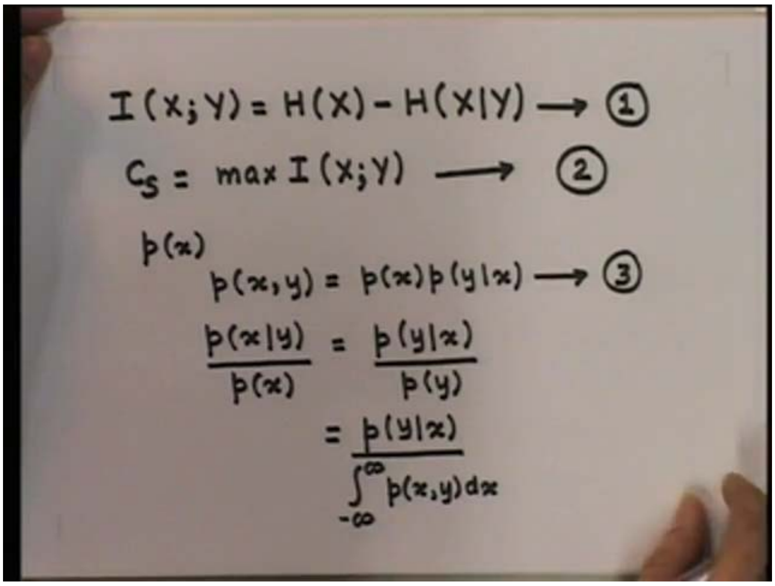
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 30
Channel Capacity of A Band Limited Continuous Channel

In the last class, we showed that a mutual information between input and output of a continuous channel. That is the information transmission over a continuous channel is related to the relative entropy of the input and equivocation of the input, with respect to output, which is uncertainty about the input after we have observed the output.

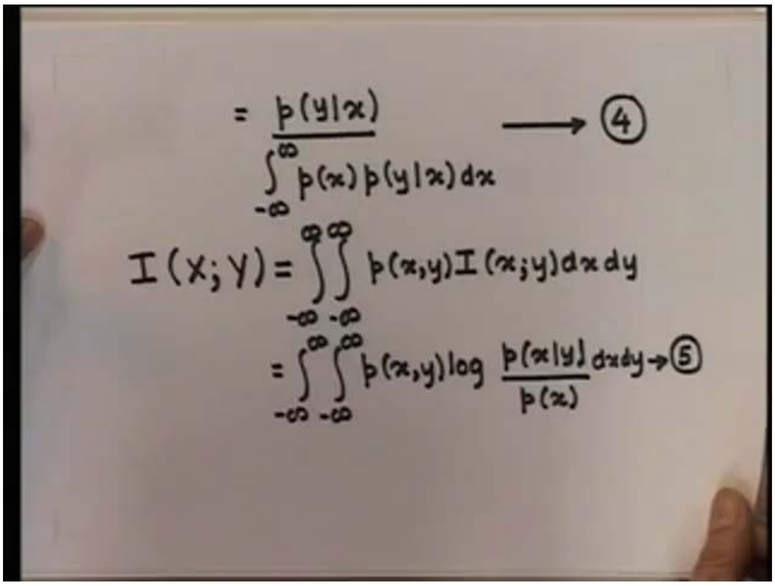
(Refer Slide Time: 01:54)



This relationship can be noted as follows. Mutual information between the input x and the output y is equal to entropy of the input less, the uncertainty about the input after we have observed the output. Now, we can define the channel capacity C s as the maximum amount of information on the average per sample or per value transmitted. So, C s is equal to maximum of the mutual information I x semicolon y.

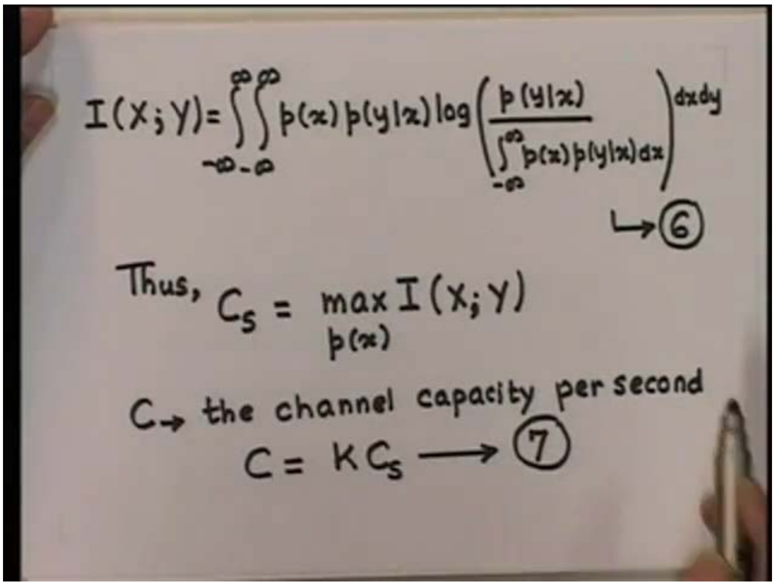
For a given channel the mutual information is a function of the input probability density p x alone. This can be shown as follows joint probability density p x y is equal to probability density p x multiplied by conditional probability density. Now, base rule says probability of x, probability distribution p x given y over probability distribution p x is equal to conditional probability distribution p of y given x over probability density p y, which can be rewritten as p d f y given x p d f p y is integral joint probability density function p x y over d x.

(Refer Slide Time: 05:02)



And this can be rewritten as p d f p y given x divided by the integral p d f p x multiplied by conditional p d f p of y given x d x. Now, we know that mutual information by definition is given as follows. Double integral of joint p d f p x y I x semicolon y integral over x and y which can be rewritten as double integration p x y log p x given y over p x d x d y. Now, if we substitute equation number three and equation number four into equation number five, we get the result for mutual information as follows.

