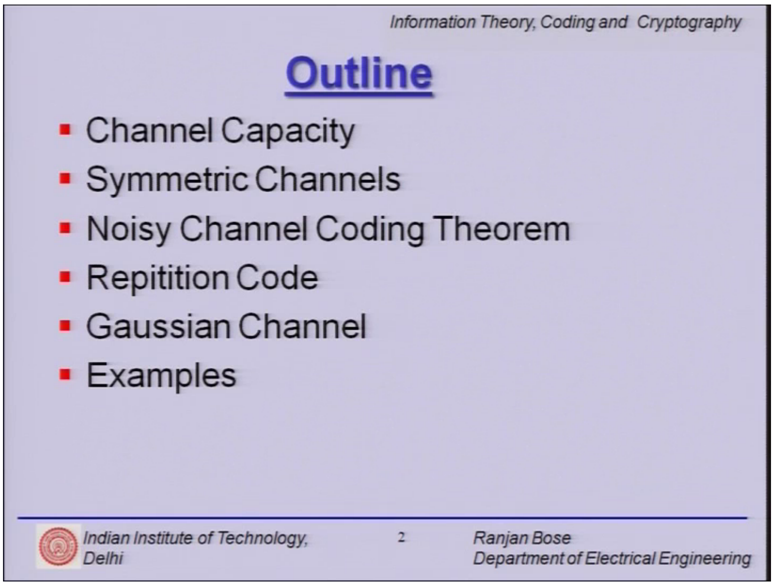
Information Theory, Coding and Cryptography
Dr. Ranjan Bose
Department of Electrical Engineering
Indian Institute of Technology, Delhi

Module - 10
Channel Capacity and Coding
Lecture – 10

(Refer Slide Time: 00:25)



Information Theory, Coding and Cryptography

Outline

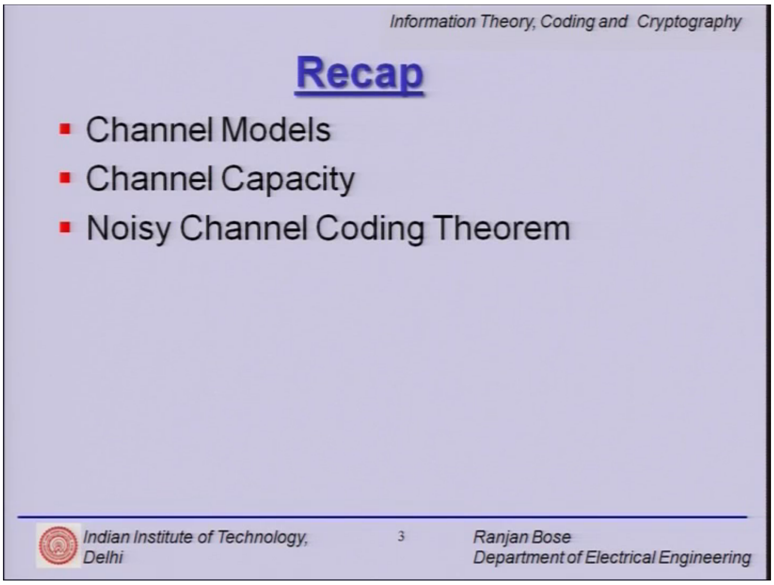
- Channel Capacity
- Symmetric Channels
- Noisy Channel Coding Theorem
- Repitition Code
- Gaussian Channel
- Examples

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

Hello and welcome to this lecture on Channel Capacity Encoding. Let us first start with the outline of today's talk. We will revisit this concept of channel capacity which we discussed in the previous lecture, then we will look at a class of channels called Symmetric Channels and also weakly symmetric channels and find out that it is quite easy to figure out the capacity of such symmetric channels.

Then we will revisit noisy channel coding theorem and we will look at a very simple example one of the simplest example of an error control code which is the Repetition Code. And finally, we will talk about Gaussian channels which are very practical and also mathematically tractable and we will sprinkle some examples here and there wherever required.

(Refer Slide Time: 01:19)



Information Theory, Coding and Cryptography

Recap

- Channel Models
- Channel Capacity
- Noisy Channel Coding Theorem

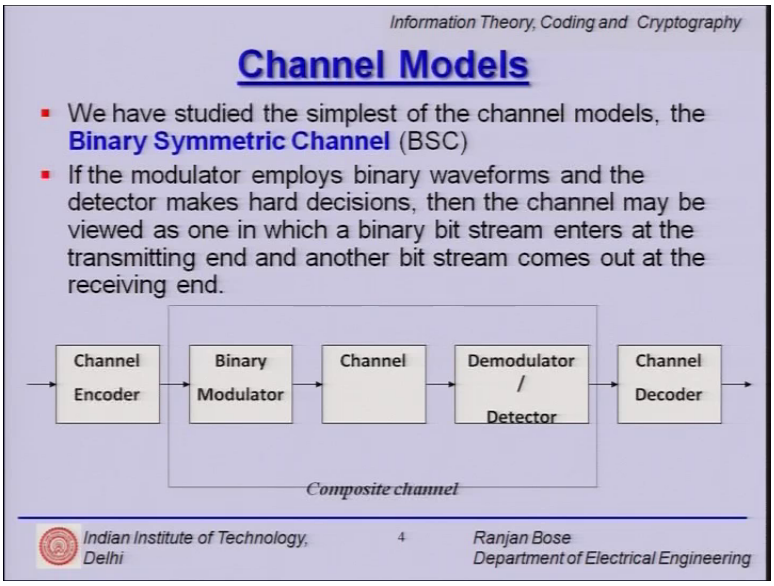
Indian Institute of Technology, Delhi

3

Ranjan Bose
Department of Electrical Engineering

So, let us start with a quick recap of what we did in the previous class. We started with a very basic channel model called the Binary Symmetric Channel. Even though it is simple, we can have several practical channels represented as binary symmetric channel including for example, the satellite channel.

(Refer Slide Time: 01:27)



Information Theory, Coding and Cryptography

Channel Models

- We have studied the simplest of the channel models, the **Binary Symmetric Channel** (BSC)
- If the modulator employs binary waveforms and the detector makes hard decisions, then the channel may be viewed as one in which a binary bit stream enters at the transmitting end and another bit stream comes out at the receiving end.

Channel Encoder → Binary Modulator → Channel → Demodulator / Detector → Channel Decoder

Composite channel

Indian Institute of Technology, Delhi

4

Ranjan Bose
Department of Electrical Engineering

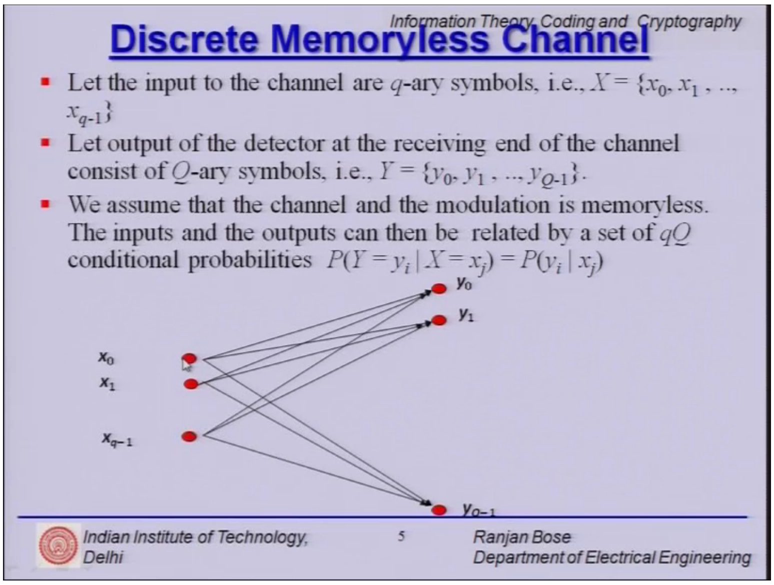
Now let us look at this structure. We always have a channel encoder whose job is to add redundancy in a known manner. So, as to overcome the ill effects of the channel, then we need to transmit the bit stream over the channel and therefore, we have a modulator. So, in this case we are writing it as a binary modulator which could be a general M-ary modulator. The job of this modulator is to take in bits and convert it into a waveform because we live in an analogue world, even though we love to deal with digital data.

The channel it could be any channel, the wireless channel, the underwater channel, the telephone line, the fiber optic channel whatever channel you choose would require the waveform to be transmitted over the channel. Now, the bits could be represented as amplitude or phase or frequency or a combination in terms of the wireless channel or even amplitude or wavelength for fiber optic channel.

So, you can have several ways to indicate whether you are transmitting a 1 or a 0. Of course, the inherent part of this channel is the noise and rest of this lecture will be dealing with reliable communication over this channel in the presence of noise when we say that the channel has noise we say it is an unreliable channel.

So, even though we are sending a 1 or a 0 or a 0 1 over this channel, what we receive; what the demodulator gives us is likely to be in error. Well depends on how good your luck is and how bad the channel is, but most of the 1‘s and 0’s will come out as a 1 and a 0, but some of them will be nearer, then the channel decoder starts a job. Its job is to ensure that it can correct from the errors and if it has been designed correctly it will recover from some of the errors. Today by the end of this lecture, we will look at a simple channel encoder and decoder a channel code called the repetition code.

