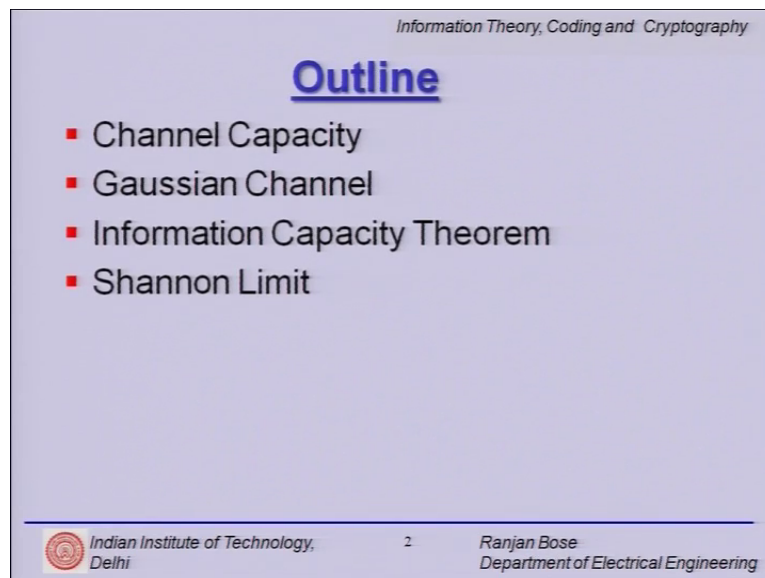


Information Theory, Coding and Cryptography
Dr. Ranjan Bose
Department of Electrical Engineering
Indian Institute of Technology, Delhi

Module - 11
Channel Capacity and Coding
Lecture - 11

Hello and welcome to the next module on Channel Capacity and Coding. We will start with the brief outline for today's lecture, we will revisit channel capacity. And then focus our attention on Gaussian channels, we will then state improve the information capacity theorem.

(Refer Slide Time: 00:51)



Information Theory, Coding and Cryptography

Outline

- Channel Capacity
- Gaussian Channel
- Information Capacity Theorem
- Shannon Limit

Indian Institute of Technology, Delhi 2 Ranjan Bose
Department of Electrical Engineering


And finally, we will spend some time on the Shannon limit.

(Refer Slide Time: 00:58)

Information Theory, Coding and Cryptography

Recap

- Channel Capacity
- Symmetric Channels
- Noisy Channel Coding Theorem
- Repetition Code

 Indian Institute of Technology,
Delhi3Ranjan Bose
Department of Electrical Engineering

Let us quickly recap what we have done so far. We have developed the general notion of the capacity of a channel. We had looked at symmetric channels and figured out how we can quickly compute the capacity for symmetric and weakly symmetric channels. We then looked at noisy channel coding theorem and we took an example of a very simple error correcting code called the repetition code.


(Refer Slide Time: 01:27)

Information Theory, Coding and Cryptography

Channel Capacity

$$C = \max_{P(x_j)} I(X; Y)$$
$$= \max_{P(x_j)} \sum_{j=0}^{q-1} \sum_{i=0}^{r-1} P(x_j) P(y_i | x_j) \log \frac{P(y_i | x_j)}{P(y_i)}$$

- $C \geq 0$, since $I(X; Y) \geq 0$.
- $C \leq \log|X|$, since $C = \max I(X; Y) \leq \max H(X) = \log|X|$.
- $C \leq \log|Y|$, since $C = \max I(X; Y) \leq \max H(Y) = \log|Y|$.

 Indian Institute of Technology,
Delhi4Ranjan Bose
Department of Electrical Engineering

So, very quickly to refresh our memory we have talked about the channel capacity C as the maximum of average mutual information between X on one side of the channel and

Y on the other side of the channel; where the maximization is done over all input probabilities and if you write it out the expression for capacity looks like this. Please note some of the interesting properties of the channel capacity. Capacity is necessarily greater than or equal to 0, since the average mutual information is greater than or equal to 0. We also established that the capacity should be less than or equal to $\log X$ where X is the cardinality and similarly later should be less than or equal to \log absolute value Y, where this sign shows the cardinality of Y.

This simply comes from like the extension of this which we should information I X semicolon into Y should be equal to H X minus H X given Y should be equal to H of Y minus H Y given X. So, this we have done already.

(Refer Slide Time: 02:41)

Information Theory, Coding and Cryptography

Noisy Channel Coding Theorem


- Let a DMS with an alphabet X have entropy $H(X)$ and produce symbols every T_s seconds.
- Let a discrete memoryless channel have capacity C and be used once every T_c seconds.
- Then if $\frac{H(X)}{T_s} \leq \frac{C}{T_c}$

there exists a coding scheme for which the source output can be transmitted over the noisy channel and be reconstructed with an arbitrarily low probability of error

- Conversely if $\frac{H(X)}{T_s} > \frac{C}{T_c}$

It is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error.

- The parameter $\frac{C}{T_c}$ is called the **Critical Rate**.

 Indian Institute of Technology,
Delhi

5

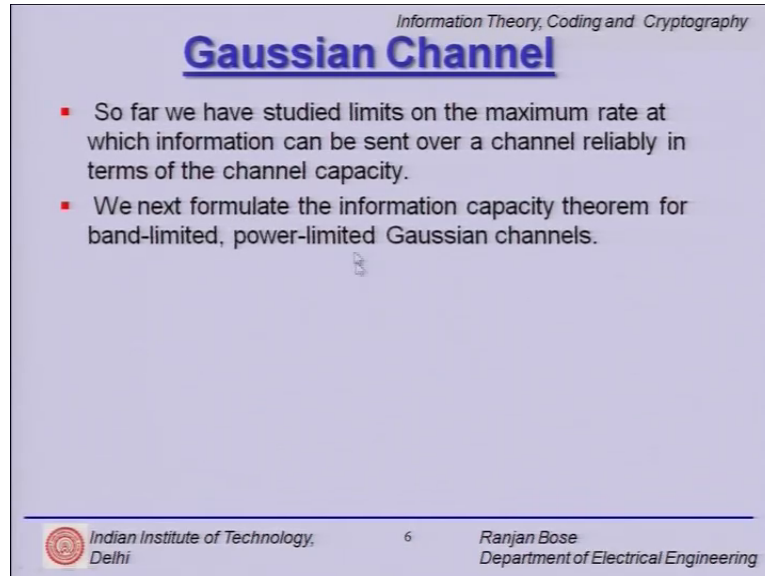
Ranjan Bose
Department of Electrical Engineering