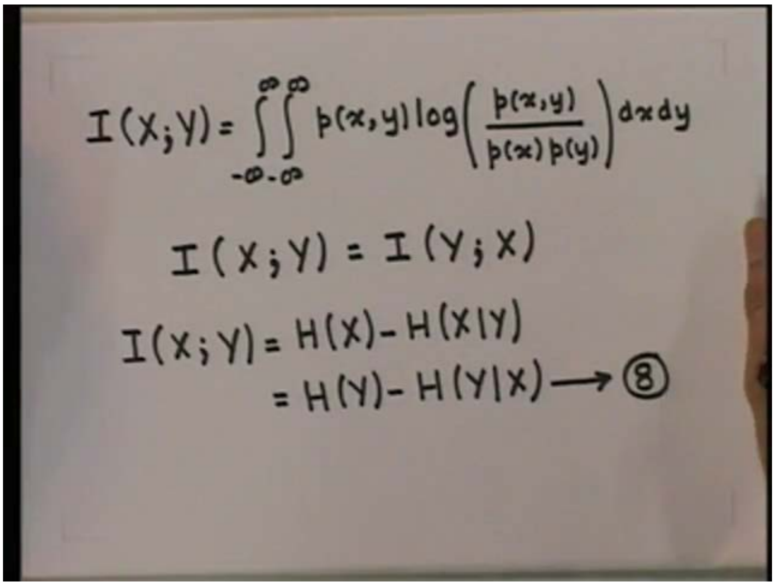
(Refer Slide Time: 07:12)



p d f p x multiplied by conditional p d f p y given x log of the quantity in the bracket conditional p d f p y given x over integral p x p y given x d x d x d y. Now, the conditional probability density p y given x is characteristic of a given channel. Hence, for a given channel mutual information I x semi colon y is a function of the input p d f p x alone. Thus, we can define the channel capacity per sample as C s is equal to maximum of mutual information over all input probability density p x.

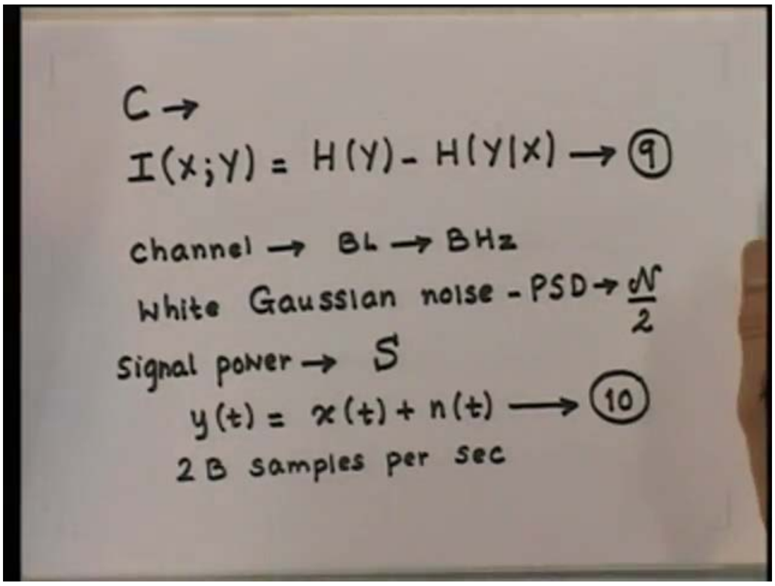
Now, if the channel allows the transmission of k values per second then C the channel capacity per second is given by C is equal to k times C s. Just as in the case of discreet variables the mutual information is symmetrical with respect to X and Y. Similarly, we can show that mutual information is symmetrical with respect to X and Y for continuous random variables too this can be seen by rewriting equation five as follows.

(Refer Slide Time: 10:24)



Joint p d f p x y log of joint p d f p x y divided by p d f p x multiplied by p d f p y integrated over x and y. Now, this equation clearly shows that the mutual information is symmetrical with respect to X and Y. Hence, we can write I X semicolon Y is equal to I mutual information between Y and X. Now, from equation one it follows that I X semicolon Y is equal to entropy of X minus equivocation of X given Y is equal to by this relationship entropy of Y less equivocation of Y with respect to X. Now, we proceed to calculate the capacity of a band limited additive white Gaussian noise channel.

(Refer Slide Time: 12:48)

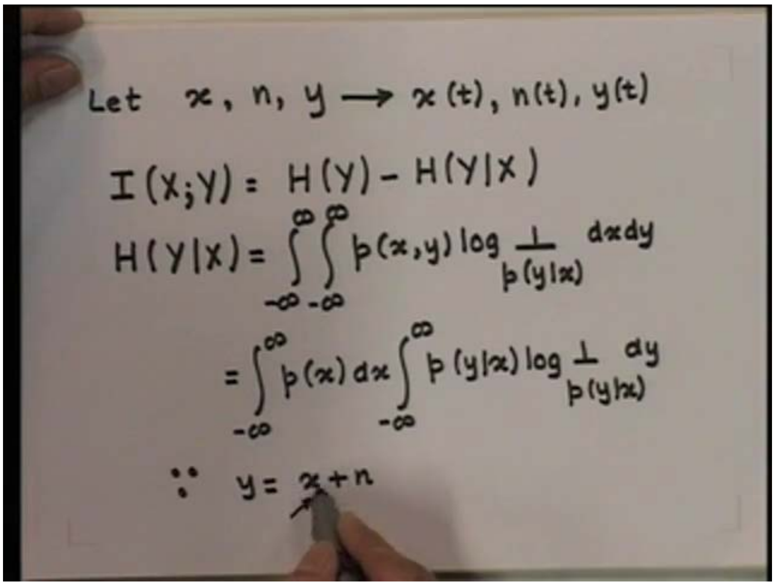


So, the channel capacity c is by definition the maximum rate of information transmission over a channel, the mutual information I X semicolon Y is given by equation eight as entropy of Y minus uncertainty of Y given X. Now, the channel capacity c is the maximum value of the mutual information per second. So, let us first find the maximum value of I X semicolon Y per sample.

We shall find here the capacity of a channel, which is band limited to b hertz and this channel is disturbed by white Gaussian noise of power spectral density given by N by 2. We assume, that this noise disturbs a channel in an additive fashion, in addition we shall constrain the signal power that is the mean value of the signal to capital S. Since, the disturbance is assumed to be additive the received signal y t is given by y t is equal to the input signal x t plus the additive noise on the channel given by n t.

Now, because the channel is band limited to b hertz both the signal x t and the noise n t are band limited to b hertz. Obviously y t is also band limited to b hertz. Now, all this signals can therefore, be completely specified by samples taken at the uniform rate of 2 b samples per second. So, let us find the maximum information that can be transmitted per sample.

