

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

Module - 09
Channel Capacity and Coding
Lecture - 09

Hello and welcome to our next lecture on Channel Capacity and Coding. Let us start with the outline of today's talk.

(Refer Slide Time: 00:28)

The slide is titled "Outline" and lists the following topics:

- Channel Models
- Channel Capacity
- Symmetric Channels
- Noisy Channel Coding Theorem
- Examples

The slide footer contains the following information:

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


We will revisit some of the channel models that we have studied and then we will ask ourselves this most important question; how many bits per second can I send through a given channel? We will talk about symmetric channels and then we will introduce this noisy channel coding theorem. So, that is the brief outline for today's talk and everywhere we will sprinkle some examples.

(Refer Slide Time: 01:04)

Information Theory, Coding and Cryptography

Communication Channels

- After efficient representation of source symbols by the minimum possible number of bits, we need to **transmit** these bit-streams over channels (e.g., telephone lines, optical fibres, wireless channels etc.).
- The term **noise** is used to designate unwanted waves that tend to disturb the transmission and processing of the wanted signals in communication systems.
- The **sources of noise** may be external to the system (e.g., atmospheric noise, man generated noise etc.), or internal to the system (e.g., thermal noise, shot noise etc.).
- In effect, the bit stream obtained at the receiver end is likely to be **different** from the bit stream that is transmitted.

 Indian Institute of Technology,
Delhi

3

Ranjan Bose
Department of Electrical Engineering

So, let us see where we were what we learnt in the previous classes is that it is important to efficiently represent information before transmission because storage and transmission both require resources and hence money. So, it is worthwhile to compress and then save or send; now we have several kinds of channels that we typically use; they could be telephone lines, optical fibers, wireless channels, underwater channels. It could be long distance, space channels that we do for intergalactic; galactic communication if you will or any other kind of channel that you can imagine.

Now, most of these channels; in fact, all real world channels are affected by this interesting thing called noise. These are these unwanted waves that tend to disturb the transmission and processing of the wanted signals in a communication signals. The sources of noise could be many it could be external or internal, it could be atmospheric noise, man generated noise, thermal noise, shot noise, you can have several kinds of noise that you can have in the channel and it really affects the communication most importantly the rate of reliable communication.

So, aim of today's lecture is to understand what limits the rate at which we can transmit information over a given channel. And this channel itself will have the noise component as a part of it. The bottom line is what we send is normally not what we receive. So, the channel ends up flipping some of the bits at random and this is what the noises supposed to do.

(Refer Slide Time: 03:16)

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.

```
graph LR; A[Channel Encoder] --> B[Binary Modulator]; B --> C[Channel]; C --> D[Demodulator / Detector]; D --> E[Channel Decoder]; subgraph Composite_channel [Composite channel]; B; C; D; end
```

Composite channel

Indian Institute of Technology, Delhi 4 *Ranjan Bose*
Department of Electrical Engineering

So, what have we studied; so, far in terms of channel models? We have looked at the binary symmetric channels and if you look at it in a more generic manner; you can have if you see this is the channel we start from the center you can see this channel.

Now, in order to send waveforms over this channel we need a modulator; if you do not put in a modulator our signal will not travel far because most of the channels are band limited. So, one of the jobs of the modulator is to position the signal in the right frequency band so that it transmits over the channel and it travels to the distance it required to be transmitted.

Before the modulator we have this channel encoder; now this is where the binary bit stream which goes into the channel encoder, the binary another binary bit stream comes in with a modulator takes in a binary bit stream and converts it into a waveform. At the other, end at the receiver end we have the demodulator which takes the analog input the modulated waveform and demodulates it and converts it back to your bit stream and then the channel decoder does its job. We have not formally introduced the channel encoder and decoder whose job is essentially to recover from the errors introduced by the channel.

But if you look at this dotted rectangle, it tells me that this composite channel has binary input and binary output it could be M-ary also in general, but just for the sake of discussion we are saying this composite channel takes in bit stream and throws out

another bit stream. So, this is the dotted rectangle that we will represent using our channel models fine.

So, any error, distortion introduced by the modulator gets coupled into that generic channel model which so far we have looked at the binary symmetric channel. And we have also looked at just the binary channel which is not necessarily symmetric.


(Refer Slide Time: 05:45)

Information Theory, Coding and Cryptography

Discrete-input, Discrete-output channel

- The composite **Discrete-input, Discrete-output channel** is characterized by
 - the set $X = \{0,1\}$ of possible inputs,
 - the set $Y = \{0,1\}$ of possible outputs and
 - a set of conditional probabilities that relate the possible outputs to the possible inputs.