A very quick recap on the noisy channel coding theorem where we started off with a discrete memory less source with an alphabet X with a given entropy H X and producing symbols every T S second. And we established that with if the source rate H X divided by T S is less than or equal to C over T C; then we have a possible coding scheme, where we can transmit information over this noisy channel which is unreliable and at the same time re construct with arbitrarily low probability of error at the receiving end.

On the other hand, if H X divided by T S source rate is greater than this quantity which we will define as the critical rate then it is not possible to transmit information reliably.

This parameter we already defined is called the critical rate ok. So, this is what the noisy channel coding theorem tells us.

(Refer Slide Time: 03:47)



The slide is titled "Gaussian Channel" in a large, bold, blue font. Above the title, in a smaller font, is "Information Theory, Coding and Cryptography". Below the title, there are two bullet points in red. The first bullet point says: "So far we have studied limits on the maximum rate at which information can be sent over a channel reliably in terms of the channel capacity." The second bullet point says: "We next formulate the information capacity theorem for band-limited, power-limited Gaussian channels." At the bottom of the slide, there is a footer with three items: on the left, the Indian Institute of Technology Delhi logo and name; in the center, the number "6"; and on the right, the name "Ranjan Bose" and "Department of Electrical Engineering".

Now, very quickly let us look at a very interesting and practical channel called the Gaussian channel.

So, we now formulate the information capacity theorem for a band limited and power limited Gaussian channel. Please note in the binary symmetric channel there was no notion of bandwidth; did it figure out? Not at all where was the notion of power? Well, it was implicit in the channel transition probabilities, but it was never explicitly stated, but in real life these are the 2 quantities we must deal with bandwidth, yes we pay a lot of money. Remember the auctions for the wireless spectrum ok. So, band limited channels are a reality, way of life and power you do not have infinite battery power, you have to outdo your competitors your mobile phone should not drain out in the middle of your conversation.

So, the channels are not only band limited, but they also power limited Gaussian we will talk about.

(Refer Slide Time: 05:03)

Information Theory, Coding and Cryptography

Gaussian Channel

- An important and useful channel is the **Gaussian Channel**
- This is a time discrete channel with output Y_k at time k .
- This output is the result of the sum of the input X_k and the noise Z_k . This noise is drawn from a Gaussian distribution with mean zero and variance σ^2 .
- Thus, $Y_k = X_k + N_k$, where $N_k \sim N(0, \sigma^2)$.
- The noise N_k is independent of the input X_k .

Indian Institute of Technology, Delhi 7 *Ranjan Bose*
Department of Electrical Engineering

So, let us have a very very simple model. So, X of k is at the transmitter site and Y of k is at the receiver site what the channel does is it adds noise which is Gaussian most specifically native white Gaussian noise.

Now, the sub k represents the samples; so, k -th sample of the symbol gets added with the k -th noise sample to get the Y_k output. So, let us talk about this discrete channel which is the time discrete channel with output Y_k at time k . And let us say the noise sometimes it is denoted Z_k literature, but here we have noted as N_k is drawn from a Gaussian distribution with mean 0 and variance σ^2 . Then we can write Y_k simply as X_k plus N_k where N_k is taken from this Gaussian distribution.

Please note the first assumption we are going to make; we are going to make several assumptions the first and foremost is that the noise N_k is independent of X_k . So, this is the first assumption we make and we will use this assumption to derive the information capacity theorem. Is this a fair assumption? Well, it should be, if we have to really work hard to make it dependent; in general the sources independent and the noise is generated independently in the circuit. So, it is a good fair believable assumption that X_k is really independent of k .


(Refer Slide Time: 07:02)

Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

- Since the transmitter is usually power-limited, let us put a constraint on the average power in X_k :
$$E[X_k^2] = P \quad k = 1, 2, \dots, K.$$
- Thus, the information capacity of the channel (same as the channel capacity) is given by
$$C = \max_{f_{X_k}(x)} \{I(X; Y) \mid E[X_k^2] = P\}$$

probability density function of X_k .

 Indian Institute of Technology, Delhi

8

Ranjan Bose
Department of Electrical Engineering

Now, the problem before us is to figure out what could be the capacity of a Gaussian channel that we have defined. This Gaussian channel we have already qualified it with being power limited and bandwidth limited. So, band limited and power limited since the transmitter is usually power limited; let us put a constraint on the average power of X_k remember X_k is what we are transmitting at time k . So, what do you mean by average? Well the expected value of X_k squared is P . So, this is the average power it is possible that some of the symbols are being transmitted with lower power some of them are be transmitted with higher power, but on an average we have the input symbols having a power of P .

So, the problem now translates to a constraint problem where we have to find out the capacity as maximization of the average mutual information $I(X; Y)$ given expected value of X_k squared is P as always the maximization will be over input probability distribution. So, this condition we have added; so, what are the 2 things we are looking at? This power limited this putting this condition expected value of X_k square is P and the probability density function of X_k is this $f_{X_k}(x)$.

So, this slide tells us that we are right now focused on the transmitter side where the symbols are being generated with X_k equal to 1 2 3 4 up to capital k . So, we proceed.