(Refer Slide Time: 04:21)



Information Theory, Coding and Cryptography

Discrete Memoryless Channel

- Let the input to the channel are q-ary symbols, i.e., $X = \{x_0, x_1, \ldots, x_{q-1}\}$
- Let output of the detector at the receiving end of the channel consist of Q-ary symbols, i.e., $Y = \{y_0, y_1, \ldots, y_{Q-1}\}$.
- We assume that the channel and the modulation is memoryless. The inputs and the outputs can then be related by a set of qQ conditional probabilities $P(Y = y_i \mid X = x_j) = P(y_i \mid x_j)$

Indian Institute of Technology, Delhi

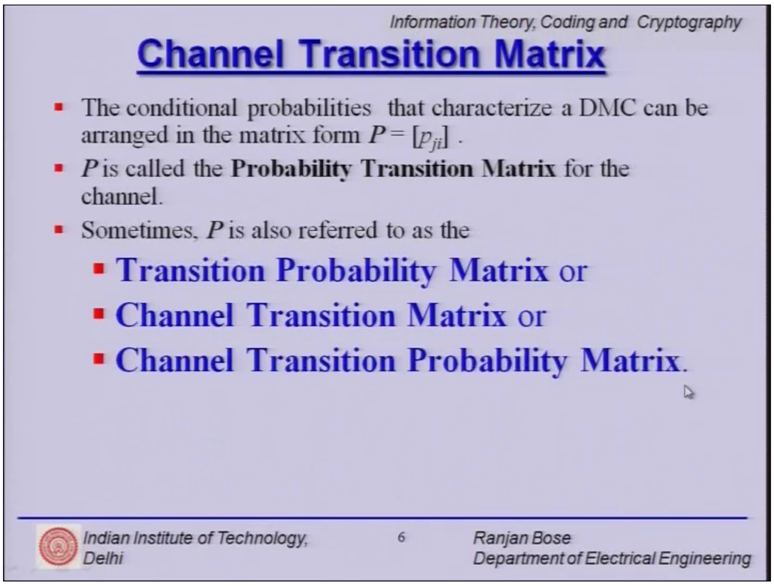
5

Ranjan Bose
Department of Electrical Engineering

Now, we looked at this Discrete Memoryless Channel where you can have 0 1 to up to small q minus 1 possible input symbols and output could be a y 0 y 1 up to y capital Q minus 1 output possible output symbols. Now what could they be? Well for me this could be frequency 1, frequency 2, up to frequency q minus 1 and output could also be frequencies or voltages or intensities.

So, it is for you to choose, but there are distinct symbols at the input and a output. The input and output are linked you can see there is a an arrow connecting each one of the nodes, at the input to the node at the output and we can write a transition probability right. So, it tells you what is the probability that when x 0 was transmitted it landed up as y 0, but it has a probability of also showing up as y 1 or being decoded as y Q minus 1. So, it will be indicated by this probability matrix.

(Refer Slide Time: 05:36)



Information Theory, Coding and Cryptography

Channel Transition Matrix

- The conditional probabilities that characterize a DMC can be arranged in the matrix form $P = [p_{ji}]$.
- P is called the **Probability Transition Matrix** for the channel.
- Sometimes, P is also referred to as the
  - **Transition Probability Matrix** or
  - **Channel Transition Matrix** or
  - **Channel Transition Probability Matrix**.

Indian Institute of Technology, Delhi

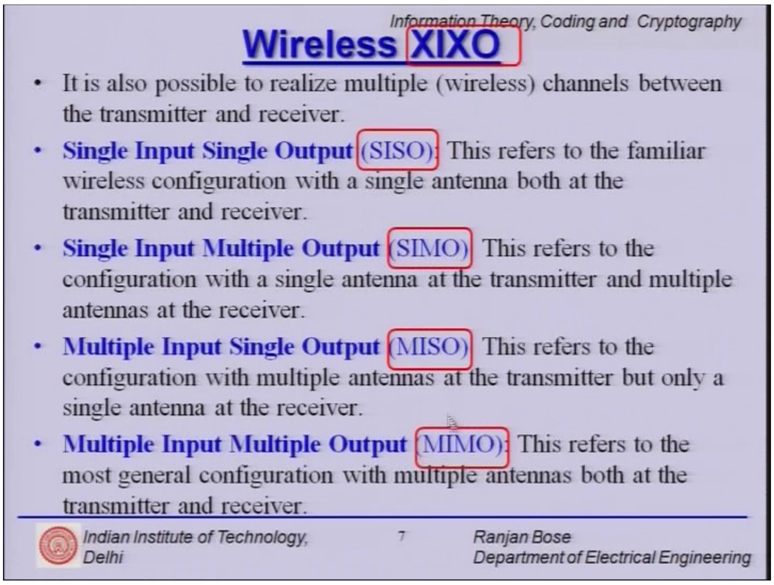
6

Ranjan Bose
Department of Electrical Engineering

So, let us spend a minute on this Channel Transition Matrix. We will learn that it has several names, but the conditional probabilities that characterize a discrete memoryless channel can be arranged in a matrix form.

So, this capital P is a set of p ji's and it connects the input to the output P is called the probability transition matrix and it is sometimes also referred to by different names like transition probability matrix or channel transition matrix or channel transition probability matrix depending on how a bit you are about writing the whole thing. You can use one of these three of tours of this channel transition matrix.

(Refer Slide Time: 06:27)



Information Theory, Coding and Cryptography

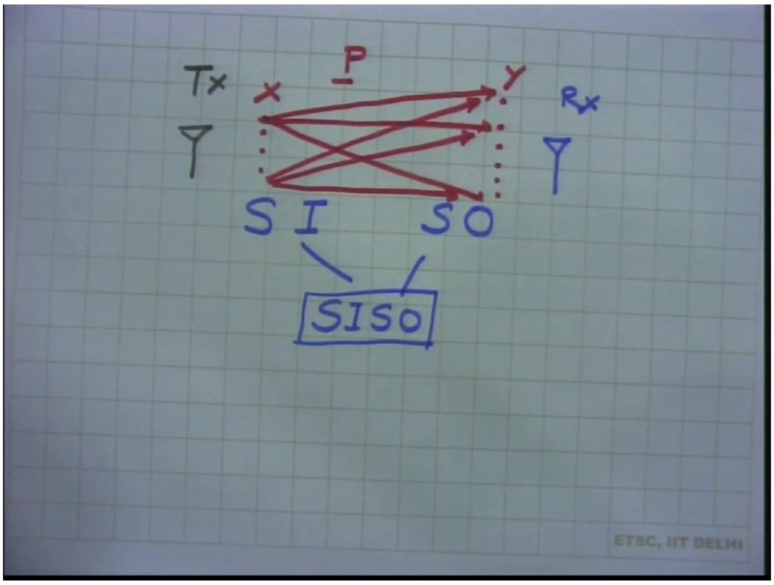
Wireless XIXO

- It is also possible to realize multiple (wireless) channels between the transmitter and receiver.
- **Single Input Single Output** (SISO): This refers to the familiar wireless configuration with a single antenna both at the transmitter and receiver.
- **Single Input Multiple Output** (SIMO): This refers to the configuration with a single antenna at the transmitter and multiple antennas at the receiver.
- **Multiple Input Single Output** (MISO): This refers to the configuration with multiple antennas at the transmitter but only a single antenna at the receiver.
- **Multiple Input Multiple Output** (MIMO): This refers to the most general configuration with multiple antennas both at the transmitter and receiver.

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

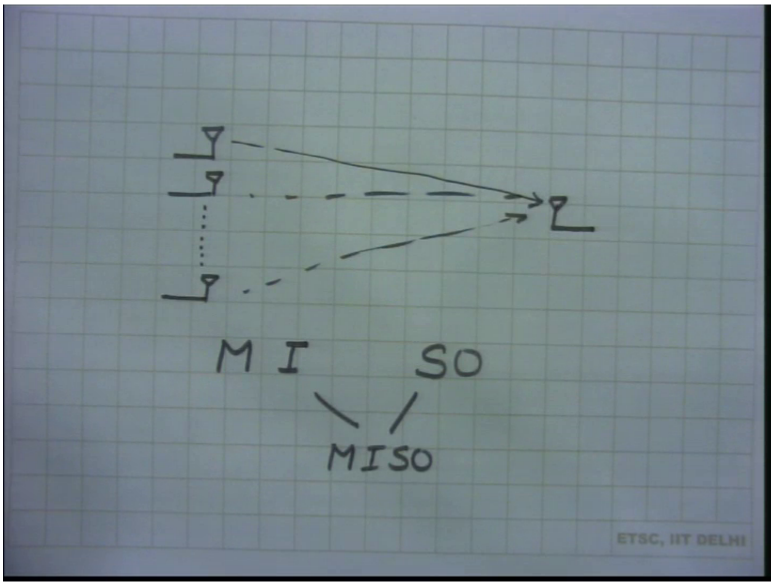
Now, just for a wireless channels, let us spend some time looking at the various channel types. Please note that this is a jump from what we were doing. Here we have four varieties like the single input, single output channel. So, here the input and output simply means that the number of antenna elements we have at the transmitter and the receiver side. Within the SISO I can have a channel transition probability matrix. So, let us look at this graphically.

(Refer Slide Time: 07:16)



Suppose we have an antenna at the transmitter and we have an antenna at the receiver and it is a wireless channel. So, we have single input, single output a SISO channel. Now within this channel, I can choose to have certain number of possible inputs X and possible number of inputs Y and they are connected together and I can have an equivalent channel probability matrix that links the input to the output for a SISO channel. But wireless gives us possibilities.

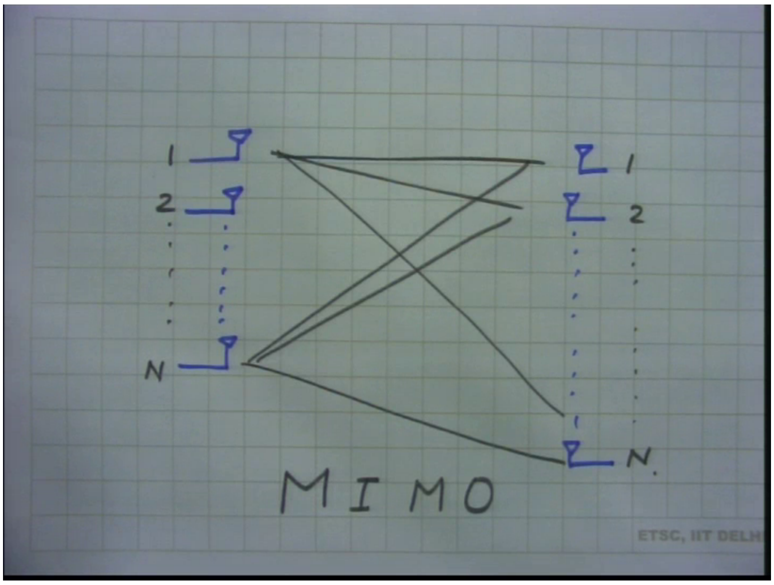
(Refer Slide Time: 08:31)



So, we can be imaginative and suppose we put more than one antenna element at the transmitter. So, if I have multiple input, but a single receiver, single output I make it as MISO and this kind of a configuration is possible when we have suppose a small form factor at the receiver. For example, this could be my handheld mobile phone and this is my base station. I can have several antennas available there.