Binary Symmetric Channel

 Indian Institute of Technology, Delhi

5

Ranjan Bose
Department of Electrical Engineering

So, let us go a little bit further; let us talk about a discrete input, discrete output channel. Clearly the modulator has been plugged inside the channel composite channel. So, if you are talking about a binary discrete input discrete output channel; there is a set X which is the possible input and set Y which is the possible output. Here we have 0 1 as 2 possible inputs and 0 1 has 2 possible outputs.

But we should not be too comfortable assuming that there will be 2 inputs and 2 outputs, you can have 2 inputs 3 outputs, 2 inputs 4 outputs, 3 inputs 2 outputs. We have to talk whether that is a useful channel or not m inputs, n outputs nothing stops me from having m possible symbols as inputs and n possible symbols as output.

But what is interesting is that each time I send a symbol across a channel there is a probability a conditional probability associated with it. So, for a simple binary symmetric channel; if you send a 0 we hope and we pray that it goes as a 0, but once in a while the channel makes an error, but what do you mean by the channel? We cannot blame the

channel is the noise in the channel and several other factors which together couple and make a 0 appear as a 1 we make a mistake at the decoding end.

Similarly, a 1 conditionally you can say that it goes to 1 as with probability 1 minus p, but once in a while the channel flips this data. But when we blame it on the channel, we clearly know that we are talking about this composite channel which has bits coming in and bits going out. Now, if you look at a discrete memoryless channel which has Q-ary symbols. So, let the inputs be $x_0, x_1, x_2, \dots, x_{q-1}$.

(Refer Slide Time: 08:00)

Information Theory, Coding and Cryptography

Discrete Memoryless Channel

- Let the input to the channel are q -ary symbols, i.e., $X = \{x_0, x_1, \dots, x_{q-1}\}$
- Let output of the detector at the receiving end of the channel consist of Q -ary symbols, i.e., $Y = \{y_0, y_1, \dots, y_{Q-1}\}$.
- 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 $P(Y = y_i | X = x_j) = P(y_i | x_j)$

6

Indian Institute of Technology, DelhiRanjan Bose
Department of Electrical Engineering

And the output again has capital Q; uppercase Q; so, Q-ary symbols. So, output could be any one of them y_0, y_1, \dots, y_{Q-1} . And clearly we have small q into big Q number of lines connecting the inputs to the outputs, this is $1, 2, 3, 4, \dots, q-1$ and here again $0, 1, 2, 3, 4, \dots, Q-1$. And if, you see how many lines are connected there are small q into capital Q and this is the general right model for a discrete memoryless channel and conditional probabilities are written on each one of them.

So, it will be great to represent these conditional probabilities using a matrix which is called the channel transition probability matrix. What does it show well it tells us that most of the time my x_0 can go as y_0 , but several times we make a mistake similarly x_{q-1} then go to y_0, y_1, \dots, y_{q-1} .

(Refer Slide Time: 09:26)


Information Theory, Coding and Cryptography

Channel Capacity

- Consider a DMC having an
 - Input alphabet $X = \{x_0, x_1, \dots, x_{q-1}\}$ and
 - Output alphabet $Y = \{y_0, y_1, \dots, y_{r-1}\}$.
- Let us denote the set of channel transition probabilities by $P(y_i|x_j)$.
- The average mutual information provided by the output Y about the input X is given by

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

7

 Indian Institute of Technology, DelhiRanjan Bose
Department of Electrical Engineering

So, if you consider these input alphabets; X and output alphabet Y alphabet means it is a set, set of symbols. So, I have small q symbols as input and here I have changed to r ; so, r symbols as output.

Now, we are more interested in finding out is the channel doing a good job how much information is it really transmitting? So, we write the expression for the mutual information average. So, average mutual information $I(X;Y)$ it is given by this double summation right this we have come across earlier, this is $P(x_j)P(y_i|x_j)$; if you multiply this out log this is the conditional probability $P(y_i|x_j)$ over $P(y_i)$, but what we realize here these are the probabilities associated with the channel the channel transition probability matrix is have these probabilities as the inputs.


(Refer Slide Time: 10:30)

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.

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

 Indian Institute of Technology, Delhi8Ranjan Bose
Department of Electrical Engineering

So, we now finally talk about the capacity of a discrete memoryless channel it is by being one of the most important parameter which will characterize a channel ok. So, what is it defined as? It is defined as the maximum average mutual information in any single use of the channel where the maximization is done over all input probabilities. So, let us understand this jargon; we define the capacity as C it is a number what is it? Well we maximize the average mutual information $I(X; Y)$, over all input probabilities there is a strong physical meaning attach to this and we will go over it shortly once we complete it.