(Refer Slide Time: 09:09)


Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

- X_k and N_k are independent random variables.
- Therefore, the conditional differential entropy of Y_k given X_k is equal to the differential entropy of N_k .
- Intuitively, this is because given X_k the uncertainty arising in Y_k is purely due to N_k . That is,
$$h(Y_k | X_k) = h(N_k)$$

Hence we can write

$$I(X_k; Y_k) = h(Y_k) - h(N_k)$$
$$I(X_k; Y_k) = h(Y_k) - h(Y_k | X_k)$$

 Indian Institute of Technology, Delhi

9

Ranjan Bose
Department of Electrical Engineering

So, we start with the differential entropy and we say that the average mutual information between X_k and Y_k . So, only focusing on the k -th symbol is equal to $h(Y_k)$ because please note Y is a continuous random variable and Y_k is a sample at time instant k . So, small h representing the differential entropy; so, $h(Y_k)$ minus $h(Y_k | X_k)$ this is just the definition of this average mutual information.

Now, we have already assumed that X_k and N_k are indeed independent random variables. So, we think a little harder and say that the conditional differential entropy of Y_k given X_k . So, we are looking at the second term on the right hand side; so, the conditional differential entropy of Y_k given X_k is nothing but the differential entropy of N_k why because, X_k and N_k are independent. So, if you go back to your original model X_k and Y_k are independent, so if you want the uncertainty $h(Y_k | X_k)$, right.

So, if I know already if I have been given X_k then the uncertainty in Y_k is coming from where only because of N_k . So, $h(Y_k | X_k)$ is nothing but $h(N_k)$. So, we come back and say that intuitively given X_k the uncertainty arising in Y_k is purely due to N_k that is to say $h(Y_k | X_k)$ is equal to $h(N_k)$. So, this is a very important assumption and it is based on X_k and Y_k and N_k being independent random variables.

So, just because we have a strong intuitive feel we could simplify this expression to $I(X; Y)$ is semicolon Y k average mutual information is nothing but the difference of the uncertainty $h(Y)$ k minus $H(N)$ k . So, if you bring it closer and compare we have got this second term $h(Y)$ k given N k being replaced by $H(m)$ k . So, we have we are moving forward.

(Refer Slide Time: 12:02)

Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

- If we assume Y_k to be Gaussian, and N_k is Gaussian by definition, then X_k is also Gaussian.
- This is because the sum (or difference) of two Gaussian random variables is also Gaussian.
- Thus, in order to maximize the mutual information between the channel input X_k and the channel output Y_k , the transmitted signal should be Gaussian.
- Therefore we can write

$$C = I(X; Y) | E[X_k^2] = P \text{ and } X_k \text{ is Gaussian}$$

Indian Institute of Technology, Delhi
10
Ranjan Bose
Department of Electrical Engineering

Now we have to make some more assumptions. So, if we assume Y_k to be Gaussian and N_k to be Gaussian then by definition X_k is also Gaussian why do we have to do all of this? Well these assumptions are being made to maximize the average mutual information in order to obtain the capacity value. So, we have to get some distribution of X_k . So, as to maximize the average mutual information, but to make any comment about the distribution X_k I must then talk about Y and N_k .

Now, N_k is Gaussian by definition and if you are assuming Y_k to be Gaussian then X_k to be also Gaussian. This is because the sum or even the difference of any 2 Gaussian random variables is also Gaussian. So, in order to maximize the mutual information between the channel input X_k and output Y_k , the transmitted signal should be Gaussian; so, these are the logical steps to follow. So, what we can write now is that capacity.

Now, we have said that we will maximize given the distribution of X_k , but that X_k which maximizes the average mutual information is Gaussian and I will comment upon why is it so. So, then C is equal to the average mutual information X semicolon Y under

now 2 constraints; one is the power limited condition that the average power of the symbols at the input is P . And again at the input we request that this X_k to be Gaussian in order to write this capacity.

So, this maximum value and its gone maximization is already done under the assumption that X_k is Gaussian that is the take home message here.

(Refer Slide Time: 14:17)

Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

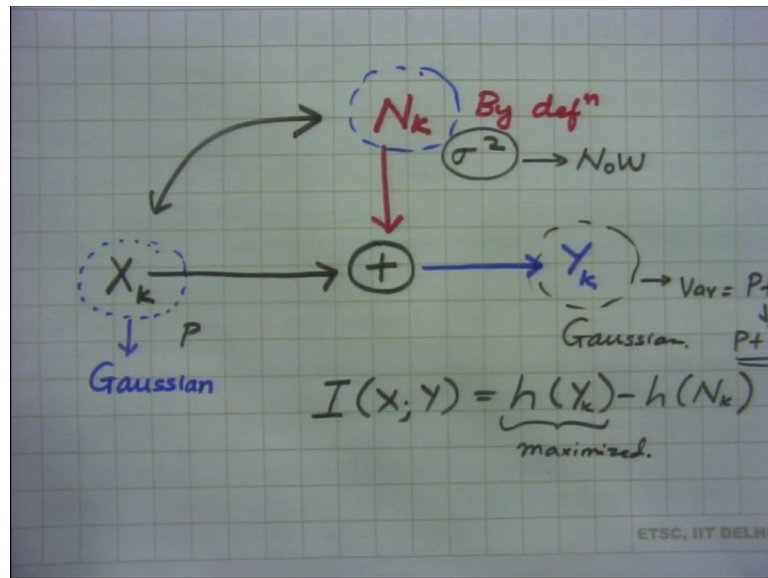
- We know that if two independent Gaussian random variables are added, the variance of the resulting Gaussian random variable is the sum of the variances.
- Therefore, the variance of the received sample Y_k equals $P + N_0W$.

Indian Institute of Technology, Delhi 11 Ranjan Bose
Department of Electrical Engineering

So, we go a little bit further if 2 independent Gaussian random variables are added; the variance of the resulting Gaussian random variable is the sum of the variances right because this is again the independence notion.

So, let us quickly look at what we are doing here. So, we have this nice Gaussian channel.

(Refer Slide Time: 14:45)



Where we have this X_k and we have our friend today N_k because without this N_k we do not have a case and we have your Y_k . Now we wanted to maximize the average mutual information in order to get the capacity; for that the conclusion was that this X_k must be Gaussian. This is Gaussian by definition and sum or difference of 2 Gaussians under variables as Gaussian. So, this guy is also Gaussian and why was this taken as Gaussian? Because we had this needs to be maximized right because capacity requires us to maximize this, this has to be maximized and this is Gaussian this is Gaussian. So, in order to maximize this Y should be Gaussian, this is Gaussian it forced us to make X Gaussian. So, that was the general train of thoughts.