(Refer Slide Time: 16:20)

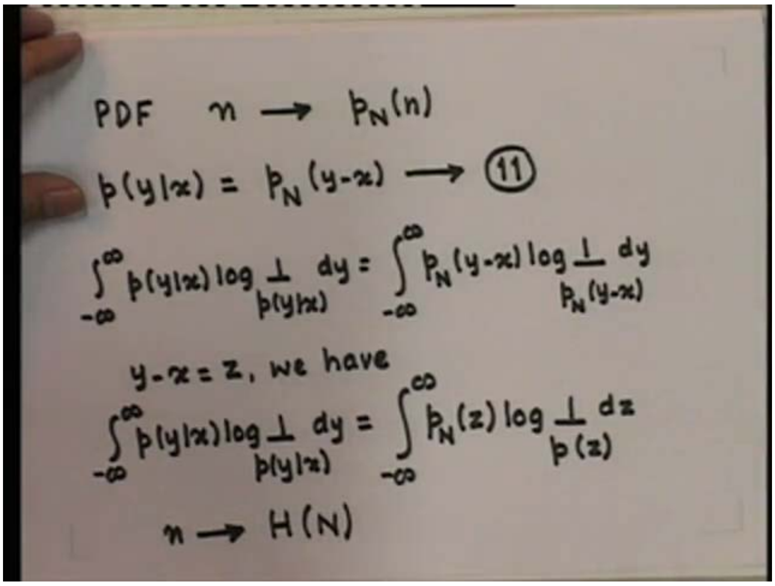


Let x n y represent samples of x t n t and y t respectively. Now, the information IX semicolon Y transmitted per sample is given by equation nine, which is rewritten here as I X semicolon Y is equal to H Y minus H of Y given X. We shall now find h of y given

x. Now by definition, H of Y given X is equal to double integral over minus infinity to plus infinity of joint p d f p x y log of 1 by conditional p d f p of y given x integrated over x and y which can be rewritten as follows.

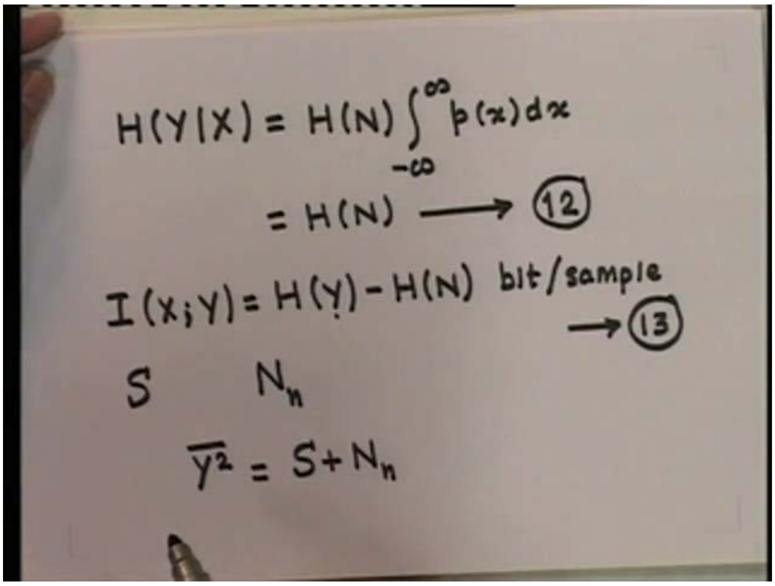
Integral of p d f p x multiplied by integral over minus infinity to plus infinity of conditional p d f p y given x log of 1 by p y given x over d y. Now, because y is equal to x plus n, what it implies that for a given x y is equal to n plus a constant x. Hence, the distribution of y when x has a given value is identical to that of n except for a translation by a constant x.

(Refer Slide Time: 19:16)



So, if p d f of noise sample n is denoted by then conditional p d f p y given x is equal to p d f p subscript n y minus x. Using this relationship we can rewrite the integral minus infinity to infinity of p y given x log of 1 by p y given x d y is equal to integral minus infinity to plus infinity of p y minus x log of p n y minus x integrate over y. Now, letting y minus x is equal to z, we have minus infinity to infinity of p of y given x log of 1 by p y given x over y equals as p z log of 1 by p conditional p d f of noise. Now, the right hand side is the entropy of the noise sample n and this is given by H of n.

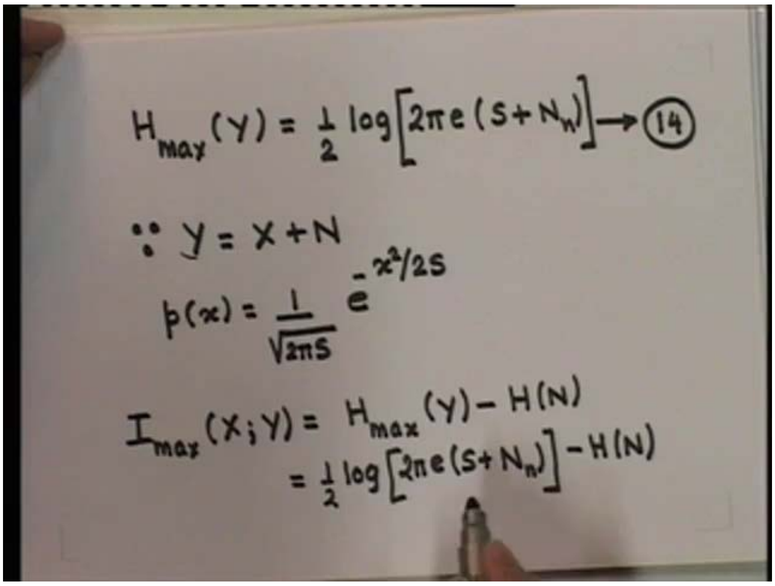
(Refer Slide Time: 21:57)



Hence, H of Y given X is equal to H of n multiplied by minus infinity to plus infinity integration of p x and this is equal to H N. In deriving equation twelve we made no assumptions about the noise, hence equation twelve is very general and applies to all types of noise. The only condition is that the noise disturbs the channel in an additive manner. Thus, we can write the mutual information between X and Y as H of Y minus H of N bits per sample.