Now, again for this pseudo channel, we can have our transition probability matrix available for this and so and so forth for this and so and so forth for this. At the same time, we can choose to have a much larger number of transmit antennas and also much larger number of receive antennas and I can have this as 1, 2 up to N, 1, 2 up to N and they are also linked. So, this is multiple input, multiple output MIMO systems.

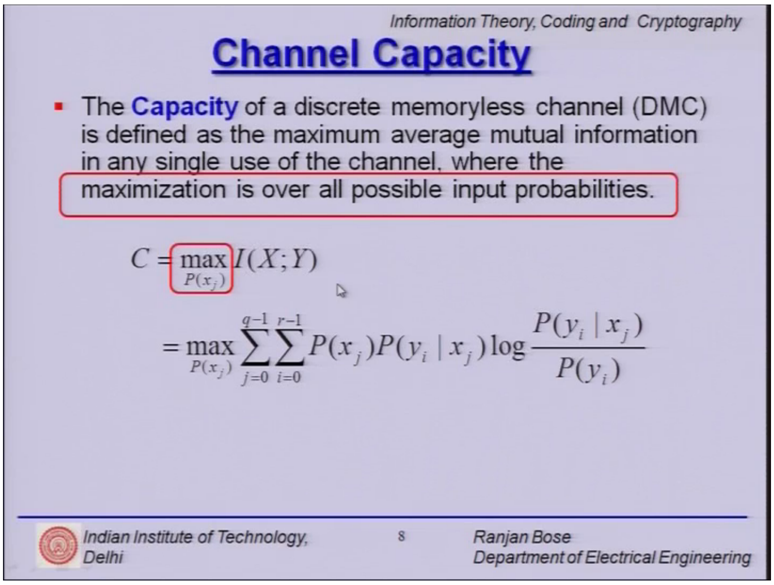
(Refer Slide Time: 09:40)



The reason why these are very popular in the current wireless standards is because they give us diversity gain. Wireless channels are inherently hostile and it is possible that one of or many of these channels are not good, but some of the other ones will be good because they would be inherently uncorrelated and thereby we can use different diversity combining techniques to get diversity gains.

But we will move back into our information theoretic concept and if we look at these slides again, I have put together this SISO, SIMO, MISO and MIMO, four possible combinations. But please note these are generally waveform channels. So, I have put the title as XIXO standing for these four combinations.

(Refer Slide Time: 11:23)



Information Theory, Coding and Cryptography

Channel Capacity

- The **Capacity** of a discrete memoryless channel (DMC) is defined as the maximum average mutual information in any single use of the channel, where the maximization is over all possible input probabilities.

$$C = \boxed{\max_{P(x_j)}} I(X;Y)$$

$$= \max_{P(x_j)} \sum_{j=0}^{q-1} \sum_{i=0}^{r-1} P(x_j)P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)}$$

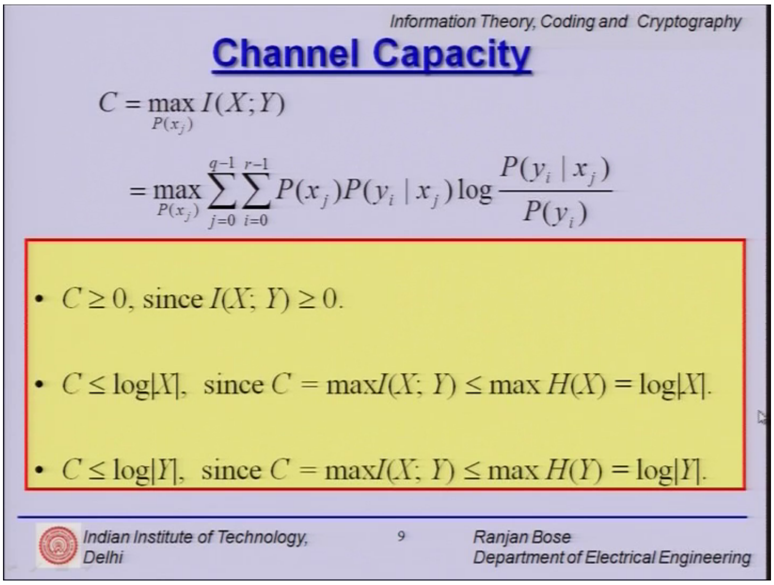
Indian Institute of Technology, Delhi 8 Ranjan Bose Department of Electrical Engineering

Now we come back to channel capacity. So, the capacity of a discrete memoryless channel is defined as the maximum average mutual information, in any single use of the channel where the maximization is over all input probabilities. So, if you look at C which is the capacity, I maximize it over all possible input probabilities and this is the average mutual information and these expressions are easily obtained from the channel transition probability matrix.

What it tells us is having observed Y at the output, what can you really say about X being transmitted. Now, that answer would depend on the input probabilities why? Because suppose the channel treats 0 differently than 1, it makes more errors for 0 versus 1, then the probabilities have to be adjusted to get the maximum out of this channel wherever the channel is weak.

Suppose it does not treat 0 very well, it makes lots of errors. I should have fewer number of 0's and if the channel treats 1 very well, it rarely makes a mistake for 1. I should have more number of 1's in the input. So, I can optimize the input distribution to get the maximum average mutual information over this channel.

(Refer Slide Time: 12:58)



Information Theory, Coding and Cryptography

Channel Capacity

$$C = \max_{P(x_j)} I(X;Y)$$

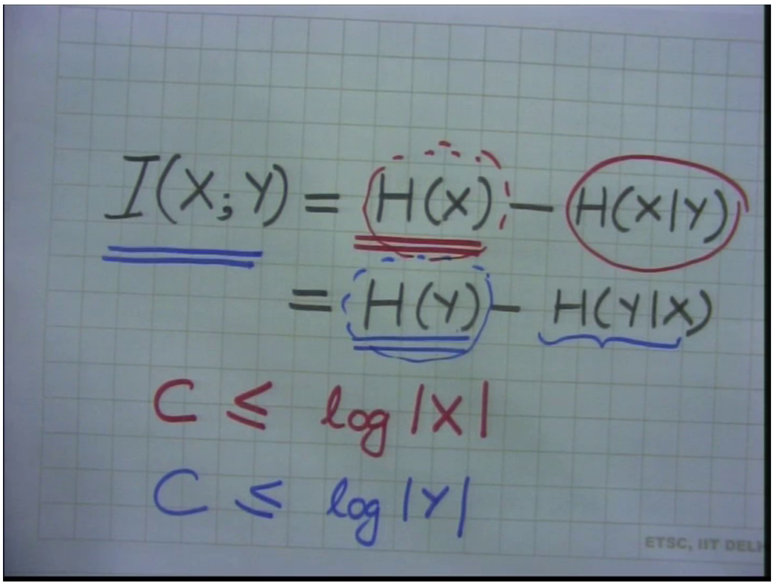
$$= \max_{P(x_j)} \sum_{j=0}^{q-1}\sum_{i=0}^{r-1} P(x_j)P(y_i \mid x_j)\log\frac{P(y_i \mid x_j)}{P(y_i)}$$

- $C \geq 0$, 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, Delhi 9 Ranjan Bose Department of Electrical Engineering

So, let us look at some properties. The first one is obvious. It is the maximum of the average mutual information. Now, since average mutual information is always greater than or equal to 0; then capacity must necessarily be greater than or equal to 0. This is a relief because every time I use the channel, the units will bits per use and worst the channel is broken and it does not communicate any information. So, it is 0. But even if the channel is doing any meaningful job, my capacity would be greater than 0. But it can never be negative. It does not make sense ok.

(Refer Slide Time: 14:03)



$$\underline{\underline{I(X;Y)}} = \underline{\underline{H(X)}} - \boxed{H(X|Y)}$$

$$= \underline{\underline{H(Y)}} - \underbrace{H(Y|X)}$$

$$C \leq \log|X|$$

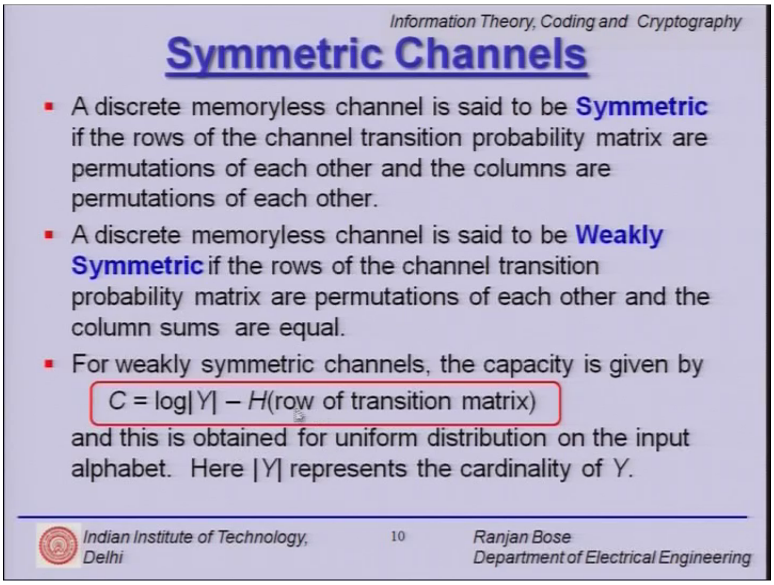
$$C \leq \log|Y|$$

Now if you remember, this average mutual information can be written as I X semicolon Y is equal to H of X minus H of X given Y. So, if you recall, alternatively you can also write as H of Y minus H of Y given X. So, clearly, you can conclude that this is the minus part. So, if you look at it your C should be less than or equal to log of cardinality of X from here right because this is the maximum I can have for H of X.

So, mutual information at best can reach the maximum attainable value for H of X and this happens when X has equiprobable outcomes and at that case the entropy of X becomes log to the base 2, the cardinality of X. So, it necessarily must be less than or equal to log cardinality of X. At the same time, I can have C less than or equal to log Y. This is the same logic at worst, I will subtract something out of it. If these are independent, then I get the 0 and the maximum attainable value is H of Y and when outcome are equi probable, then I get log cardinality of Y.