Now, we do this independence thing further and we do that the power here at the input they have taken the average value to be P and this is sigma squared and this has a variance P ; then this is Gaussian sum of 2 Gaussians, but we will ensure that this variance should be equal to P plus sigma squared. So, noise power can be written as N naught W . And therefore, we can write P plus N or W very simple observation the beauty of Gaussian distribution right.

(Refer Slide Time: 17:25)


Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

- We know that if two independent Gaussian random variables are added, the variance of the resulting Gaussian random variable is the sum of the variances.
- Therefore, the variance of the received sample Y_k equals $P + N_0W$.
- It can be shown that the differential entropy of a Gaussian random variable with variance σ^2 is

$$\frac{1}{2} \log_2(2\pi e \sigma^2)$$

Therefore, $h(Y_k) = \frac{1}{2} \log_2[2\pi e(P + N_0W)]$

 Indian Institute of Technology, Delhi 11 Ranjan Bose
Department of Electrical Engineering

And therefore, so we come back to the slide and say that it will be shown that the differential entropy of a Gaussian random variable with variance sigma squared is half log to the base 2 pi e sigma squared ok.

So, you can show this and what is very interesting is that if I give you 2 random variables with equal power; so, variance then it is a Gaussian random variable that maximizes the entropy; it is a very interesting part. So, what is the physical significance of this? suppose I make an observation some funny vibration is going on bridge or a measuring some traffic condition and I have no clue about what distribution to put in.

Then the best bit is to model it as Gaussian because Gaussian captures the maximum information. The entropy is maximum for a Gaussian random variable that is why we chose Gaussian to be the distribution for Y_k and hence X_k and what is that differential entropy where for the Gaussian it is $\frac{1}{2} \log_2(2\pi e \sigma^2)$.

So, for Y_k which is nothing but the sum of X_k and N_k the variances add up and I have half log to the base 2; $2\pi e$ and instead of sigma squared effective sigma square is a sum of the 2 variances. So, P is the variance of X_k and N_0W is the variance of noise which is the noise power fine. So, very simply no hard mathematics we are combining intuition with a little bit of mathematics to get this differential entropy for $h(Y_k)$.

Remember our job is to find out $h(Y_k)$ then $h(N_k)$ and find the difference and that difference should be the capacity of this Gaussian channel. So far so good we are trudging along.

(Refer Slide Time: 19:47)

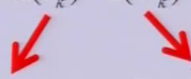
Information Theory, Coding and Cryptography


Capacity of Gaussian Channel

Therefore, $h(Y_k) = \frac{1}{2} \log_2 [2\pi e(P + N_0W)]$

And, $h(N_k) = \frac{1}{2} \log_2 [2\pi e(N_0W)]$

$C = h(Y_k) - h(N_k)$


 $C = \frac{1}{2} \log_2 [2\pi e(P + N_0W)] - \frac{1}{2} \log_2 [2\pi e(N_0W)]$

 Indian Institute of Technology,
Delhi

12

Ranjan Bose
Department of Electrical Engineering

So, what have we found so far? We have this deficient entropy of $h(Y_k)$ as follows and for the noise which is also Gaussian distributed or assumption $h(N_k)$ is half log to the base $2\pi e$ and N_0W the noise power.

Now, we come to finding the capacity which is the difference and we write it out explicitly. So, the expressions for $h(Y_k)$ and $h(N_k)$; so, we write it down as a difference.

(Refer Slide Time: 20:33)


Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

$$C = h(Y_k) - h(N_k)$$
$$C = \frac{1}{2} \log_2 [2\pi e(P + N_0W)] - \frac{1}{2} \log_2 [2\pi e(N_0W)]$$

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N_0W} \right) \text{ bits per channel use}$$

13

 Indian Institute of Technology, DelhiRanjan Bose
Department of Electrical Engineering

Now log difference of log is just tailor made for simplification and so, what we do is we take half log to the base 2 and then we take it common and we take the ratio. So, $2\pi e P$ plus N_0W divided by $2\pi e N_0W$. So, this $2\pi e$ cancels out and I am left with P plus N_0W divided by N_0W and we are left with finally, $1 + P/N_0W$ as argument of the log.

So, C comes out to be half log 2 $1 + P/N_0W$ and clearly I must put the units of this and it is bits per channel use ok. So, I am still uncomfortable because per use every time I use a channel, but suddenly I remember that I have to give an answer in terms of bits per second. I can always say every time you use the channel, but that does not really make much sense for the user. The user would like to know; what is the capacity in bits per second; especially when we have introduced this notion of power limited and bandwidth limited.

So, we have to take that last final step to change this bits per use bits per channel use to something like bits per second. So, we need to find out how many times per second can I really use this channel ok. So, this point is bothering me still and I need to make a fix for this.

(Refer Slide Time: 22:20)

Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N_0 W} \right) \quad \text{bits per channel use.}$$

- We are transmitting $2W$ samples per second, i.e., the channel is being used $2W$ times in one second.
- Therefore, the information capacity can be expressed as

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right) \quad \text{bits per second}$$

Indian Institute of Technology, Delhi 14 Ranjan Bose
Department of Electrical Engineering

So, what are we doing? We have so far derived that in terms of bits per channel use capacities half log to the base 2 $1 + \frac{P}{N_0 W}$ and since the log to the base basis 2 the answer is N bits. We are transmitting $2W$ samples per second because the bandwidth is W . So, we know that if we have a bandwidth of W , we can have in terms of the micro sampling theorem, we can use the channel being used $2W$ times in 1 second. So, then if we are using it fruitfully; then we can say that we have finally, the last piece in the jigsaw, I have the answer of times per second the channel can be used. So, bits per channel use in to use per second will give me bits per second.

