1, that is 0 which means no information can be transmitted across this channel that is 0 bits per symbol. That is the information that is the rate at which information can be transmitted is 0.

So, these are the interesting results and therefore, there are several cases. Of course, this is only for the Binary Symmetric Channel and there are several other channels for instance several such as Erasure channel, Typewriter channel so on and so forth. There are several other channel for which such interesting results can be derived. Of course, the most fundamental result is the capacity, the channel capacity result which can be applied to a variety of channels to then derive what is a fundamental rate at which information can be transmitted across the channels.

There is one other important channel which is employed frequently in practice that is the Gaussian channels, additive white Gaussian noise channel, alright which basically concerns itself. What is the maximum rate at which information can be transmitted over a channel which is effected by Additive Gaussian Noise? So, that requires some more theory in the form of differential entropy etcetera which were going to start looking at in the subsequent models towards basically finally characterizing what is the capacity of this one of the most popular and one of the most relevant channels for a communication system, that is the Additive White Gaussian Noise Channel.

So, we will stop here and I would like to request you again to take a go through this module to understand the importance of this result on channel capacity and also, the implications and if possible to also try to go through the aspect of the proof of this channel capacity on your own. Although we are not going to cover it, maybe you can look at text books such as for instances text book one Information Theory, but Claude E and Thomas etcetera and try to get a glimpse of this fundamental of the proof of this fundamental result and also, the capacities of various other channels. So, we will stop here.

Thank you very much.