So, in any case your average mutual information is a upper bounded by these two values. So, this is what we see in this slide and this property of channel capacity. So, there is a sanity check, if you have done a lengthy calculation and quickly you are getting a value which is not consistent. It is greater than log of cardinality of X or log of cardinality of Y. You immediately must check your answer.

(Refer Slide Time: 16:17)



Information Theory, Coding and Cryptography

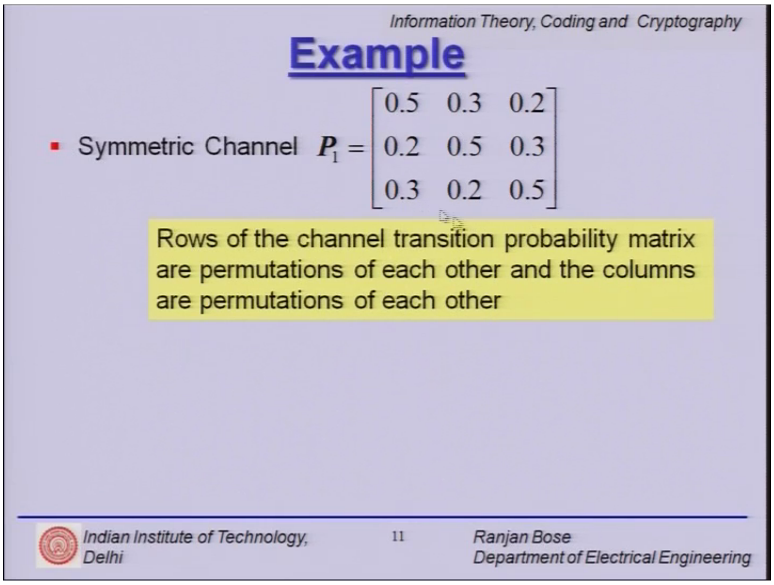
Symmetric Channels

- A discrete memoryless channel is said to be **Symmetric** if the rows of the channel transition probability matrix are permutations of each other and the columns are permutations of each other.
- A discrete memoryless channel is said to be **Weakly Symmetric** if the rows of the channel transition probability matrix are permutations of each other and the column sums are equal.
- For weakly symmetric channels, the capacity is given by $\boxed{C = \log|Y| - H(\text{row of transition matrix})}$ and this is obtained for uniform distribution on the input alphabet. Here $|Y|$ represents the cardinality of Y.

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

Now, we look at some special cases, some very friendly channels. So, we talk about symmetric channel and weakly symmetric channels in this slide. So, a discrete memoryless channel is said to be symmetric, if the rows of the channel transition probability matrix are permutations of each other and so, are the columns. Whereas a weekly symmetric channel is such that the rows of the channel probability matrix are permutation while the column sums are equal. Why is this important? Because for weakly symmetric channels; the capacity can simply be written as log of cardinality of Y minus H of row of transition matrix. This is actually you can if you look a little hard is, this is H Y given X that is the answer to this one.

(Refer Slide Time: 17:12)



Information Theory, Coding and Cryptography

Example

- Symmetric Channel $P_1 = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{bmatrix}$

Rows of the channel transition probability matrix are permutations of each other and the columns are permutations of each other

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

So, let us look at a simple example. Consider this channel which we hope is symmetric, but we can always test it out. So, what is the test? Well the test is that a symmetric channel, the rows of the channel transition probability matrix is permutations. So, if you look at row 1, row 2 and row 3; they are also permutations of each other. If you look at the columns; column 1, column 2, column 3, there are also permutations of each other. So, rows of the channel probability matrix are permutations and so other call. So, clearly this is a symmetric channel.

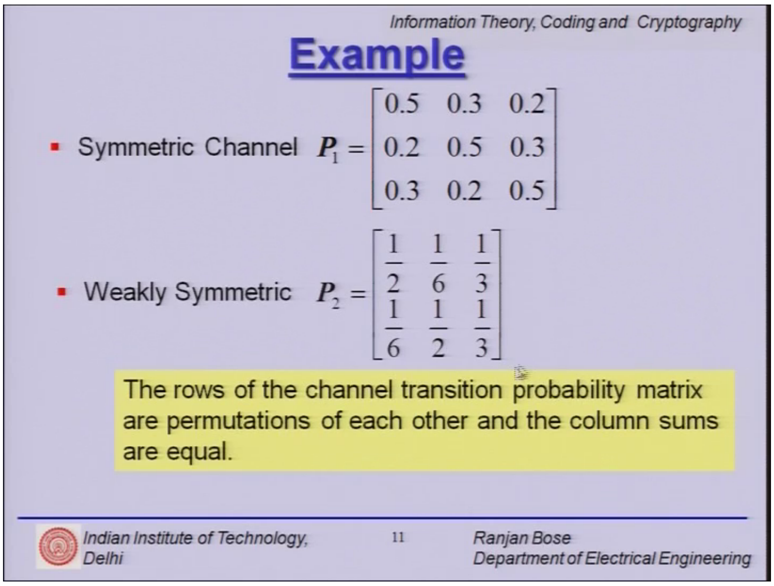
Student: Sir unique permutations or of the different (Refer Time: 17:54).

What is the question? The question is whether they are unique permutation, they are just permutations; so, you just standard 0.5, 0.2, 0.3, 0.3, 0.2, 0.5 and so and so forth.

Student: Each row (Refer Time: 18:09) to 1.

Yes.

(Refer Slide Time: 18:14)



Information Theory, Coding and Cryptography

Example

- Symmetric Channel $P_1 = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{bmatrix}$
- Weakly Symmetric $P_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$

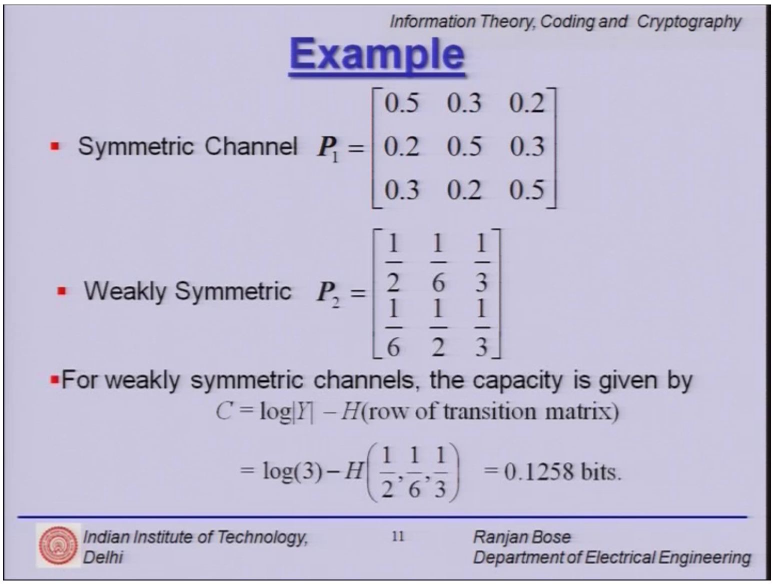
The rows of the channel transition probability matrix are permutations of each other and the column sums are equal.

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

Now, let us look at another simple example. This is an example for a weakly symmetric channel. So, here what is interesting is that the input, there are 2 input symbols and there are 3 possible output symbols and they are connected with this channel transition values and if you look at it, the rows of the channel transition probability matrix of permutations. So, if you see half 1 by 6, 1 by 3; this 1 by 6, half 1 by 3 and I can have several such permutations possible and this is the step 1. Check 1 is satisfied the rows of the transition probability matrix are permutations. But do the columns add up to some same value.

So, 1 by 2 plus 1 by 6 is 2 by 3, 1 by 6 plus 1 by 2 is again 2 by 3, 1 by 3 plus 1 by 3 is again 2 by 3. So, the column sums are equal and hence we can easily write it as weakly symmetric channels.

(Refer Slide Time: 19:28)



Information Theory, Coding and Cryptography

Example

- Symmetric Channel $P_1 = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{bmatrix}$
- Weakly Symmetric $P_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$
- For weakly symmetric channels, the capacity is given by

$$C = \log|Y| - H(\text{row of transition matrix})$$

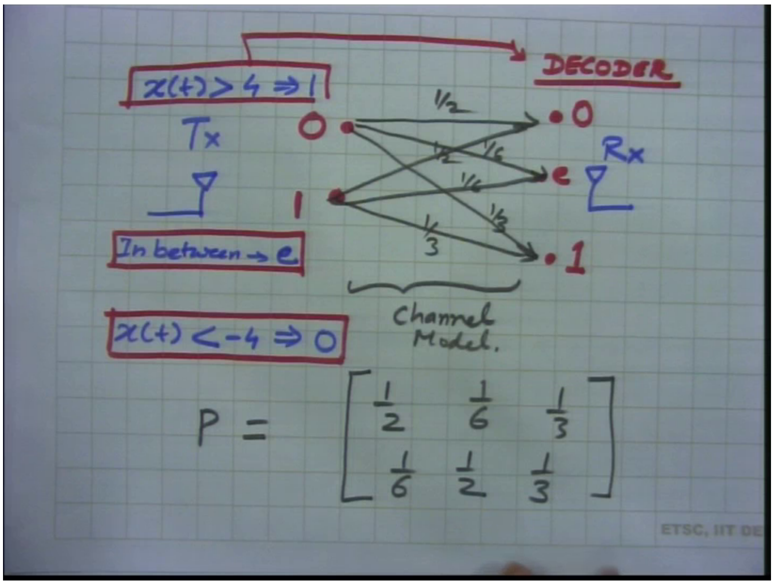
$$= \log(3) - H\left(\frac{1}{2}, \frac{1}{6}, \frac{1}{3}\right) = 0.1258 \text{ bits.}$$

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

Student: Sir, but how the transition matrix is found biasing the channel or input output code or (Refer Time: 19:34).

So the question being asked is how is a channel transition probability matrix found? How do we ascertain it? It is a pretty simple way to do it. We can do it experimentally or by simulations. So, let us look at this example. Suppose I have a channel it could be a wireless channel so that we can all relate to it.