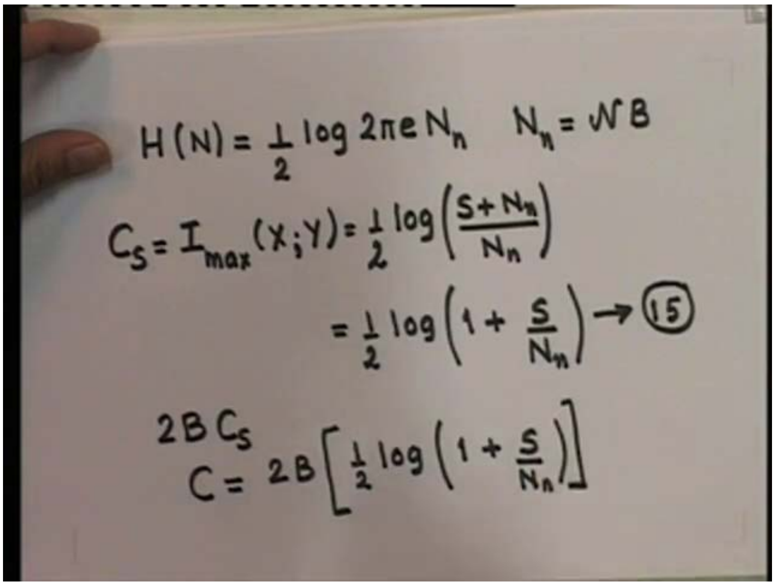
We have assumed that the mean square value of the signal x t is constrained to have a value capital S and the mean square value of the noise is denoted by N subscript small n. Now, we shall assume that the signal x t and the noise n t are independent. In such a case the mean square value of Y will be the sum of the mean square value of X and N. Hence, mean square value of Y is equal to S plus N subscript small n. Now, for a given noise that is H N is given I X semicolon y that is the mutual information is maximum, when H Y is maximum.

(Refer Slide Time: 24:57)



Now, we have also seen that for the given mean square value of Y which is indicated here H Y will be maximum, if Y is Gaussian and the maximum entropy h max Y is given by half log of 2 pi e multiplied by S plus N n, this relationship we have derived earlier. Now, because Y is equal to X plus N and N is Gaussian Y will be Gaussian only if X is Gaussian. Now, as the mean square value of x is s this implies that the p d f of x is equal to 1 by root of 2 pi s multiplied by exponential e raise to minus x squared by 2 S and the maximum value of the mutual information is given as follows. H max Y minus H N is equal to half log 2 pi e S plus N n minus entropy of the noise.

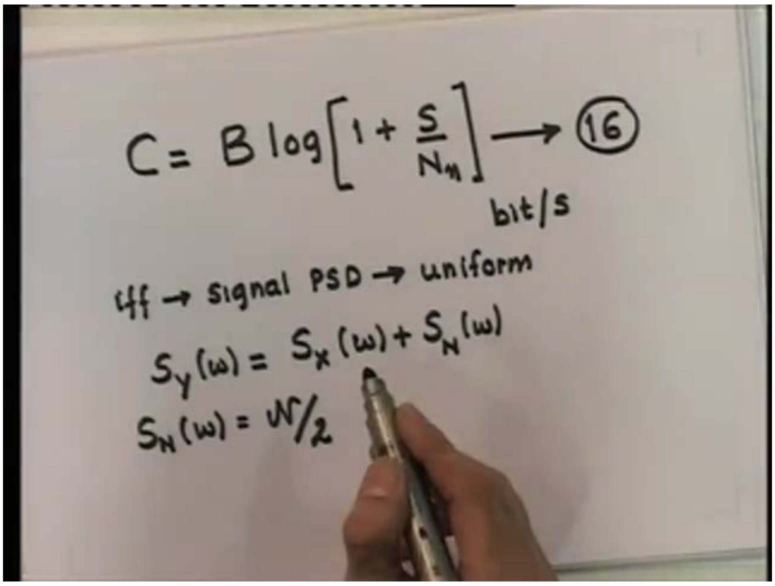
(Refer Slide Time: 27:15)



Now, for a white Gaussian noise with mean square value of N subscript n the entropy of the noise is given by H of noise is equal to half log 2 pi e N subscript n, where noise power N subscript n is equal to bandwidth multiplied by twice the power spectral density. Therefore, channel capacity C s per sample is equal to half log S plus N n over N subscript n is equal to half log 1 plus signal to noise ratio.

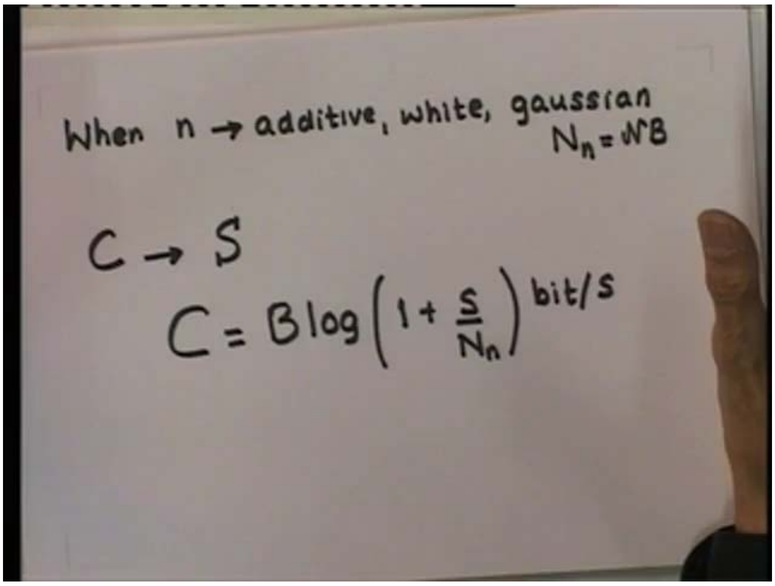
Now, the channel capacity per second will be the maximum information that can be transmitted per second equation fifteen represents, the maximum information transmitted per sample. Now, if all the samples are statistically independent the total information transmitted per second will be two B times C s. If the samples are not independent then the total information will be less than this quantity because the channel capacity is C represents the maximum possible information transmitted per second.

(Refer Slide Time: 30:04)



We get channel capacity as C is equal to twice B multiplied by C s, which is half log of 1 plus signal to noise ratio, which is simplified to C is equal to B times log of 1 plus signal to noise ratio, this is bit per second. Now, the samples of a band limited Gaussian signal are independent if and only if the signal power spectral density is uniform. Obviously, it would transmit information at the maximum rate given by equation sixteen. The power spectral density of signal y t must be uniform. Now, the power spectral density of y is given by S y is equal to power spectral density of input plus power spectral density of noise assuming signal and noise are independent because power spectral density of noise is y denoted by N by 2 the power spectral density of x t must also be uniform. Thus, the maximum rate of transmission that is C given in terms of bits per second is attained when the input signal x t is also a white Gaussian signal.

(Refer Slide Time: 32:18)



So, to recapitulate when the channel noise is additive white and Gaussian with the mean square value given by N subscript small n, which is equal to twice the power spectral density multiplied by the bandwidth. Then the channel capacity C of a band limited channel under the constrain of given signal power S is given by C is equal to b times log of 1 plus signal to noise ratio. B is the channel bandit in hertz and this maximum rate is realized only if the input signal is a white Gaussian signal.