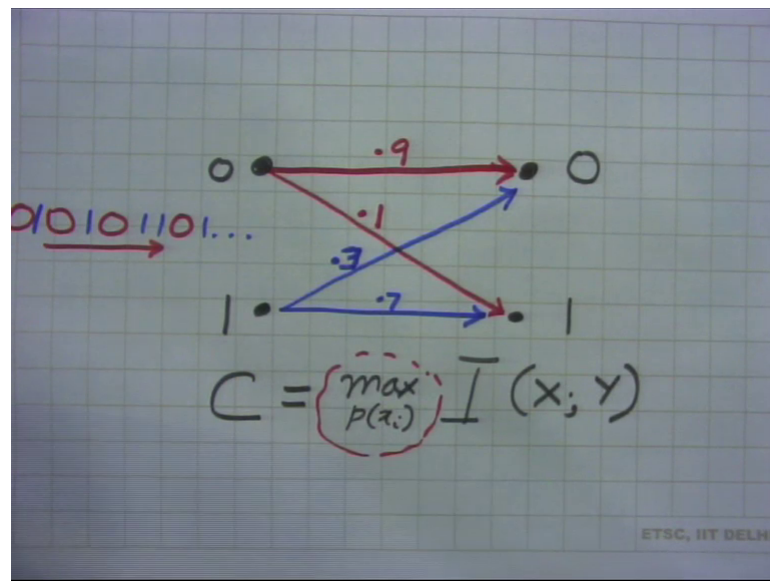
So, please recall that average mutual information is symmetric; so, it as well could be $I(Y; X)$. So, what have we written down? We maximize overall input probabilities $P(x_j)$ and this is the expression for the average mutual information. So, we are giving the benefit of the doubt to the channel we say skew the probabilities to the extent you want to, choose the probabilities at the input the way you want to do what you want.

But give me what is the best you can do in terms of maximizing the mutual information in the average sense. And whatever will be the value will be the capacity of the channel; what will be the units? Well, this average mutual information has the units of bits provided the loggers of the base 2, but what we are seeing is for any single use of the channel; that is every time we use the channel how many bits can effectively send every time I use the channel.

So, it makes sense to include that sentiment into the definition and the units. So, C will be defined as maximum average mutual information with the units bits per use. So, for every time we use the channel some bits get transferred and that will be the capacity. So, please note maximization is done over the input probabilities; let us look at a little bit of intuition that is attached to this.

So, let us draw a simple channel. So, I have got a 0 and a 1 and I have a 0 and a 1.

(Refer Slide Time: 13:26)



And clearly this channel makes a mistake once in a while; same with my 1, most of the time a 1 goes as a 1, but sometimes the one appears as a 0 at the decoder we make mistakes. But what is interesting is that the channel is partial let us say the channel makes a mistake.

So, with the 10 percent probability as 0 becomes a 1, but 90 percent of the time 0 appears as a 0. On the other hand my 1 is not a favorite of the channel. So, 70 percent of the time or we can flip it; 70 percent of the time a 1 appears as a 1, but 30 percent of the time the 1 goes as a 0.

So, clearly this channel is partial can this be a real life channel? Well sure it can be suppose we assign different amount of energy per bit for 0 and a 1. Then I might end up making more errors when transmission transmitting as 1 rather than transmitting a 0. So,

this binary channel it is clearly not symmetric is a candidate that we want to investigate further.

Now, clearly I can calculate the capacity of this channel. So, we would like to maximize over $P(x)$ right and $I(X; Y)$. Now we would like to build in the intuitive part into this; it really does not make sense for us to have this calculation done when probabilities are half and half, this is very clear because if the channel is more favorable towards transmission of a 0.

Then it will be advantageous to have more 0s in the input bit stream than 1s because the channel is treating 0 better than 1. Should I have all 0s? That is not right because even 4; 0s it is of making an error. So, I must adjust the ratios of 0s and 1 into the input bit stream. So, clearly there will be more 0s and there will be fewer ones right such that the mutual information is maximized.

So, this q in this channel forces me to have such a distribution at the input probabilities which will maximize this quantity. It will become clear very soon that the ratio of 1s and 0s should not be the ratios of 0.3 and 0.9 because if you have say m inputs and n outputs; this logic will not hold true.

(Refer Slide Time: 17:00)

Information Theory, Coding and Cryptography

Example

- Consider a BSC with channel transition probabilities
 $P(0|1) = p = P(1|0)$. $\rightarrow P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$

By symmetry, the capacity, $C = \max_{P(x)} I(X; Y)$ is achieved for $p = 0.5$


The capacity of a BSC is

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

Let us define the entropy function

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p).$$

Hence we can rewrite the capacity of a binary symmetric channel as $C = 1 - H(p)$.