(Refer Slide Time: 19:58)



$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

So, there is a transmitter and then there is a receiver and we say that if the amplitude is if amplitude of x t is greater than 4 volts, we say it 1 ; if x t less than minus 4 volts, we say a 0 and suppose it is anywhere in between we say it is an erasure fine.

So, this is my rule. So, I carry out lot of experiments. So, I say randomly I send a million 1's. What is a 1? It is a waveform with an amplitude greater than 4 or 5 whatever I wish to transmit. And at the decoder, I take a call and I do a count so, for every 1 that I sent. So, I am transmitting a 0 or a 1, but what I receive is a 0 erasure or a 1 depending upon these 3 conditions. So, this is the decoder side. The decoder we will use 1 of these 3 rules to say and measure.

So, I know what pseudo random bit sequence. I am sending at the transmitter and I find out what is the decision it is taking. So, it is possible that half the time when I sent a 0, it is actually received as a 0. So, I write half, but what happens to the other half of the time right? So, well I say that one-third of the time when I send a 0, it lands up at 1. How does it know? Well the transmitter knows what it is sending, the receiver figures out what it decoded, the compared notes and then they calculate on an average one-third of the time, it lands up as a 1. What was sent? A 0.

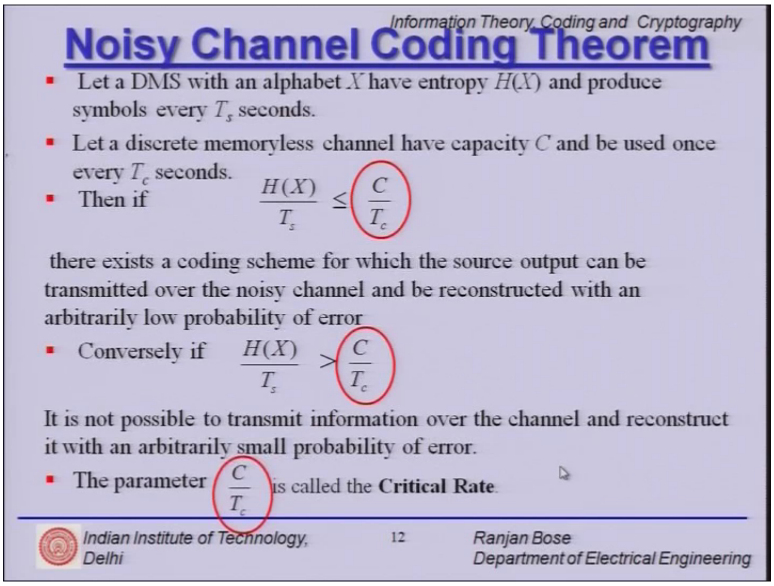
So, we get 1 by 3 as the transition probability 0 to 1. What happens the rest of the 1 6'Th of the time. Well it is an erasure. So, I write 1 by 6 here and I can do the same experiment with my 1. So, I transmit hundreds of thousands of 1,s and a see how many times actually 1 is received over this channel and maybe it is 1 by 3, 1 by 2, 1 by 6 to be anything and then I write down my channel transition probability matrix.

So, I can do it experimentally or I can make a MATLAB program for my channel model. So, I can plug in my channel model. It could be a Rayleigh fading model or a recent model or a simple Gaussian model with additive white Gaussian noise and I can conduct this experiment and see what are the numbers I get. So, getting this channel transition probability matrix is easy, I can do it by measurement or by simulations. And needless to say if I make my channel more hostile, if it is a poor channel, I plug in more noise; I will start making more errors and this transition probability matrix values will change.

So, coming back to our example, we observed that the second matrix is weakly symmetric and you already know what is the capacity for weakly symmetric channels and I am curious to find this out for this guy and for weakly symmetric channels, it is log cardinality of Y minus H row of transition matrix. What is H? H is like half log to the base 2 1 by 2 plus 1 by 6 log to the base 2 1 by 6 plus 1 by 3 log to the base 1 by 3 with a negative sign in. In the front and I will get a certain value. In this particular case it is point 0.1258 bits.

Physically what does it mean? It means that on an average having observed Y, we can only communicate 0.1258 bits of information over this channel. This is what is it worth I can again look at this and note that even if I make any transitions, so even if I juggle them; make a permutation of it, it is still remains weakly symmetric and nothing will change. So, I can make this as 1 by 6, 1 by 3, 1 by 2 and still the capacity will not change because this H of rows of transition matrix will not change. Because for the receiver, it does not matter whether I call this as a 0 erasure and 1 or 0 1 eraser; it does not matter. It is just 2 symbols being received as 3.

(Refer Slide Time: 26:25)



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 $\frac{H(X)}{T_s} \le \frac{C}{T_c}$

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 12 Ranjan Bose Department of Electrical Engineering

Now, let us look at this noisy channel coding theorem. So, let us have a discrete memoryless source with an alphabet X. So, alphabet is a set of symbols with entropy H of X. So, clearly we know the probabilities of occurrences of each of the symbols.

Now, we have introduced this parameter T s every second because we are tired of saying per use. So, we would like to bring in this time component and we say that this channel, the source is generating a symbol every T a second whereas, what we have is a channel with capacity C. Now we know what is c and this channel is used once every T c second.

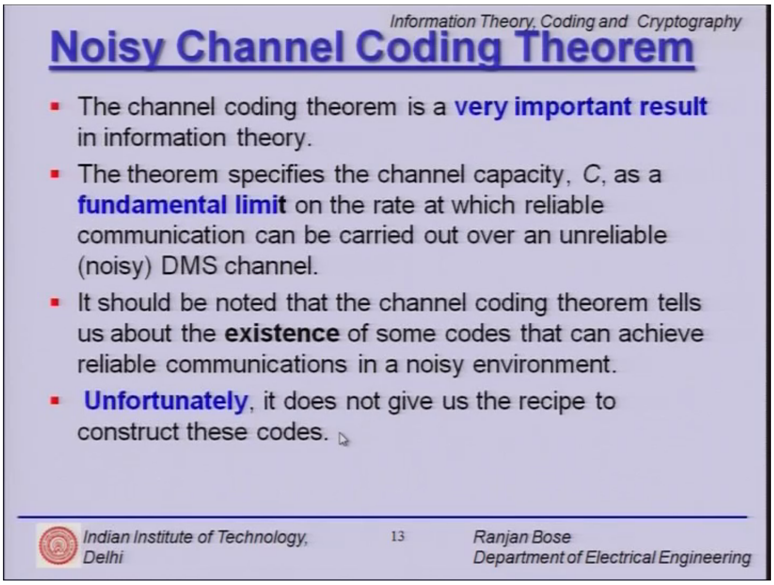
Then if H X over T s which is the source rate and C over T c is how frequently we are using the channel. So, capacity divided by the time, we use the channel; if we have the condition that C over T c is greater than H X over T s, then there exist a coding scheme. What is this coding scheme? Is an error control coding scheme. What is an error control coding scheme? I have a smart way of adding redundancy in a known manner.

Please recall in source coding, we were removing the redundancy; in channel coding, we are adding redundancy though in a known manner. We will have a mathematical structure. The noise does not know about this structure. So, noise at random flips bits and using that mathematical or the algebraic structure, I should be able to reconstruct the code that is in a Laymans language, what a coding scheme does. Coding scheme is error control coding scheme sometimes. This is also called as channel coding scheme because the errors are introduced by the channel.

So, coming back to the noisy channel coding theorem, it says that if the source rate H of X divided by T of s is less than this magical quantity C over T c, then there exists a coding scheme for which the output can be transmitted over this unreliable noisy channel and can be reconstructed with an arbitrarily low probability of error. How low you define it, 10 raise power minus 10, 10 raise power minus 15 be my guest. You can go ahead and do that. Conversely if we have the source rate H of X divided by T s greater than C over T c, then sorry we cannot guarantee you much. You cannot really pin down, the probability of error.

So, this interesting parameter C over T c is called the critical rates. It is figures out here. So, if you are less than the critical rate, you are in business. On the other hand, if you exceed this critical rate sorry pretty hard because your probability of error cannot be pinned down.

(Refer Slide Time: 29:43)



Information Theory, Coding and Cryptography

Noisy Channel Coding Theorem

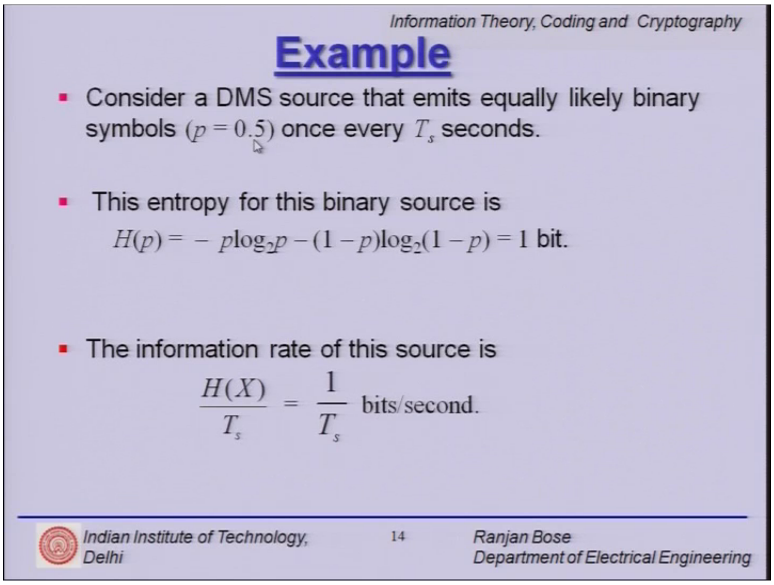
- The channel coding theorem is a **very important result** in information theory.
- The theorem specifies the channel capacity, *C*, as a **fundamental limit** on the rate at which reliable communication can be carried out over an unreliable (noisy) DMS channel.
- It should be noted that the channel coding theorem tells us about the **existence** of some codes that can achieve reliable communications in a noisy environment.
- **Unfortunately**, it does not give us the recipe to construct these codes.

Indian Institute of Technology, Delhi 13 Ranjan Bose Department of Electrical Engineering

So this is a very important result. It puts in a fundamental limit on the rate at which reliable communication is possible over an unreliable channel ok, but the sad part is that it tells us about the existence of these codes, but it does not give us a recipe to construct these codes.