So, finally, I multiply this by $2W$; so, I here it was half when I multiplied with $2W$ I am end up with W ; $W \log$ to the base 2 $1 + \frac{P}{N_0 W}$ and please note here we had bits per second and $2W$ samples per second. So, if you use these 2 together I end up with bits per second.

So, finally, I am quite comfortable with this notion of capacity which is $C = W \log$ to the base 2 $1 + \frac{P}{N_0 W}$ bits per second; this makes sense this I can go and sell my product, I can relate to bits per second. So, this is pretty much the capacity of a Gaussian channel.


(Refer Slide Time: 24:14)

Information Theory, Coding and Cryptography

Capacity of Gaussian Channel

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right) \text{ bits per second}$$

- This basic formula for the capacity of the band-limited, AWGN waveform channel with a band-limited and average power-limited input was first derived by Shannon in 1948.
- It is known as the Shannon's third theorem, the **Information Capacity Theorem**.

 Indian Institute of Technology, Delhi15Ranjan Bose
Department of Electrical Engineering

This basic formula even though we are saying its basic, we will very soon realize how versatile it is and how much insight it can provide. This basic formula for the capacity of the band limited power limited, additive wide Gaussian, waveform channel was first derived by Shannon in 48; 1948 and it is also known as Shannon's third theorem and also called the information capacity theorem.

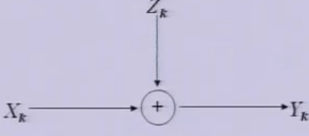
(Refer Slide Time: 24:50)

Information Theory, Coding and Cryptography


Information Capacity Theorem

- The information capacity of a **band-limited, power-limited** Gaussian channel can be expressed as

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right)$$



```
graph LR; Xk[Xk] --> Sum((+)); Zk[Zk] --> Sum; Sum --> Yk[Yk]
```

 Indian Institute of Technology, Delhi16Ranjan Bose
Department of Electrical Engineering

So, a quick recap of what we have done so far. So, we had this nice band limited, power limited channel X_k and the transmitter Z_k or N_k which ever notion you want to put in

is added and I get Y_k and the capacity we have derived as W locked the base to 1 plus P over N naught W . So, we have introduced the notion of band limited and power limited. Now, it is a very interesting formula let us see; what does it tell us. So, let us write it down.

(Refer Slide Time: 25:39)

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right)$$
$$C = W \log_2 (1 + \text{SNR})$$

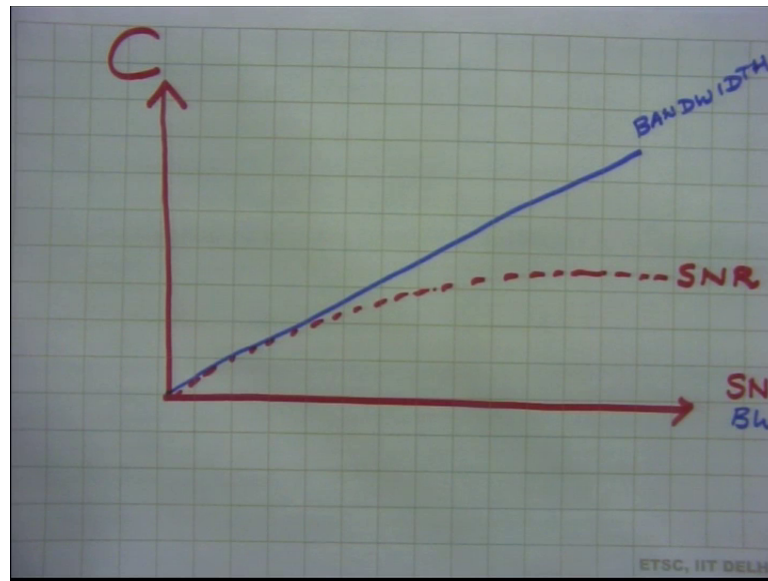
Linearly with W
Log. with SNR !!!

ETSC, IIT DEL

So, the first observation that we make is that this W is at 2 places at the same time N naught W is the noise power. So, this is effectively W log to the base 2 1 plus SNR. This is the signal power divided by the noise power, this is your SNR. It tells me that the capacity grows linearly with W , but logarithmically with SNR; this is an interesting observation.

The capacity does not treat bandwidth and power equally; in fact, the capacity is partial towards bandwidth and is less partial towards in terms of giving importance to SNR. So, if you just plot it.

(Refer Slide Time: 27:38)



And if you have this capacity on this axis you have SNR versus your bandwidth. So, it kind of grows linearly with bandwidth. On the other hand just I am superimposing to bring out the full import, it grows logarithmically with SNR. So, if I have money to put I will rather put my money on bandwidth because it gives me a much better return in terms of the capacity than the SNR.

In fact, if you look at the practical systems the 2 G wireless systems invested more in terms of SNR right, but then we made a transition from 2 G to 3 G; for example, CDMA, CDMA uses excess bandwidth and much lower power. So, it invested more in terms of bandwidth than power in order to get a better capacity. This is simple way to justify why CDMA was chosen over the FDMA, TDMA systems when we transited from 2 G to 3 G ok, but there are many other implications in terms of bandwidth and power tradeoffs.


So, other thing that tells me is that we can trade off bandwidth for power in terms of performance, but we will come to that shortly. So, we go back to our slide.

(Refer Slide Time: 29:45)

Information Theory, Coding and Cryptography

Information Capacity Theorem

- The **Information Capacity Theorem** is one of the important results in information theory.
- In a single formula one can see the trade off between the channel bandwidth, the average transmitted power and the noise power spectral density.
- Given the channel bandwidth and the SNR the channel capacity (bits/second) can be computed.
- This channel capacity is the fundamental limit on the rate of reliable communication for a power-limited, band-limited Gaussian channel.
- It should be kept in mind that in order to approach this limit, the transmitted signal must have statistical properties that are Gaussian in nature.
- Note that the terms channel capacity and information capacity have been used interchangeably.