 Indian Institute of Technology, Delhi
9
Ranjan Bose
Department of Electrical Engineering

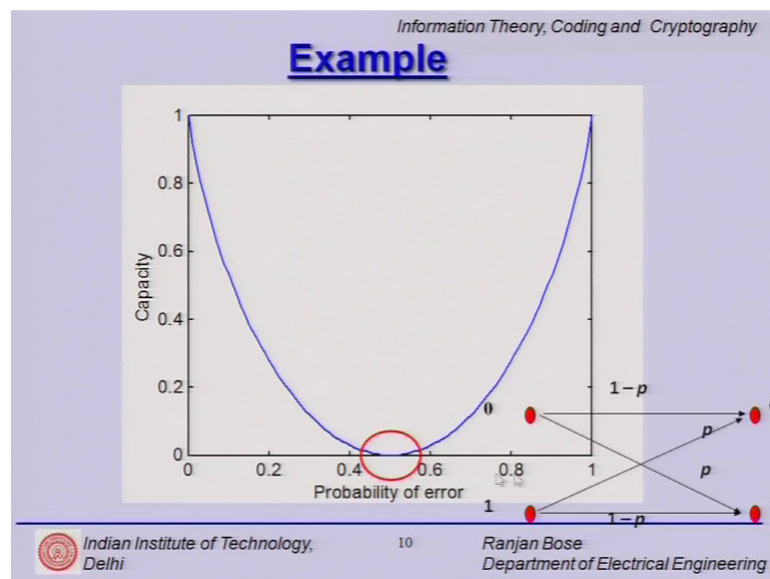
So, we come back to our slide and we say that we have this binary symmetric channel. In this case we have put the symmetry; so the probability of error is same for 1 and for 0.

In that case, it is clearly we must have the input probabilities as 0.5 and 0.5 for both 0s and 1s because channel is treating both 0 and 1 equally well. The probability of error of a 0 is the same as the probability of error for a 1. So, we would rather have the input probabilities this is this maximization; obviously, we know that this will be maximized when input probabilities they are equiprobable.

And if we do; so, we can write the capacity of the channel as follows C equal to $1 - H(p)$ where $H(p)$ is the binary entropy function. This comes from a simple observation of the entropy function that we have defined earlier. And we have seen that this capacity which is obtained by putting in the input probabilities is 0.5 and 0.5 right; it can be now written as $C = 1 - H(p)$ where $H(p)$ is the binary entropy function.

So, this is a toy example nonetheless a very useful example; it tells me that a capacity the capacity for this binary symmetric channel goes as $1 - H(p)$. So, what does it mean in reality?

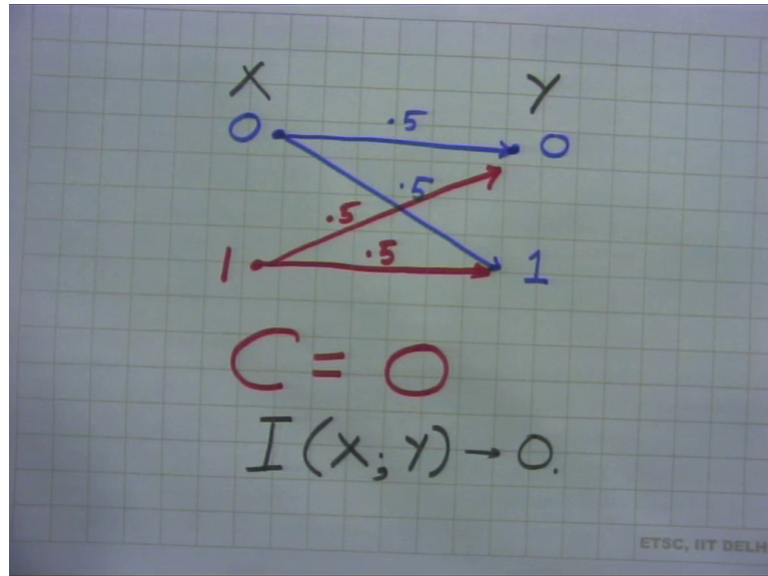
(Refer Slide Time: 18:53)



We have our favorite binary symmetric channel, but let us look at the entropy function right. And once we subtract this entropy function $1 - H(p)$ we have the capacity ok. So, on the y axis we have the capacity for this binary symmetric channel let us get it out of the way. And on the x axis we have the probability of error which is the crossover probabilities small p .

Intuitively it is clear that if you start from a very low probability of error; the capacity is high how high? Let us say we have an ideal channel; so, we draw an ideal channel.

(Refer Slide Time: 19:50)



So, we are just putting the transition probabilities such that 0 always goes as a 0 and a 1 always goes as a 1. So, this is your ideal channel we would like the world to be like this, but it is never. So, and then a lot of people who make money out of error control codes will be out of business.

So, let us let this be a toy problem. So, this says that every time I use the channel; I sent a 0 and it is received as a 0, every time I send a 1 it is received as a 1 and this channel never makes a mistake. So, intuitively this channel can transmit 1 bit per use that is every time I use the channel, I am able to send 1 bit. So, this is shown. So, if you go back to your graph here right here top when the probability of error the probability of flipping a 1 into a 0 or vice versa is 0 we have the capacity 1.