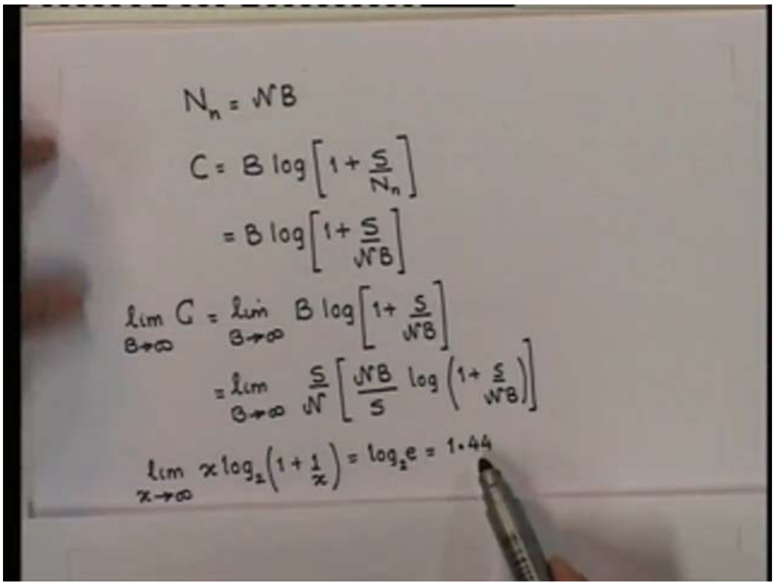
Using signal space, this theorem can be verified in a way similar to the clues for the verification of the channel capacity of a discreet case, which we had done earlier in the course. This channel coding theorem indicated by Shannon's equation is the maximum error free communication rate achievable on an optimum system without any restrictions, except for bandwidth restricted to B hertz signal power restricted to capital S. And Gaussian white channel noise power given by N subscript small n.

If we have any other restrictions, this maximum rate will not be achievable. For example, if we take a binary channel that is a channel which is restricted to transmit only binary signals we will not be able to attain Shannon's rate even if the channel is optimum. The channel capacity formula indicates that the transmitted rate is a monotonically increasing function of the signal power S.

Now, if we use a binary channel however, we can show that increasing the transmitted power beyond a certain value buys very little advantage. Hence on a binary channel increasing S that is the signal power will not increase the error free communication rate beyond some value. This does not mean that the channel capacity formula has failed, it simply means that when we have large amount of signal power with a finite bandwidth then the binary scheme is not the optimum communication scheme.

Shannon’s result provide us the upper theoretical limit of error free communication, but it does not tell us how to achieve this. Now, let us look at the capacity of the channel in the case of infinite bandwidth. Superficially, this Shannon’s equation seems to indicate that the channel capacity goes to infinity as the channel bandwidth b goes to infinity.

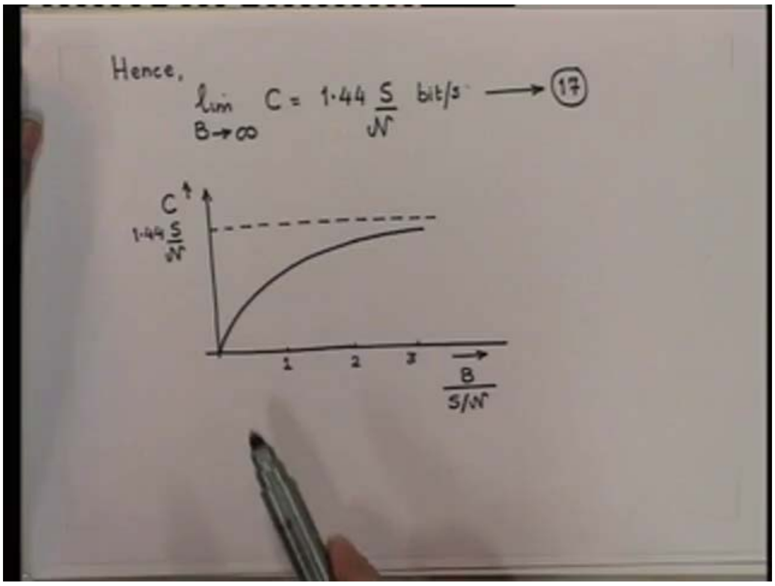
(Refer Slide Time: 36:37)



This however is not true, for white noise the noise power is given as follows. Hence, as B increases the noise power also increases, now it can be shown that S B tends to infinity the channel capacity C approaches a limit as follows. C is given by this relationship which we just derived, this can be rewritten as follows, where the noise power is written in terms of power spectral density and the bandwidth of the channel.

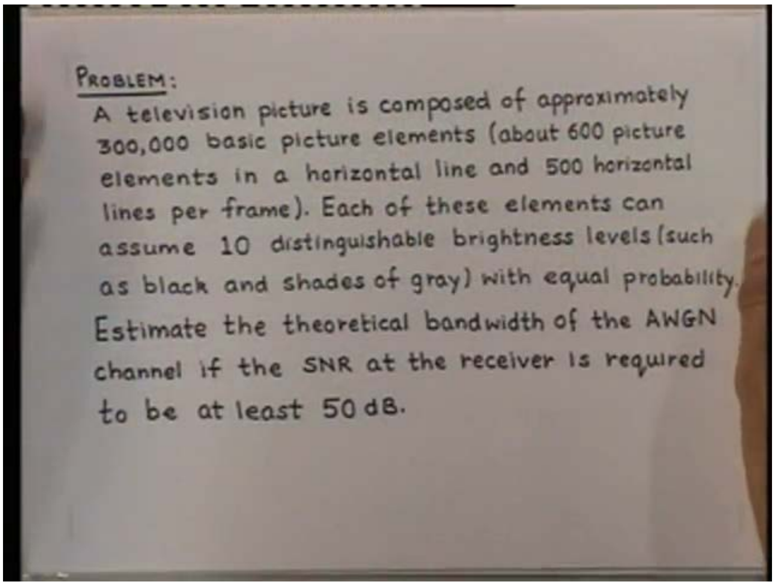
Now, what is desired is to calculate the channel capacity when B tends to infinity, which is the calculation of this term on the right hand side. Now, this term on the right hand side can be rewritten as follows. Now, we can use the relationship that limit of x tending into infinity of the value x log to the base 2 of 1 plus 1 by x is equal to log of e to the base e which is equal to 1.44.