So, we still have to look hard and find and keep a research going regarding finding a better and better codes. Since these codes, we do not have a formula, a method, a recipe to come up with this codes; those guys who come came up with very good codes got their names attached to it. So, we will see to it we will come to hamming code reed Solomon code, reed Solomon code, reed Muller code and several other code. So, all the inventors got the credit because there is no hard and fast rule to really construct it. There are some rules, but in general it has relied on the ingenuity of the inventor.

(Refer Slide Time: 30:50)



Information Theory, Coding and Cryptography

Example

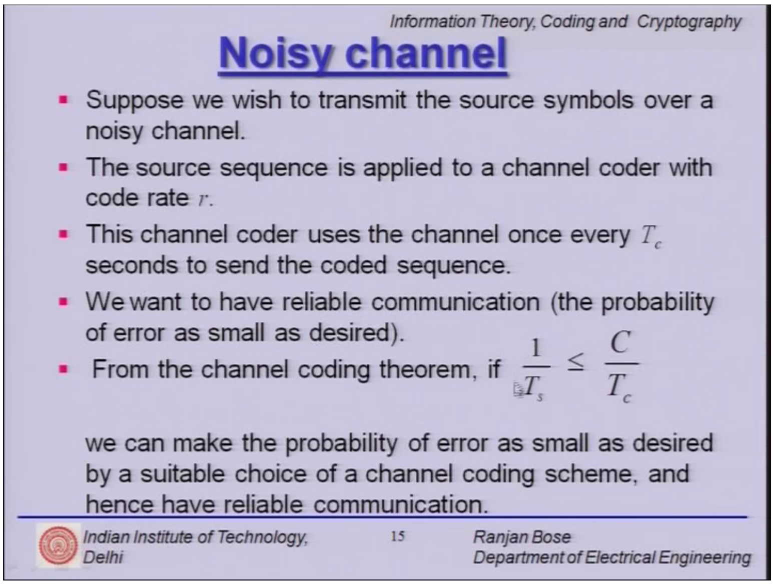
- Consider a DMS source that emits equally likely binary symbols ($p = 0.5$) once every T_s seconds.
- This entropy for this binary source is
  $H(p) = -\ p\log_2 p - (1-p)\log_2(1-p) = 1$ bit.
- The information rate of this source is
  $$\frac{H(X)}{T_s} = \frac{1}{T_s} \text{ bits/second.}$$

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

So, let us look at a simple example. Let us have a simple discrete memoryless source that has equally likely outputs. So, memoryless source it is a binary source. So, every T s seconds, it throws out a 1 or a 0. So, it could be my friend tossing a fair coin and every time he sees a head, he says 1; every time he sees a tail he says 0 and so and so forth. But he has been instructed to toss the coin and report the result every T s seconds.

Now, we know very well that the entropy of this binary source is 1 bit ok. It is the maximum possible value because both the outcomes are equally likely. So, the information rate of this source is simply entropy divided by T s is 1 over T s bits per second. So, this is the entropy, the source rate of this guy and we call this as the information rate.

(Refer Slide Time: 31:57)



Information Theory, Coding and Cryptography

Noisy channel

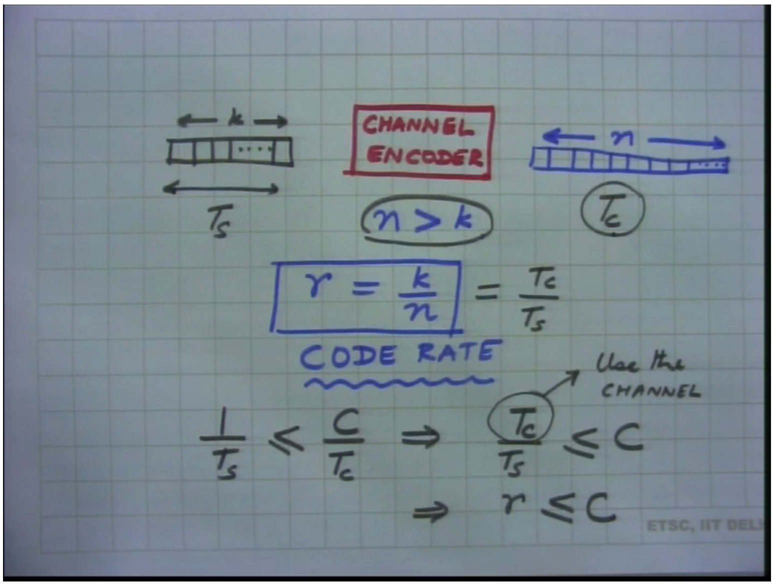
- Suppose we wish to transmit the source symbols over a noisy channel.
- The source sequence is applied to a channel coder with code rate r.
- This channel coder uses the channel once every T_c seconds to send the coded sequence.
- We want to have reliable communication (the probability of error as small as desired).
- From the channel coding theorem, if $\frac{1}{T_s} \le \frac{C}{T_c}$

we can make the probability of error as small as desired by a suitable choice of a channel coding scheme, and hence have reliable communication.

Indian Institute of Technology, Delhi 15 Ranjan Bose Department of Electrical Engineering

Now, this guy who is generating this 1,s and 0's my source wishes to transmit this over a noisy channel. The source sequence is applied to the channel code rate with code rate r. Now what is this code rate r?

(Refer Slide Time: 32:28)



$$\boxed{r = \frac{k}{n}} = \frac{T_c}{T_s}$$

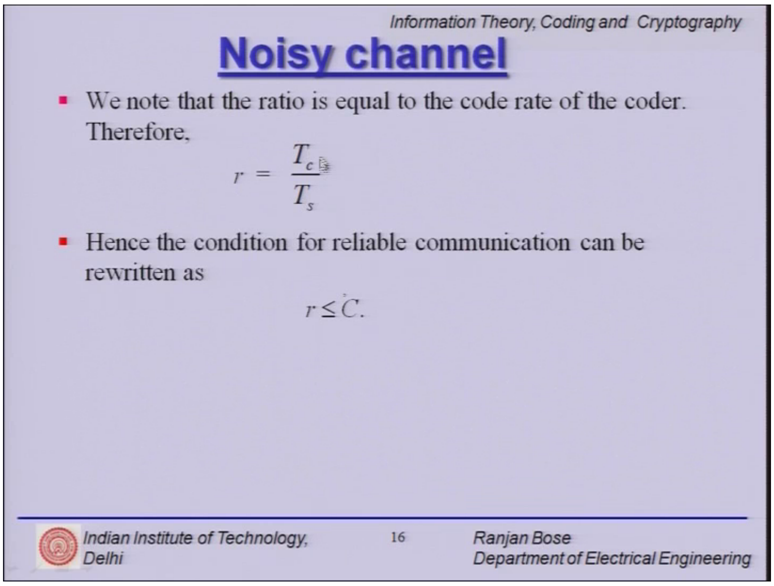
$$\frac{1}{T_s} \leq \frac{C}{T_c} \Rightarrow \frac{T_c}{T_s} \leq C$$

$$\Rightarrow r \leq C$$

Suppose we have certain k bits of data to be transmitted. So, suppose I have this k bits. So, my channel encoder will add n minus k redundant bits and converted into a longer bit stream. So, n is greater than k and this code rate r is equal to k over n. This is the code rate we are referring to. It will become obvious why we need to use this code rate shortly.

So, this channel encoder's job is to pad or introduce redundant bits. How many n minus k so, that it is able to recover from errors. Now let us go back to the slide and we see that the source sequence is applied to a channel coder with code rate r. This channel code is used to uses the channel every T c seconds to send the coded sequence. So, what we do is this channel coding theorem says that if I have to do reliable communication, then this source rate 1 over T s should be necessarily less than or equal to C over T c ok.

(Refer Slide Time: 34:20)



Information Theory, Coding and Cryptography

Noisy channel

- We note that the ratio is equal to the code rate of the coder. Therefore,

$$r = \frac{T_c}{T_s}$$

- Hence the condition for reliable communication can be rewritten as

$$r \le C.$$

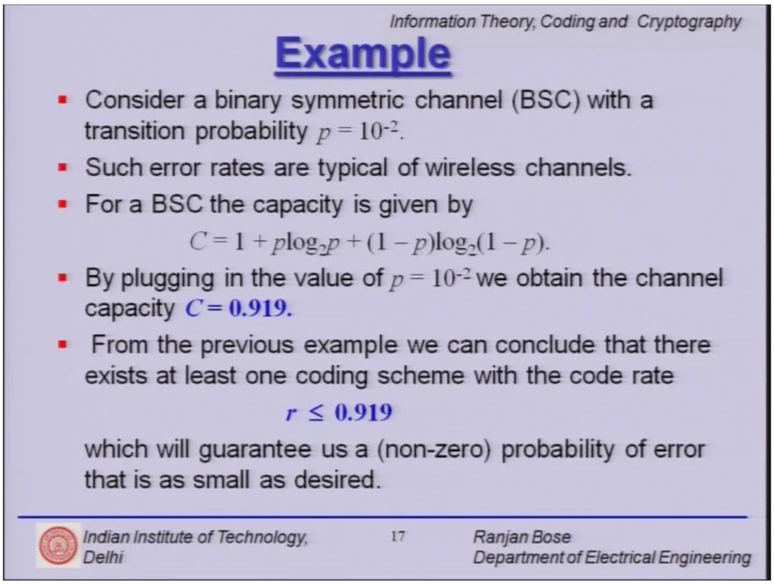
Indian Institute of Technology, Delhi 16 Ranjan Bose Department of Electrical Engineering

Now, C is the channel capacity I need to find out. But what do we mean by C over T c? If you juggle this, I get 1 over T s is necessarily less than or equal to c over T c for me to conduct communication reliably, but if you look at it, it means T c over T s should be less than or equal to C, but what is my T c? T c is the rate at which I use the channel.

Now clearly this T c should be less than T s, why is it so? Because my source is generating k bits but I have to transmit n bits n is greater than k. So, that I avoid the overflow, in the time that my source is generating k bits; I should be able to pump out n bits. But who decides, how fast am I using the channel? Well I am using the channel once every T c seconds and how fast is my friend generating those bits T s. So, the only way I can have this T c less than T s is, when I can pump out more number of bits, then what the channel encoder is getting.