But here the magic begins the moment we have even a small probability of error; it drops drastically there is a sharp drop. So, by that time your probability of error is 0.1 or even less you have dropped 20 percent and it keeps dropping. So, this is fitting our intuition as the probability of error increases, the channel is becoming worse I am able to send fewer and fewer bits across the channel per use, but this is on an average.

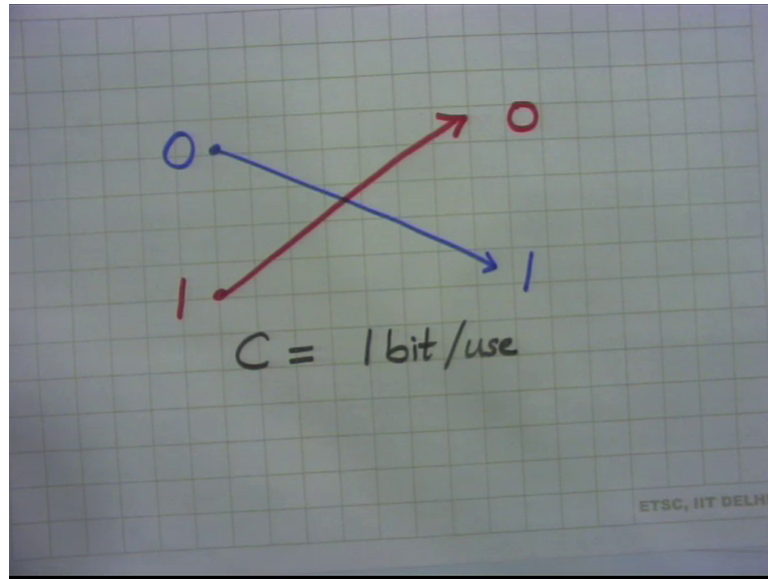
So, please know the units is bits per use now you cannot say hey how are you sending 0.6 bits every time I use the channel; it is on an average. So, when I am down to point 2 I am really down to this thing; so, the capacity is falling.

But then it starts slowing down and when we hit 0.5; we hit rock bottom and the capacity is 0 what does it mean? We look at the channel once again. So, this 0 goes as a 0, but it is equally likely that it goes as a 1. So, we have 0.5 and 0.5 at the same time if you have a 1; it goes as a 1 half the time and rest half of the time it shows up as a 0.

Now, the capacity of this channel is 0 what it means is that this channel fails to communicate any useful information from the X to the Y. So, this mutual information between X and Y; the average mutual information goes to 0. Intuitively it means that the channel is practically broken; what do you get? Half the time you get a 0, half the time you get a 1 regardless of what you are sending.

So, might as well throw away the channels, take a fair coin at the receiver and toss it and whatever you record is mathematically equivalent to what was being received through this channel; it is a useless channel the capacity is 0. So, this we go back to our slide this rock bottom point C equal to 0 happens for probability 0.5. But then the magic continues and we see that we go up slightly going up gradually what is happening is even we make the channel even worse are we? We are increasing the probability of error even further, but the capacity improves to the extent that it goes up and finally, we get a channel which always makes a mistake is it a bad channel? Not really.

(Refer Slide Time: 24:39)



So, this channel always makes a mistake. Again if we do the calculation, the capacity is equal to 1 bit per use; it is back to your ideal channel. All we have to do is flip the bit at the receiver; just flip the bit and you are home safe. You have been able to achieve 1 bit for per use, it never really makes a mistake always making a mistake that amounts to never really making a mistake. So, this is again a practically useful channel.

(Refer Slide Time: 25:27)

Information Theory, Coding and Cryptography

Example

- From the plot we make the following observations.
- For $p = 0$ (i.e., noise free channel), the capacity is 1 bit/use, as expected.
- Each time we use the channel, we can successfully transmit 1 bit of information.
- For $p = 0.5$, the channel capacity is 0, i.e., observing the output gives no information about the input.
- It is equivalent to the case when the **channel is broken**.
- We can, might as well, discard the channel and toss a fair coin in order to estimate what was transmitted.

Indian Institute of Technology, Delhi

11

Ranjan Bose
Department of Electrical Engineering

So, these are the observation that are now given here that we look at the 2 cases when p is equal to 0, noise free channel we have the capacity 1 bit per use. But the worst case

and probability of error is 0.5 the channel is practically broken and it is really not communicating any useful information across this channel ok.

And we have also seen that between 0.5 and 1. The capacity still increases because you are not really making the channel any worse. So, since p is a monotonically decreasing function of SNR the capacity of binary symmetric channel is monotonically increasing function of SNR.