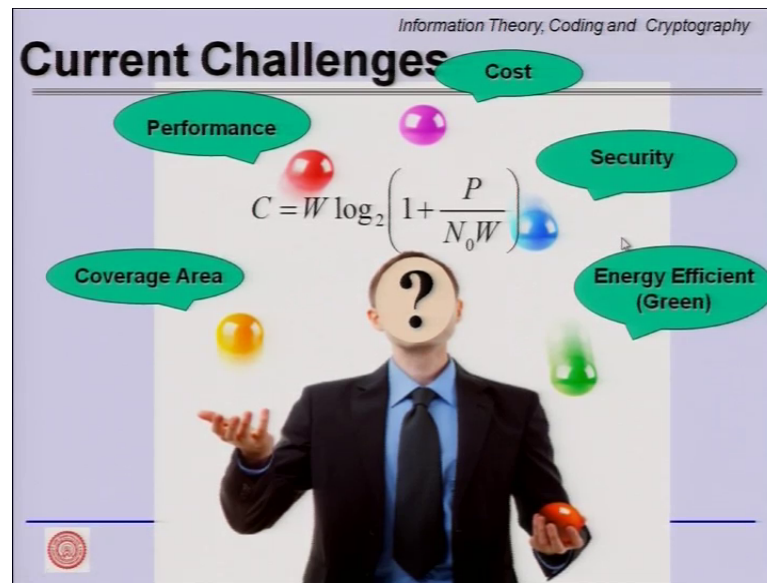
 Indian Institute of Technology,
Delhi17Ranjan Bose
Department of Electrical Engineering

And come to some of the interesting take home messages from the information capacity theorem. The information capacity theorem is one of the very important results in information theory. In a single formula one can see the tradeoff between the channel bandwidth the average transmitted power and the noise power spectral density is a one formula that links. Before this there was no clear idea as to how to look a tradeoff between say channel bandwidth or the average transmitted power or how the noise power relates to the capacity, right.

Given the channel bandwidth and the SNR the channel capacity can be computed. So, I can give you actual numbers this channel capacity is the fundamental limit on the rate of reliable communication for a power limited, bandwidth limited Gaussian channel. So, it gives a theoretical limit on the rate of reliable communication, but it should be kept in mind that will meet some assumptions.

And one of the most important assumption is that the transmitted signal must have statistical properties that are Gaussian in nature; this was not always true and therefore, this is only a theoretical limit. But in order to achieve the capacity X should tend towards Gaussian. Note that the return channel capacity and information capacity have been used interchangeably so far. So with, sometimes talking about channel capacity of the Gaussian channel, sometimes we are talking about the information capacity of the Gaussian channel; well the theorem is called the information capacity theorem.

(Refer Slide Time: 31:50)



So, let us look at a holistic view first; so, what are the current challenges? So, just for the sake of discussion let us see we are going to design a wireless communication system, the next generation may be 5 G and what are the challenges in front of us?.

So, the designer has to look at several things right from the coverage area to the performance in terms of bit error rate, frame error rate, delay jitter took the cost security aspects becoming very important and energy efficiency it has a fancy name today called green communications. So, these are all the things that a person must juggle so, as to get things right and sell the product in the market.

But how does this relate to our information capacity theorem? Well if you see already we have how to trade off the power versus bandwidth versus performance. So, one thing that we have not really made it very clear explicitly is how the performance gets related? And that we will shortly link to, but before that let us understand that these are the current challenges and the number of users you can support, what is a coverage you can provide, what is the performance you can give in terms of bit error rate to the different users, how much battery power will the user have to expend, how much bandwidth will you allocate to the user all of this gets captured by one single formula given by the information capacity theorem.

(Refer Slide Time: 33:44)


Information Theory, Coding and Cryptography

The Shannon Limit

- Consider a Gaussian channel that is limited both in power and bandwidth.
- We wish to explore the limits of a communication system under these constraints.
- Let us define an ideal system which can transmit data at a bit rate R_b which is equal to the capacity, C , of the channel, i.e., $R_b = C$. Suppose the energy per bit is E_b .
- Then the average transmitted power is

$$P = E_b R_b = E_b C.$$

$$\frac{C}{W} = \log_2 \left(1 + \frac{E_b C}{N_0 W} \right)$$

 Indian Institute of Technology,
Delhi19Ranjan Bose
Department of Electrical Engineering

Now, we look at a very very interesting thing called the Shannon limit. So, again this is the follow from the information capacity theorem, let us consider a Gaussian channel once again which is limited both in power and bandwidth. Now we wish to explore the limits of communication system under these constraints. So, let us define an ideal some system since we are looking at the limits which can transfer data rate data at bit rate R_b which is equal to the capacity.

Please recall capacity they have been finally, able to nail it down in terms of the units bits per second. And this capacity if you remember your source rate should be less than this capacity so, as to get reliable communication let us say we are running at capacity. So, R_b which is the data rate in terms of bits per second is equal to C fine.

So, this is whatever we are going to derive under this condition will be the limit; if you go even 0.1 bits per second above this. We enter into the red territory where reliable communication may not be guaranteed. Now we make another assumption which is the energy per bit now this is again a very practical thing we always deal with energy per bit if you have to compare apples to apples in terms of higher order modulation schemes.

So, the average transmitted power is nothing but E_b into R_b because E_b is energy per bit and R_b is bits per second. So, we get say joules per second which is watts in terms of the power and are we has already been put equal to C ; so, we have E_b into C . So, P is equal to E_b into C ok. So, now, if you recall the information capacitive theorem we had

the C equal to W. So, that W has been put on the left hand side now with C over W is equal to log to the base 2; 1 plus what do we have here? We had P over N naught W, but P we have substituted as E b into C.

So, now, the differently expressed information capacity theorem running at capacity is C over W equal to log to the base 2 1 plus E b over N naught C over W. Why am I doing this? I have got the C over W and C is equal to R b that is equivalently R b over W at 2 places and I have got E b over N naught as at 2 places.