Now what is this ratio? This ratio is nothing, but r because if you look at it right. So, I send k bits in T s and n bits in T c. So, every T c second, I use the channel right. So, this is nothing, but T c over T s. So, if you look at it, we can write r less than or equal to C. So, coming back to the slide, the condition for reliable communication can be written as r, the code rate r which is the ratio of before channel encoder to after channel encoder necessary less than 1, is less than equal to C.

(Refer Slide Time: 36:55)



Information Theory, Coding and Cryptography

Example

- Consider a binary symmetric channel (BSC) with a transition probability $p = 10^{-2}$.
- Such error rates are typical of wireless channels.
- For a BSC the capacity is given by

$$C = 1 + p\log_2 p + (1-p)\log_2(1-p).$$

- By plugging in the value of $p = 10^{-2}$ we obtain the channel capacity $C = 0.919$.
- From the previous example we can conclude that there exists at least one coding scheme with the code rate

$$r \leq 0.919$$

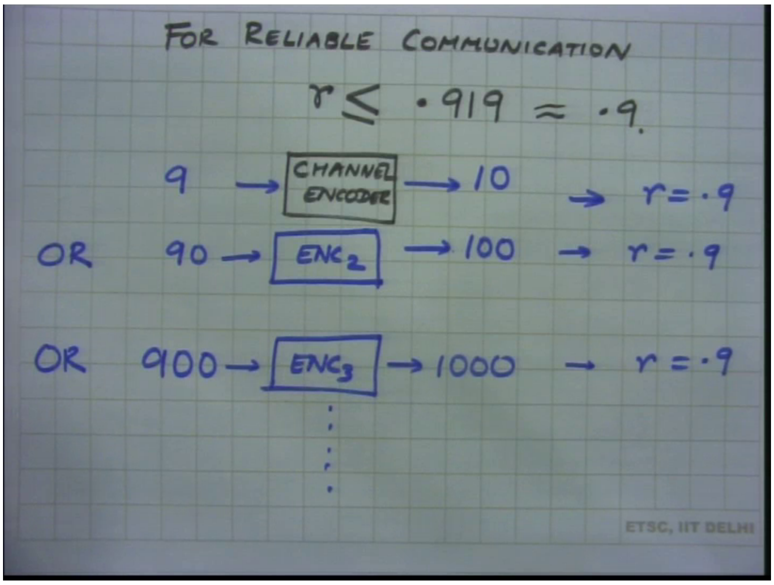
which will guarantee us a (non-zero) probability of error that is as small as desired.

Indian Institute of Technology, Delhi 17 Ranjan Bose Department of Electrical Engineering

So, we continue with this example. Consider this binary symmetric channel with probability of error 10 raise power minus p. So, one in every hundred bits gets flipped. So, this is a typical error rate of a wireless channel. So, for a semi binary symmetric channel, we have the capacity; we already calculated it as 1 minus h p and this is the binary entropy function with a negative sign and if you plug in the value of p is equal to 0.01, you get a capacity 0.919 ok. So, with this low probability of error, this channel capacity tells us that we can use this channel every single time and effectively transmit 0.919 bits.

So, we know that from the previous example, we conclude that this exists at least one coding scheme with rate less than 0.919 which will guarantee me arbitrarily low error rates, so for this simple channel ok. Now, how do we get this? What is the best way to pad or add redundancy to get this 0.919? Let us just understand the import of this.

(Refer Slide Time: 38:33)



FOR RELIABLE COMMUNICATION

$r \leq .919 \approx .9$

$9 \rightarrow \boxed{\text{CHANNEL ENCODER}} \rightarrow 10 \rightarrow r = .9$

OR $90 \rightarrow \boxed{\text{ENC}_2} \rightarrow 100 \rightarrow r = .9$

OR $900 \rightarrow \boxed{\text{ENC}_3} \rightarrow 1000 \rightarrow r = .9$

⋮

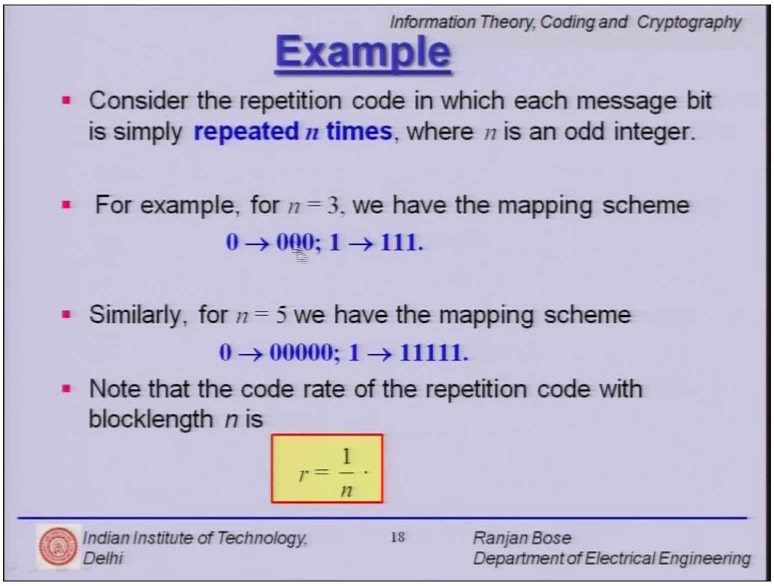
So, what we have seen is for reliable communication. You have seen that your r should be less than or equal to 0.919 ok. So, just for the sake of discussion, let us say its 0.9? What does this mean? It means that my friend, the channel encoder which is supposed to pad or introduce redundancy can take 9 bits and give out 10 bits, then I have r equal to 0.9. So, possibly it is adding a parity bit fine or it can take 90 bits and throughout 100 bits, again I meet the condition, but this is a more complex encoder because there are many more ways of adding 10 bits to 90 bits, then adding 1 bit to 90 bits or you can keep this argument going. I can take 900 bits at 1000 bits and make it 1000 bits.

Again my r is equal to 0.9. So, this is my encoder 2, encoder 3 and so and so forth. I do not have a recipe. Nobody tells me, how to add their bits? What is the best way to do things? This is my problem. I know all of this methods; there is a some way to add 1 bit to 9 bits or 10 bits to 90 bits or 100 bits to 900 bits and I will be able to get as low probability of error as possible, but how? Where to add? What is the best way to do it? Maybe this may not do or we have to figure things out. What is the best way to do it?

Student: (Refer Time: 40:56) because it was still we have to use the channel 1 (Refer Time: 40:59) T c seconds. So, it is like for 1 bit. You have to map with 1.1 widths and then send it.

Question being asked is where does the T c part figure into this? We assumed. So the assumption of T c was that we are going to, if you look at this diagram again; it says that once every T c seconds, but I can scale this ratios again. So, if I have to use 10 k bits as input and 10 n bits, then this T c is and T s both go down the 1 by time. It does not really make a difference. It is a ratio which is important fine. So, that is not the point.

(Refer Slide Time: 41:48)



Information Theory, Coding and Cryptography

Example

- Consider the repetition code in which each message bit is simply **repeated n times**, where n is an odd integer.
- For example, for $n = 3$, we have the mapping scheme

$$0 \to 000;\ 1 \to 111.$$

- Similarly, for $n = 5$ we have the mapping scheme

$$0 \to 00000;\ 1 \to 11111.$$

- Note that the code rate of the repetition code with blocklength *n* is

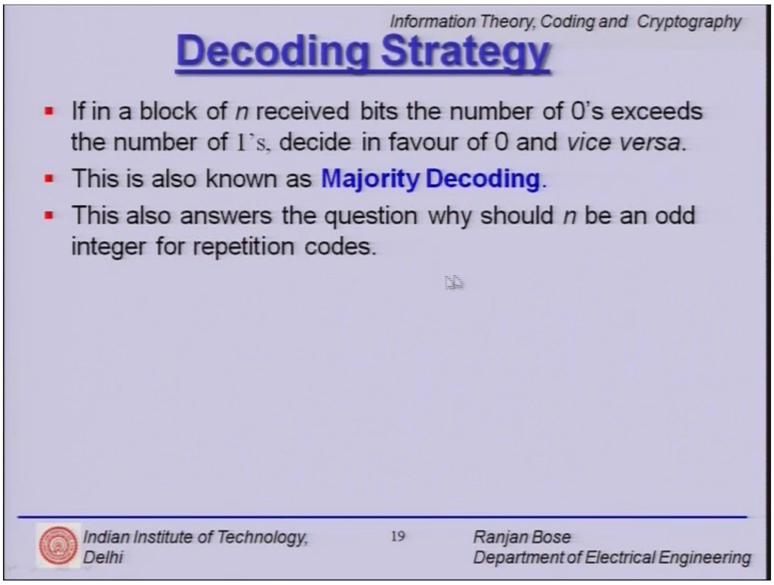
$$\boxed{r = \frac{1}{n}.}$$

Indian Institute of Technology, Delhi 18 Ranjan Bose Department of Electrical Engineering

So, coming back to our example, suppose we use a repetition code; Repetition code simply tells you that every bit is repeated n times. So, for example, if I have to send a 0, I do not send simply a 0; I send a 0 0 0. If you have to send a 1, I do not send a 1, I send a 1 1 1. So, this is a repetition code of length 3 also called block length n, small n is the block size and it is a simple block code.

So, for n is equal to 5, 0 is sent not as a 0, but it is repeated 5 times. 1 is not said as a 1, repeated 5 times right. So, the code rate is 1 over n where n is the length of the block code.