(Refer Slide Time: 26:20)

Information Theory, Coding and Cryptography

Another Example

- Consider a Binary Erasure Channel (BEC) with transition probability matrix given by

$$P = \begin{bmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & \varepsilon & 1-\varepsilon \end{bmatrix}$$

p 0
 $1-\varepsilon$
 ε
 0

$1-p$ 1
 ε
 $1-\varepsilon$
 e
 1

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

Let us look at another example and this is also a very very instructive example; it is called the binary erasure channel. Here we have 2 symbols as the input, but we have 3 symbols as a possible output. So, you are small q and big Q that we were talking about in the earlier part of the lecture, they are not necessarily the same; now can this happen? Well why not.

Suppose, my channel sends 0 as minus 5 volts and 1 as plus 5 volts, but over the transmission period; there is attenuation, there is noise. So, we put in a guard band and we say even though plus 5 is 1 and minus 5 is 0. We will put in a guard band and say anything between plus 2 and minus 2 we will not take a call, we will say that it is erased it is an erasure.

Anything lower than minus 2 volts will be a 0 anything above plus plus 2 volts will be a 1. So, this is could be a simple strategy which will say that yes at the receiver; I will

either get a 0 or a 1 or do not know. So, there are 3 possible outputs and this is an example of a binary erasure channel.


Now can we apply; what we have learnt regarding the capacity for this erasure channel.

(Refer Slide Time: 28:00)

Information Theory, Coding and Cryptography

Another Example

- $C = \max_{P(x_i)} I(X;Y) = \max_{P(x_i)} (H(Y) - H(Y|X)) = \max_{P(x_i)} (H(Y) - H(\varepsilon))$
- Note that given the input X , there are two possible outcomes (Y) with probabilities $\{1 - \varepsilon, \varepsilon\}$.
- For the output Y , we have three cases with $P_Y(0) = p(1 - \varepsilon)$, $P_Y(\varepsilon) = \varepsilon$ and $P_Y(1) = (1 - p)(1 - \varepsilon)$. Therefore,
- $H(Y) = H(p(1 - \varepsilon), \varepsilon, (1 - p)(1 - \varepsilon)) = H(\varepsilon) + (1 - \varepsilon)H(p)$
- Hence,
- $C = \max_{P(x_i)} (H(Y) - H(\varepsilon)) = \max_p ((1 - \varepsilon)H(p) + H(\varepsilon) - H(\varepsilon))$
 $= \max_p ((1 - \varepsilon)H(p))$
- We know that the maximum value of $H(p) = 1$ for $p = 0.5$. Therefore, the capacity of a BEC is $C_{BEC} = 1 - \varepsilon$


15
Ranjan Bose
Department of Electrical Engineering

So, the formulas are the same we would like to find out the capacity C and it is nothing, but maximizing the average mutual information over input probabilities ok. But we know that $I(X;Y)$ can be expanded as $H(Y) - H(Y|X)$; this is the entropy of Y minus conditional entropy $H(Y|X)$. But we look back and see what is this entropy in Layman's language; it is the uncertainty; entropy is the uncertainty. What is the uncertainty of Y given X ? So, we go back to the question; suppose I give you X ; what is X ? Well either a 0 or 1; so, given X what is the uncertainty of Y . So, here Y can be a 0 or an erasure it cannot be a 1.

So, there is a probability that it will become a 0 with $1 - \varepsilon$ and probably that will be the erasure with probability ε . So, what is the entropy? Well, entropy there are 2 probabilities. So, $(1 - \varepsilon) \log \frac{1}{1 - \varepsilon} + \varepsilon \log \frac{1}{\varepsilon}$ which is the entropy function and so this will be nothing, but $H(\varepsilon)$ same is the logic for 1.

So, what are we trying to do? We are trying to find this quantity $H(Y|X)$; suppose X was 1 then again there are 2 possibilities either we will get a 1 or we will get a relation

with an associated probabilities what is the entropy there? Again H of ϵ ; so, for both 0 and 1 we have H of ϵ . So, on an average H of Y given X is nothing, but H of X ϵ .

So, we are now the problem is now reduced to finding out the maximum of this quantity H of Y minus H of ϵ over $P \times j$ clearly the second quantity is a constant, it depends only on the erasure probability. So, we are now focused on H of Y ; so, half the journey is done; just by making an observation. Note that the given the input X ; we have already argued that 2 outcome 1 minus ϵ and ϵ are there.

Now, we look at the other side let us look at this output Y ; we have 3 cases. So, we have 3 cases well the output is 0 erasure 1 ok. So, let us assume that the probabilities at the input are p and $1 - p$ what do we mean by maximize over input probabilities? It is telling us to find these values p what should be p and $1 - p$ that is the maximization problem.