(Refer Slide Time: 38:07)



Using this we get the final result for the channel capacity when B tends to infinity as given by this expression on the right hand side of the equation number seventeen. And if we plot this we get this graphical representation of the channel capacity C with respect to bandwidth B for given signal power S, and for the given noise power spectral density N. Now, it is evident that a capacity can be made, this relationship is depicted in the figure as shown here, where the channel capacity C is plotted against bandwidth B normalise for the given signal power S and power spectral density N. It is evident that the capacity can be made infinite only by increasing the signal power S to infinity for a given optimum system. For finite signal and noise powers the channel capacity will always remain finite.

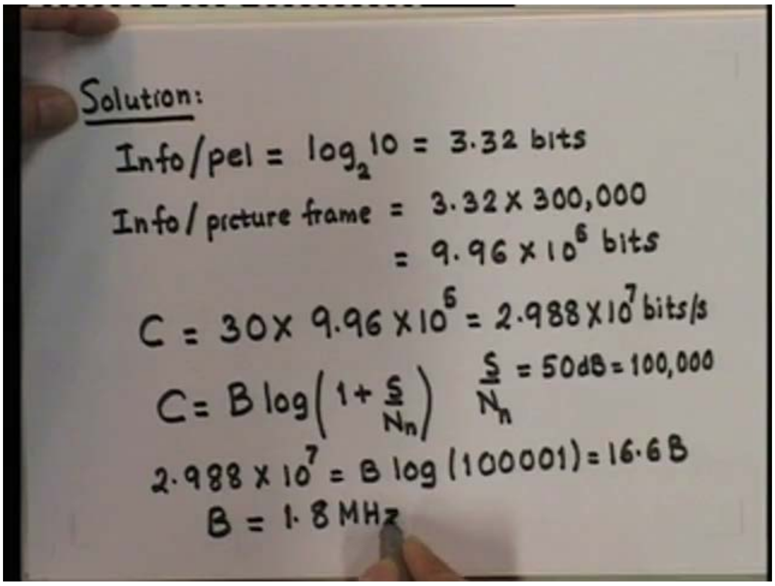
(Refer Slide Time: 39:39)



Now, based on this idea we will solve a problem as follows, let us assume that we are given a television picture, which is composed of approximately 300,000 basic picture elements about 600 picture elements in a horizontal line and 500 horizontal lines per frame. Each of these elements in a television picture can assume 10 distinguishable brightness levels, such as black and shades of grey with equal probability.

Now, we are required to calculate or estimate the theoretical bandwidth of the additive white Gaussian noise channel, if the signal to noise ratio at the receiver is required to be at least 50 db. So, the solution to this problem is as follows. First, let us calculate information per picture frame, information per picture element that is known as pel is equal to log of 10 to the base 2 assuming all the brightness levels are equal probable, which gives 3.32 bits per pel.

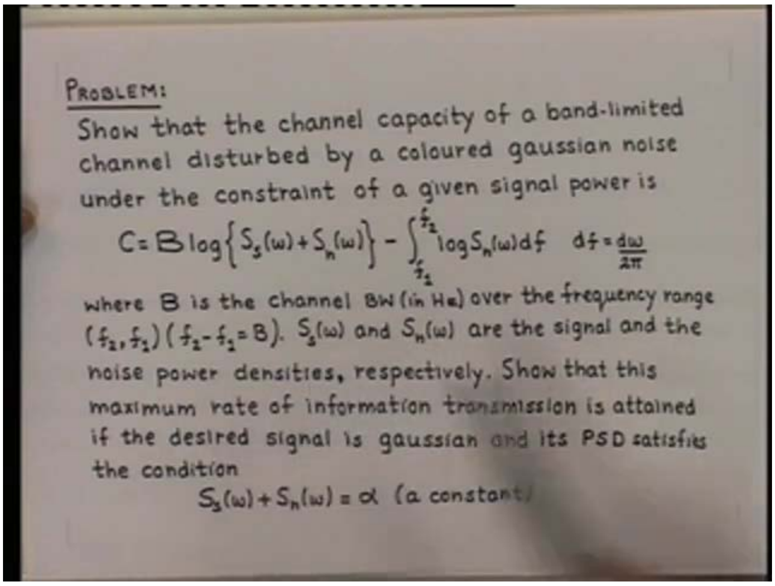
(Refer Slide Time: 40:39)



Now, information per picture frame is equal to information per element multiplied by the number of pels in a picture frame, which is 300000 and this is equal to 9.96 multiplied by 10 is to 5 bits. Now, for 30 picture frames per second we need a channel with a capacity C equal to 30 multiplied by this quantity which is equal to 2.988 multiply by 10 is to 7 bits per second. Now, we know that for a additive white Gaussian noise channel capacity C is equal to bandwidth multiply by log of 1 plus signal to noise ratio. We have been given that signal to noise ratio that is S by N subscript small n is equal to 50 db, which corresponds to 100,000.

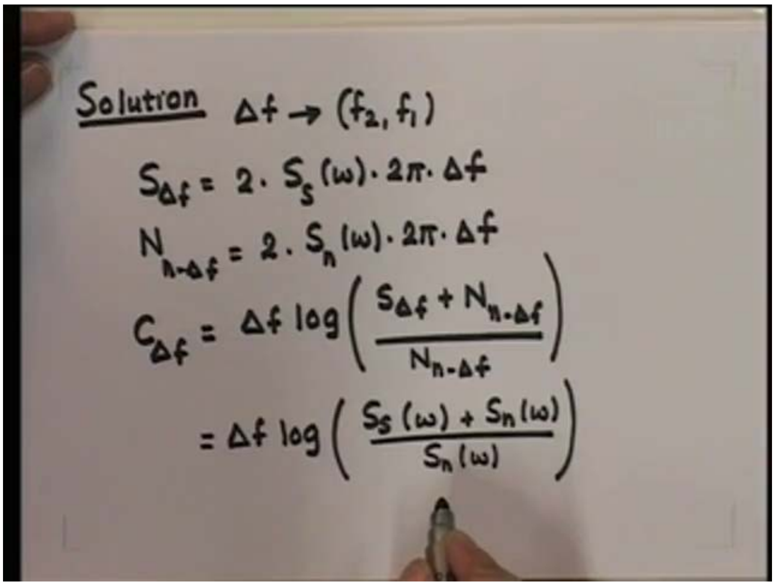
Therefore, using this value we get the following equation and from this, we can calculate the theoretical estimation of the bandwidth of the additive white Gaussian noise channel as b equal to 1.8 megahertz. Having calculated the channel capacity for the case of additive white Gaussian noise, let us calculate the channel capacity of a band limited channel disturbed by additive coloured Gaussian noise.