(Refer Slide Time: 42:40)



Information Theory, Coding and Cryptography

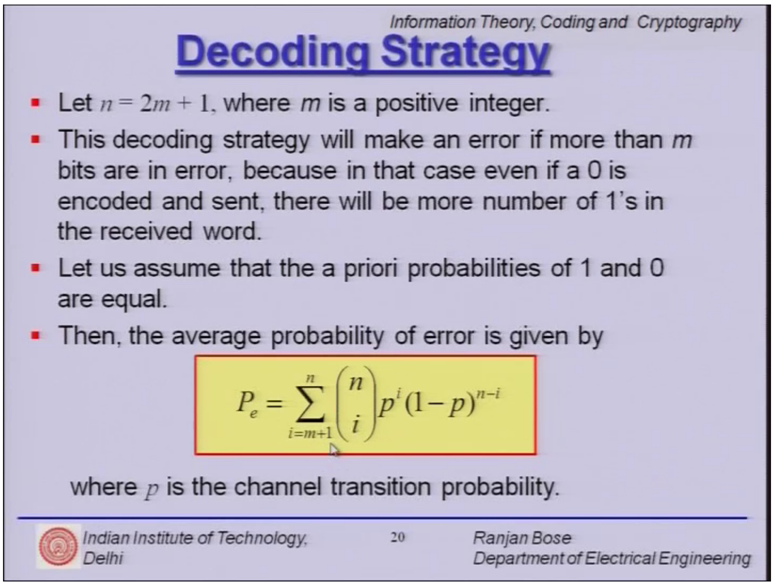
Decoding Strategy

- If in a block of *n* received bits the number of 0's exceeds the number of 1`s, decide in favour of 0 and *vice versa*.
- This is also known as **Majority Decoding**.
- This also answers the question why should *n* be an odd integer for repetition codes.

Indian Institute of Technology, Delhi 19 Ranjan Bose Department of Electrical Engineering

Now, what is the decoding strategy? Well a simple decoding strategy is the majority decoding strategy where whatever you receive, you count the number of 1's or 0's. If the number of 1's exceeds 0's, you declare a 1's was sent. If the number of 0's exceeds 1, then declared a 0 was sent because the probability of flipping of more 1's than 0's will skew the decision in the other direction. This also tells you why n typical is odd for repetition code.

(Refer Slide Time: 43:18)



Information Theory, Coding and Cryptography

Decoding Strategy

- Let $n = 2m + 1$, where *m* is a positive integer.
- This decoding strategy will make an error if more than *m* bits are in error, because in that case even if a 0 is encoded and sent, there will be more number of 1's in the received word.
- Let us assume that the a priori probabilities of 1 and 0 are equal.
- Then, the average probability of error is given by

$$\boxed{P_e = \sum_{i=m+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}}$$

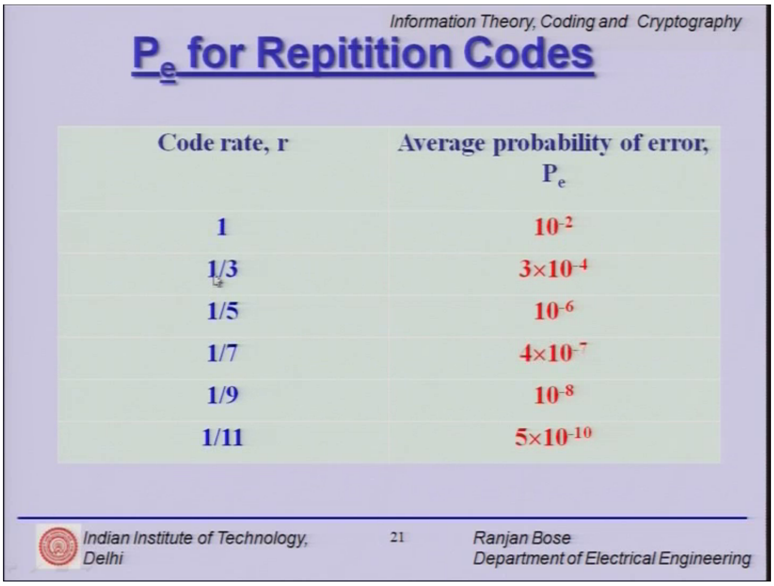
where p is the channel transition probability.

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

So, if n is a odd, n is equal to 2 m plus 1 where m is a positive integer and we use this majority decoding logic and let us assume that the a priority a priori probabilities of 1's and 0's are equal, then we can always write the average probability of error as follows.

Well it is a sum of mutually exclusive events. Error will happen, error means despite using this repetition code, you will end up making mistakes. When will it make a mistake? If more than the half plus 1 bits are in error. So, m plus 1 bits are in error and they can be permitted as follows or m plus 2 bits are in error going up to all the n bits are in errors. All these cases, the make this repetition code will make a mistake.

(Refer Slide Time: 44:19)



Information Theory, Coding and Cryptography

P_e for Repitition Codes

Code rate, r	Average probability of error, P_e
1	10^{-2}
1/3	3×10^{-4}
1/5	10^{-6}
1/7	4×10^{-7}
1/9	10^{-8}
1/11	5×10^{-10}

Indian Institute of Technology, Delhi

21

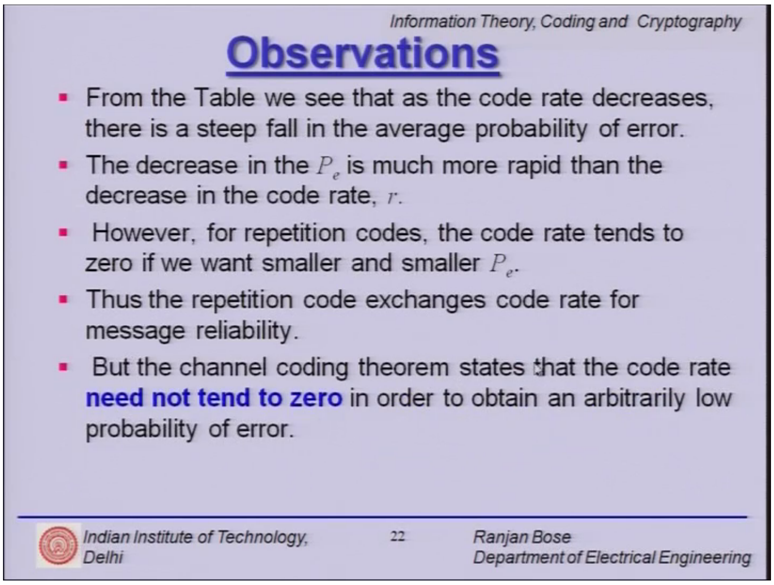
Ranjan Bose
Department of Electrical Engineering

We will go over this example in greater detail, but just to understand this point and relating it to the noisy channel coding theorem, you look at this code rates. Code rate r is equal to 1 means you are not using any error control code. You are not repeating it at all. So, it is a raw probability of error.

If you do 1 by 3, your probability of error is going down. Repetition code of length 5 gone down, 4 orders of magnitude. We have hope somewhere Shannon promised that we could have arbitrarily low probability of error and we are getting that and it appears that we have a strategy going with this repetition code. The only sad part is this code rate is also growing going to close to 0.

Shannon never said that code rate must also go to 0. It should said that it should only be less than C, C was just 0.9 and here we are dropped to 0.33, 0.20, 0.15, 0.11. It is dropping drastically something is wrong or this is just a toy example. It is not good enough. We can find better codes than that.

(Refer Slide Time: 45:45)



Information Theory, Coding and Cryptography

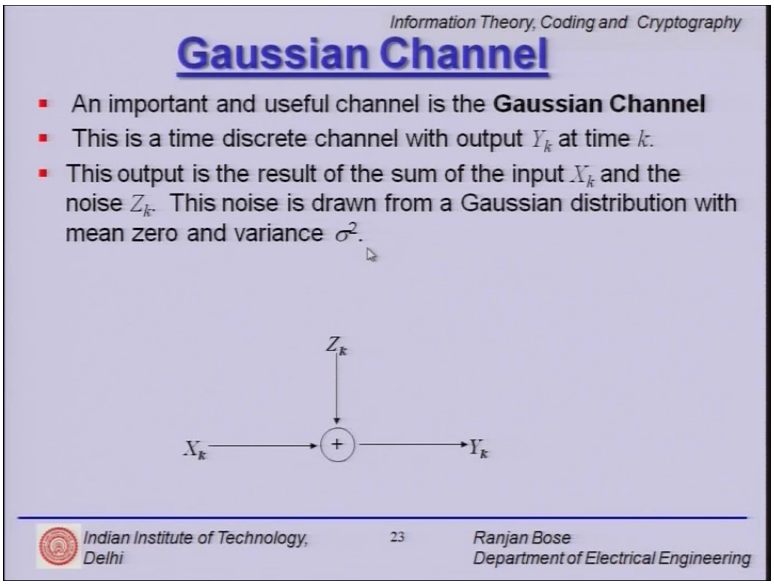
Observations

- From the Table we see that as the code rate decreases, there is a steep fall in the average probability of error.
- The decrease in the P_e is much more rapid than the decrease in the code rate, r.
- However, for repetition codes, the code rate tends to zero if we want smaller and smaller P_e.
- Thus the repetition code exchanges code rate for message reliability.
- But the channel coding theorem states that the code rate **need not tend to zero** in order to obtain an arbitrarily low probability of error.

Indian Institute of Technology, Delhi
22
Ranjan Bose
Department of Electrical Engineering

So these are the observation that the probability of error is decreasing rapidly right, but the for the repetition code, if you want to have smaller and smaller probability of error the code rate is also falling, but the code rate need not tend to zero that is the bottom line. The channel coding theorem the biggest contribution is that you can have very efficient codes where the code rate need not tend to zero, in order to obtain arbitrarily low probabilities of error. So, repetition code was an example, but probably not the best way to illustrate how we can tie it to the channel coding theorem.

(Refer Slide Time: 46:29)



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.

Indian Institute of Technology, Delhi
23
Ranjan Bose
Department of Electrical Engineering

Now, very quickly we will talk about the Gaussian channel and this is just a precursor to our next lecture where we will look at the channel capacity of a Gaussian channel. So, just let us have a quick view on this Gaussian channel. So, X k is the input and Z k is the noise and we get the output. This Z k is drawn from a Gaussian distribution with mean 0 and a variance sigma square.

We are curious to find out what is the capacity of this Gaussian channel, but we will take up this problem in the subsequent class. So, in summary we have looked at the different kinds of channel, channel models right starting from binary symmetric channel to discrete memoryless channel. We have looked at the channel coding theorem and we have discussed some examples and found out the need for a an error control code and why this code rate must be less than 1 and we linked it to the rate at which we use the channel and linked it to the source rate. So, with that we come to the end of our lecture and we will take it up in the next class.