(Refer Slide Time: 37:01)

$$C = R_b$$

$$\frac{C}{W} = \log_2 \left(1 + \frac{E_b}{N_0} \frac{C}{W} \right)$$

$$\frac{R_b}{W} = \log_2 \left(1 + \frac{E_b}{N_0} \frac{R_b}{W} \right)$$

$$\downarrow \quad \downarrow \quad \downarrow$$

$$x \quad \quad \quad x \quad y$$

$$xy = \log_2 (1 + xy)$$

$$y = \log_2 (1 + xy)$$

So, if we write it out we can express it as; please note C equal to R b in this case. So, it is equivalently R b over W equal to log to the base 2 1 plus R b W times E b, where N not. So, you can always write this quantity as X this quantity as Y and this quantity again is X. So, you have X is equal to oh thank you this is a Y. So, we have Y equal to log to the base 2 1 plus xy.

So, it will be if I write it neatly it becomes Y equal to log to the base 2 1 plus xy and this we should be able to plot. But we will have physical significance for both E b over N naught and R b over W. So, R b over W is the normalized rate and E b over N naught is kind of a SNR.

(Refer Slide Time: 38:53)


Information Theory, Coding and Cryptography

The Shannon Limit

- This equation can be re-written in the following form

$$\frac{E_b}{N_0} = \frac{2^{C/W} - 1}{C/W}$$

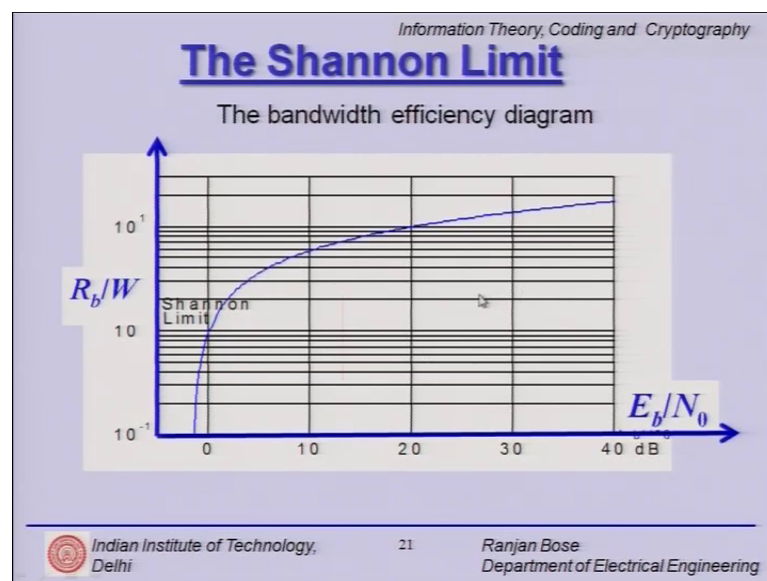
- The plot of the bandwidth efficiency $\frac{R_b}{W}$ versus $\frac{E_b}{N_0}$ is called the **Bandwidth Efficiency Diagram**

 Indian Institute of Technology, Delhi 20 Ranjan Bose
Department of Electrical Engineering

So, if you go back to the slide we can write now equation E_b/N_0 is equal to $2^{C/W} - 1$ over C/W ; this is if you want take 2 layers of power on both sides.

So, now we plot the bandwidth efficiency R_b/W versus E_b/N_0 . So, we have is R_b/W which we designated as Y ; versus E_b/N_0 which we designated X and whatever we get is called the bandwidth efficiency diagram.

(Refer Slide Time: 39:30)



So, let us mark this axis first; on the X axis we put E_b over N_0 this is what we put as X, on the Y axis we have R_b over W and if you make the plot you get this blue line as the Shannon's limit. This represents the condition that R_b is equal to C anything below it will be R_b less than C anything above it is R_b bigger than C . So, we will back to this diagram again.

(Refer Slide Time: 40:13)

Information Theory, Coding and Cryptography


The Shannon Limit

- For infinite bandwidth, the ratio $\frac{E_b}{N_0}$ tends to the limiting value

$$\left. \frac{E_b}{N_0} \right|_{W \rightarrow \infty} = \ln 2 = 0.693 = -1.6 \text{ dB.}$$

- This value is called the **Shannon Limit**.
- Note that the Shannon limit is a **fraction**.
- This implies that for very large bandwidths, reliable communication is possible even for the case when the signal power is **less than the noise power!**
- The channel capacity corresponding to this **limiting value** is

$$C \Big|_{W \rightarrow \infty} = \frac{P}{N_0} \log_2 e.$$



*Indian Institute of Technology,
Delhi*

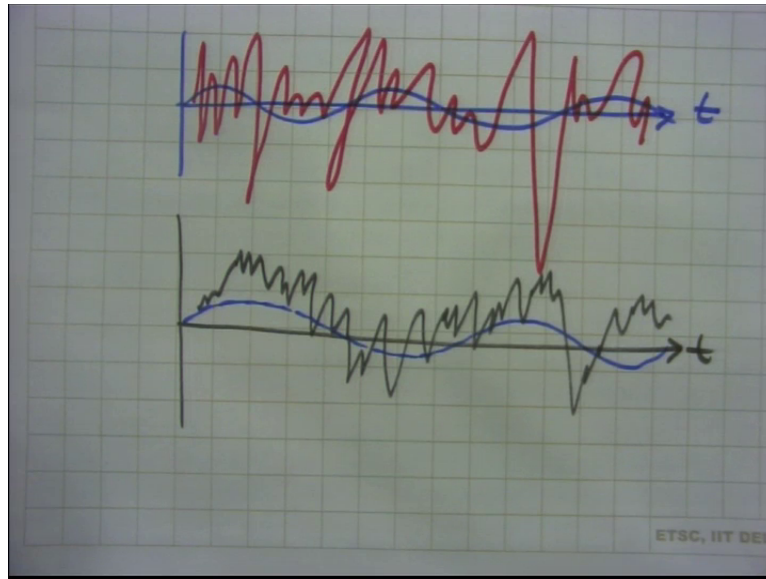
22

*Ranjan Bose
Department of Electrical Engineering*

But let us look at this limit those called the Shannon limit which we will come very quickly it is also a little counterintuitive. So, for infinite bandwidth; so, what do you mean by infinite bandwidth? W is the bandwidth, infinite means this R_b over W goes very low. So, I go down this limit. So, if you see if I go down this curve tends to SM totally become it tends to some value asymptotically this limit is the Shannon's limit that will talk about. So, for infinite bandwidth the ratio E_b over N_0 tells to a limiting value E_b over N_0 as W tends to infinity is nothing but $\ln 2$ is 0.693 in terms of dB it is minus 1.6 dB this is called the Shannon's limit.