(Refer Slide Time: 44:15)



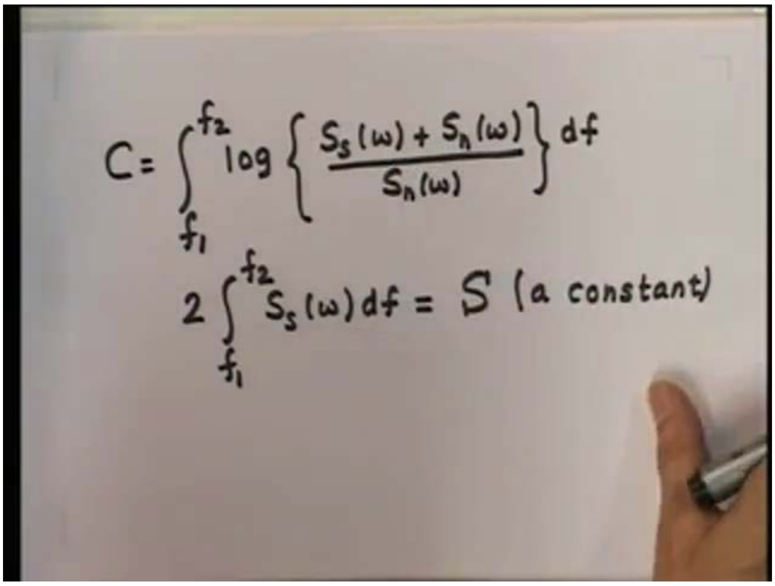
So, the problem is to show that the channel capacity of a band limited channel disturbed by a coloured Gaussian noise under the constraint of a given signal power is given by this expression, where B is the channel bandwidth in hertz over the frequency range f 2 to f 1. S s and S n are the signal and the noise power densities respectively and we have to show that this maximum rate of information transmission is attained if the desired signal is Gaussian, and its power spectral density satisfies the following condition. The solution to this problem is as follows.

(Refer Slide Time: 45:03)



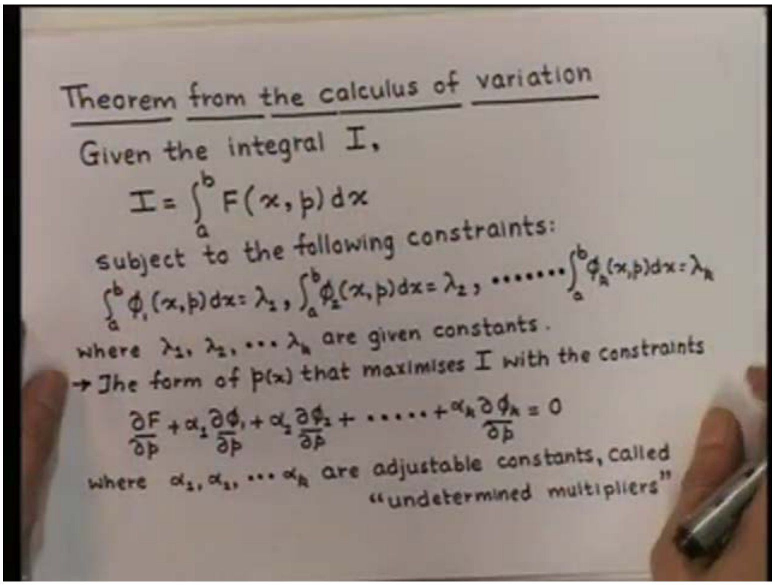
Consider a narrowband delta f where delta f tends to 0 over the range f 2 f 1. So, that we may consider both signal and noise power density to be constant, that is band limited white over the interval delta f. In this case the signal and noise power over the band delta f is given as follows S delta f is equal to twice f and the noise power over the band delta f is equal to. Now, the maximum channel capacity C over this band delta f is given by C delta f is equal to delta f log of signal power plus noise power divided by noise power, which can be rewritten as delta f log of signal power densities plus noise power density divided by noise power density.

(Refer Slide Time: 47:12)



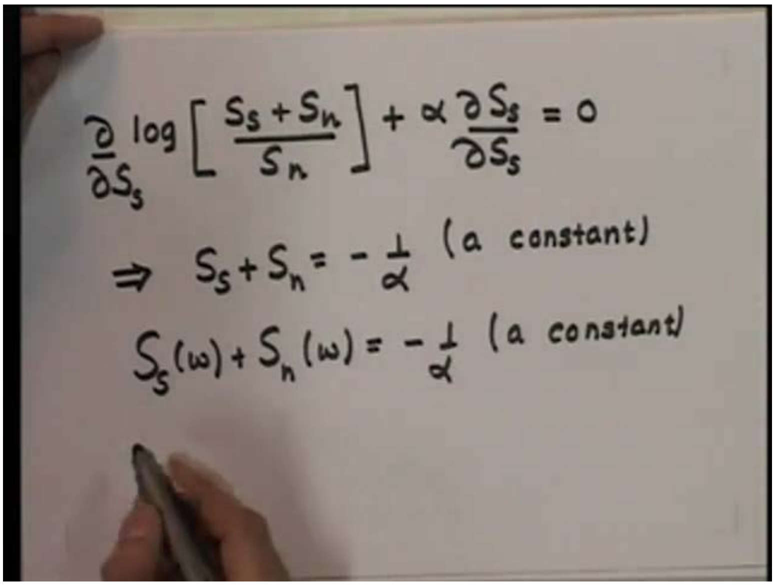
Now, the capacity of the channel over the entire band f 2 to f 1 is given by the integral C is equal to integral from f 1 to f 2 of the quantity log of signal power density plus noise power density divide by noise power density integrated over frequency f 1 to f 2. Now, we wish to maximize C, where the constraint is that signal power is constant that is twice integral of f 1 to f 2 of signal power density is equal to capital S a constant.