So, what are the probabilities of the output $P(Y=0)$ is; $p(1 - \epsilon)$ P probability of e , this is the erasure is ϵ and probability of 1 is $1 - p$; $1 - \epsilon$, it clearly comes from this one; you can find out the probabilities of 0 which is coming from 0, probability of erasure this is p into ϵ plus $1 - p$ into ϵ right. And again probability of 1 which is nothing but $1 - p$ into $1 - \epsilon$ ok; so, you have these 3 probabilities why are we trying to find out these probabilities? If we have these probabilities at Y , then we can easily find out H of Y . So, if you have these probabilities; so, I plug in this in the formula and now capacity is maximizing over input probabilities $H(Y) - H(\epsilon)$ and what is $H(Y)$? $H(Y)$ we can easily find out right and you can get $1 - \epsilon H(p) + H(\epsilon) - H(\epsilon)$ which was there

So, now we have to maximize over $p(1 - \epsilon) H(p)$. So, we have to maximize this to clearly this $H(p)$ will be maximum because ϵ is a constant $H(p)$ will be a maximum for p is equal to half equiprobable case and. So, maximization problem leads us to the value of p is equal to 0.5 and somebody could have made a very simple observation here and say look as far as the channel is concerned it is symmetric with respect to 0 and one there is no reason why a higher value of 0 probability of 0 will maximize this capacity versus one. So, intuitively this p should be 0.5 and 0.5, but we should never take it right in the beginning 0.5 and 0.5 we should do some basic analysis

and thinking before assigning either equiprobable probabilities or probabilities which will maximize

So, our initial analysis has told us that to maximize H of p in order to maximize this right hand side I must choose p is equal to 0.5. So, if I plug in 0.5 H p becomes one and I get capacity as $1 - \epsilon$. So, this simple expression has a lot of intuition attached to it what is it well this result is quite intuitive let us say we transmit n bits or we use this channel n times. So, it is like a 1 1 1 0 0 0 one some random bit stream is coming in and I transmit it

But each time I send a bit either it is correctly received or with ϵ probability it is lost. So, if we keep repeating this experiment many many number of times then what I get is that if n is the total number of bits that are sent on an average n times ϵ bits will get lost and $1 - \epsilon$ times n bits will actually get correctly received

So, this channel has a leak it loses out n ϵ bits for every n bit sent. So, how many bits is it really transmitting well for every n bits that I sent $1 - \epsilon$ into n bits actually get passed. So, on an average per channel use how many bits can this channel send $1 - \epsilon$.


So, if we increase the value of ϵ we are decreasing the capacity in the limiting case when ϵ is 0 it is an ideal channel the capacity is one bit per use and if it is a worst case ϵ is one each time I send a bit it is lost the capacity goes to 0. So, this is a very intuitive example which tells us how effectively we can use the capacity.

(Refer Slide Time: 36:30)

Information Theory, Coding and Cryptography

Living with errors !

- All real-life channels are **affected by noise**.
- Noise causes discrepancies (errors) between the input and the output data sequences of a digital communication system.
- For a typical noisy wireless channel, the probability of bit error may be as high as 10^{-2} .
- This means that, on an average, 1 bit out of every 100 bits that are transmitted over this channel gets flipped.
- For most applications, this *level of reliability* is far from adequate.

 Indian Institute of Technology, Delhi

17

Ranjan Bose
Department of Electrical Engineering

Now, please note that all real life channels are affected by noise and this noise is actually causing these errors or erasures or mistakes in decoding. So, what is a typical value how many errors does a channel introduce? Well, it is pretty easy to measure for example, it should depend on the signal to noise ratio and if you have a typical noisy wireless channel, the probability of error for a bit may be as high as 10 raised to the power minus 2. It means that on an average 1 bit gets flipped out of every 100 bits that I send; it could be better, it could be one in 1000, but that is the best you can push your luck.

So, this is the real life effort can I improve it? Yes I can do it by increasing my signal to noise ratio, but then my battery will get drained out much faster. So, in a practical situation if I do not do much I will get roughly one bit flipped in a 100 or a 1000. Now is this is this good or bad will I be able to carry on a single conversation without a call drop on my mobile phone with this? So the answer is no, this is outrageous this level of error will not let me carry out a single conversation completely nobody will pay for this level of errors. So, what will people pay for?

Well, if you analyze people have different level of tolerances for different kinds of data. Suppose we are sending voice then an error of 10 raised to the power minus 4 is acceptable because I noise can interpolate noise are not so sensitive. So, it is a human being thing. So, 10 raised to the power minus 4 which is still much much lower than 10 raised to the power minus 2 is ok; people might want to pay for 10 raised to the power

minus 4 noise. But if I am sending data nothing less than 10^{-6} probability of error would be acceptable, people will just not pay there will be too many errors in the SMS that you receive, your calls will get dropped. So, people would definitely not pay for data with errors 10^{-4} .