Now, comes the counterintuitive part the Shannon's limit is a fraction what does it mean? This implies that for very large bandwidth reliable communication is possible even when the signal power is less than the noise power. Now if you do not understand the full import of it let us draw and tell you.

(Refer Slide Time: 41:43)



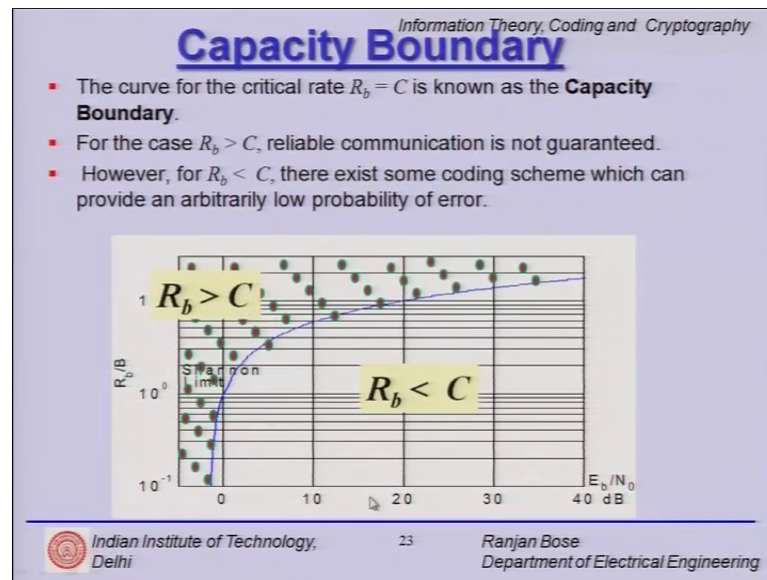
That suppose for a certain case; I have this as my signal and suppose I have this has noise; then when we add the signal and noise we get something like this. What Shannon's limit tells us is given infinite bandwidth I should be able to recover the signal from this seemingly junk data.

So, this shows that even if the noise power is larger than the signal power, if the SNR is minus 1.6 dB fine still I can get reliable communication and how reliable? Well you name it the trailer tends to understand you got it, but you give me infinite bandwidth; well how will I use that infinite bandwidth? I will use some kind of a stronger and stronger and stronger error control code to do this fine.

We come back to our slide the channel capacity corresponding to this limiting value right C with W tending to infinity. Well, what is it? It is P over N naught log to the base 2 e this is a constant what the capacity for infinite bandwidth is nothing but P over N not just decided only by the signal power and then noise power.

So, the capacity is completely determined by the signal to noise ratio for infinite bandwidth; I just keep does not keep growing. So, it is not that I keep increasing the bandwidth and my capacity will keep increasing; this is a very interesting result if you ponder over it. So, just if somebody gives a lot of bandwidth and keeps giving you more and more bandwidth it does not mean that you can keep increasing your capacity.

(Refer Slide Time: 44:28)



So, let us look at that Shannon's limit and this plot for this R_b equal to C condition once again. So, this is the capacity boundary; so, this blue line is called the capacity boundary. On the X axis again you have E_b over N_0 on the Y axis if R_b over W bandwidth here sometimes we sometimes you note bandwidth by b , where it is R_b over W .

And what you have is this region this red dots which is R_b greater than C . R_b greater than C means that reliable communication is not possible you cannot limit your error rate. So, it is all over this place right, but above this blue line. So, this is your R_b greater than C . On the other side is the region where we can design practical systems where R_b is less than C .


And again below this blue line all of this region right is your R_b less than C here we can have a reliable communication with arbitrarily low probability of error. And this blue line which is the capacity boundary, which separates this region of for practical communication and other region.

(Refer Slide Time: 45:57)

Information Theory, Coding and Cryptography

Capacity Boundary

- The curve for the critical rate $R_b = C$ is known as the **Capacity Boundary**.
- For the case $R_b > C$, reliable communication is not guaranteed.
- However, for $R_b < C$, there exist some coding scheme which can provide an arbitrarily low probability of error.
- The bandwidth efficiency diagram shows the trade-offs between the quantities $\frac{R_b}{W}$, $\frac{E_b}{N_0}$ and the probability of error, P_e .
- Note that for designing any communication system the basic design parameters are
 - the bandwidth available,
 - the SNR and
 - the bit error rate (BER).

 Indian Institute of Technology, Delhi 24 Ranjan Bose
Department of Electrical Engineering


So, basically the bandwidth efficient; so, diagram shows the tradeoff between the quantities R_b over W and E_b over N_0 and in terms of the probability of error. So, for designing any communication system; so the basic design parameters are the bandwidth available would you pay money for it, the SNR well that will decide the battery life and it is expensive also and the bit error rate. So these are the 3 things we must check.

(Refer Slide Time: 46:30)

Information Theory, Coding and Cryptography

Summary

- Channel Capacity
- Gaussian Channel
- Information Capacity Theorem
- Shannon Limit

 Indian Institute of Technology, Delhi 25 Ranjan Bose
Department of Electrical Engineering

So, let us now summarize what we have studied so far. We started off revisiting the channel capacity then we jumped directly to the Gaussian channel which is a practical channel as far reaching consequences. Then, we looked at the information capacity theorem we stated and derived the information capacity theorem. And we understood what do we mean by the Shannon limit.

That brings us to the end of this module.