(Refer Slide Time: 48:23)



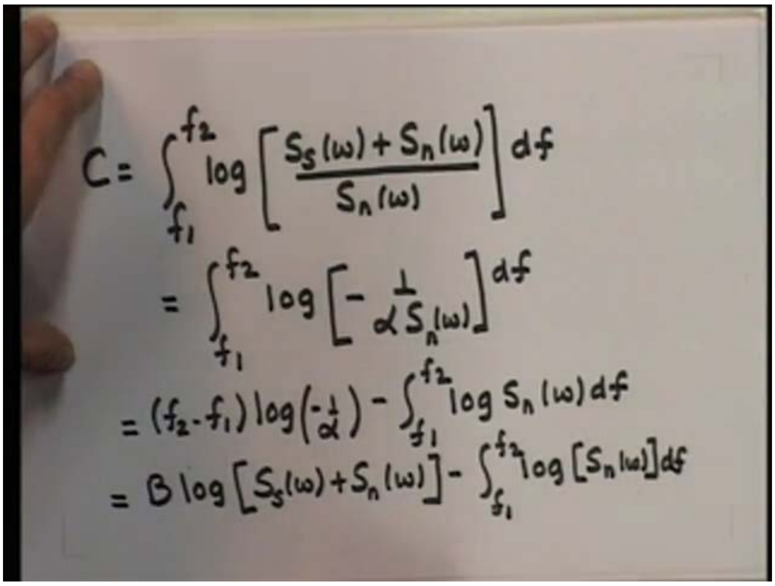
Now, to solve this problem we use the theorem from the calculus of variation, which we had seen earlier given the integral I of the form subject to the following constraints. Where lambda 1 lambda 2 up to lambda k are given constants the form of p that maximises I with the constraints is given by this equation, where alpha 1 alpha 2 and alpha k are adjustable constants called undetermined multipliers.

(Refer Slide Time: 49:03)



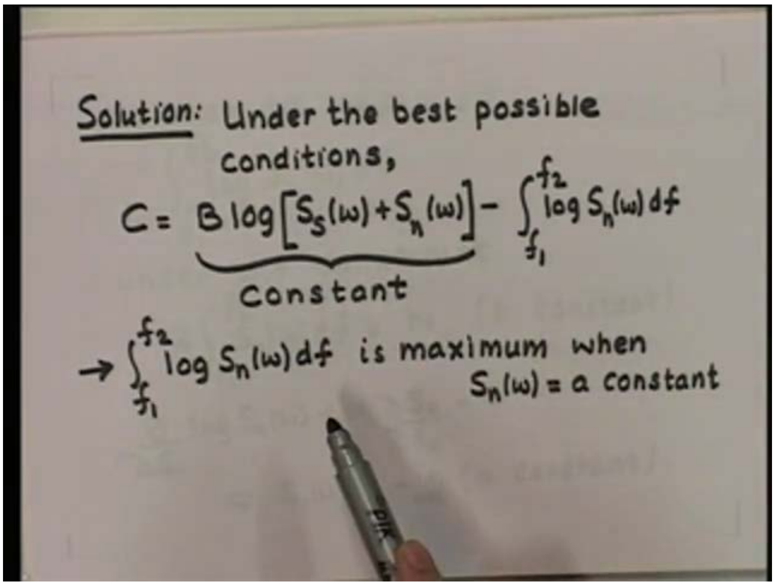
Using this relationship we can get the following equation from this equation, it is not very difficult to show that signal power density plus noise power density is equal to minus 1 by alpha, which is a constant since alpha is a constant. Thus signal power density and noise power densities are related by this relationship. This shows that to attain the maximum channel capacity, the signal power density and the noise power density must be a constant that is a white.

(Refer Slide Time: 50:24)



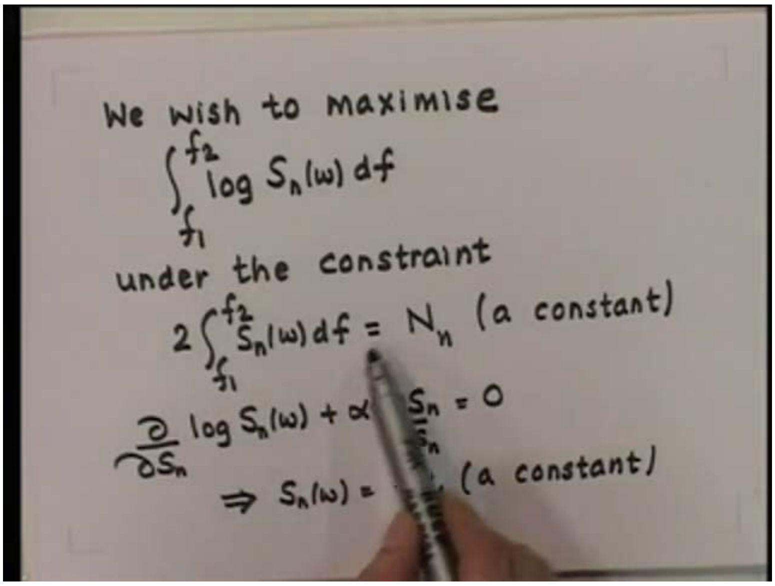
Under this condition we can calculate it the channel capacity as, is equal to and this can be simplified as which is equal to B times log of signal power density plus noise power density minus integral of the quantity log of noise power density. So, this is the relationship which we were seeking. Now, using the results in the earlier problem we can show that the worst kind of Gaussian noise is a white Gaussian noise that is constrained to a given mean square value. Now, the solution to this problem can be found as follows.

(Refer Slide Time: 51:50)



Under the best possible condition we have shown from the earlier problems result that C is equal to this quantity, where this quantity is a constants. So, the worst case will be when this quantity is a maximum. Now, we will show that this quantity is maximum when power spectral density of noise is a constant, when the noise is constrained to have a certain mean square value.

(Refer Slide Time: 52:27)



So, the problem is to maximize this quantity under the constraint of given noise power. So, again using the theorem from the calculus of variation we get the result as follows and from this we get that the power spectral density of the noise is equal to minus 1 times alpha where alpha is a constant so this term turns out to be a constants.

Thus we have shown that for a noise with a given power the integral is maximum, when the noise is white this shows that the white Gaussian noise is the worst possible kind of noise. Now, during the course of our study we have studied source coding theorem for a discreet memory less source. According to this theorem average code word length must be at least as large as the source entropy for perfect coding that is perfect representation of the source.

However, in many practical situations there are constraints that force the coding to be imperfect there by resulting in unavoidable distortions. For example, constraint imposed by communication channel may place an upper limit on the permissible code length and therefore, average code word length assigned to the information source. As another example the information source may have a continuous amplitude as in the case of speech.

And the requirement is to quantise the amplitude of each sample generated by the source to permit it's representation by a code word of finite length as impulse code modulation, in such cases the problem is referred to as source coding with fidelity criterion. And the branch of information theory that deals with it is called rate distortion theory. In the next class, we will begin our study with the rate distortion theory and see how it helps us to design efficient lossy compression schemes.