Now, if you go to much more sensitive data for example, medical data I have a digitized an X-ray sample and the doctor is looking for white dots and any white dot might signals the start of a disease; well we would not like to take a risk there. So, we would push the tolerance to 10^{-7} , the point is we are looking at much lower orders of magnitude lower probabilities of error for any practically salable system, a system for which people will pay money for that.

So, we will come back to these numbers again and again and we will have to find a way to do reliable communication over unreliable channel. Those the unreliable channel has given with this figure of 10^{-2} or 10^{-3} , which is the uncoded the raw error rate in a channel. We have to do some magic to make this 10^{-2} down to 10^{-5} ok, but let us just look at a few more things about channels and then we will get back to it.

(Refer Slide Time: 40:37)

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 $C = \log|Y| - H(\text{row of transition matrix})$ and this is obtained for uniform distribution on the input alphabet. Here $|Y|$ represents the cardinality of Y .

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

So, quickly since this on the agenda we talked about the symmetric channel; a discrete memoryless channel is said to be symmetric if the rows of the channel transition probability matrix are permutations of each other and columns are permutations of each

other. So, just by looking at the matrix the channel probability matrix, the transition probability matrix we can find out whether the channel qualifies to be a symmetric or not.

And the other hand 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. So, there are 2 conditions symmetric has a certain kind of requirement and weakly symmetric which is a subset of symmetric channels it has a other requirement. Why are we doing this well if you have identified from observing the channel transition probability matrix that a channel is indeed weakly symmetric. Then the capacity of the weakly symmetric channel is simply given by \log of Y minus H ; row transition matrix where this absolute value Y represents the cardinality of Y .

So, the point is we do not have to go through that big double summation and maximization to find out the capacity of a weakly symmetric channel. If you have identified that a particular channel is weakly symmetric, you go ahead and apply this simple formula.

(Refer Slide Time: 42:15)

Information Theory, Coding and Cryptography

Living with errors !

- Acceptable bit error rates for various applications
- Different applications require different levels of reliability (which is a component of the quality of service).

Application	Probability of Error
Speech telephony	10^{-4}
Voice band data	10^{-6}
Electronic mail, Electronic newspaper	10^{-6}
Internet access	10^{-6}
Video telephony, High speed computing	10^{-7}

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

So, let us look at the probability of error that we said people are willing to pay for. So, this is based on experiments talking to people, doing surveys other scientific techniques and the acceptable right. Now this acceptability will change from people to people, country to country, people are more tolerant, less tolerant depends on how much you are

paying for are you on a gold plan or a silver plan people have different tolerances. But, typical probabilities of error that people are willing to pay for speech telephony 10^{-4} , voice band data 10^{-6} , electronic mail pretty much lot of internet stuff that we download 10^{-4} , and video telephony high speed computing medical data 10^{-7} or 10^{-8} .

(Refer Slide Time: 43:18)

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
20
Ranjan Bose
Department of Electrical Engineering

Now, we come to a very interesting theorem called the noisy channel coding theorem. So, what is it? We start with our favorite discrete memoryless source which is an alphabet X . So, alphabet is a set of symbols and it has an entropy H of X . So, clearly there is a probability associated with each of the symbols.

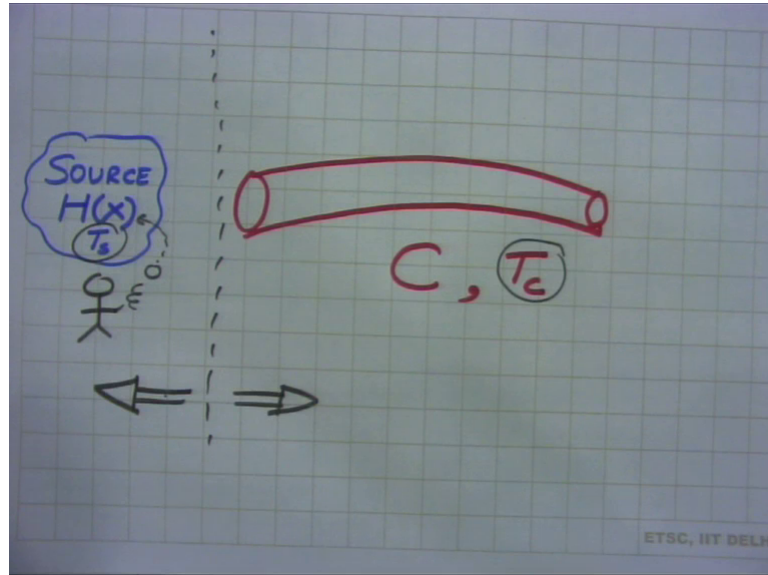
But for a change, we have introduced time in our equations because we do not have all days to send those 2 symbols. We have to get done with the job fast because the other symbols to be transmitted. So, we introduced this T_s which is kind of the time taken to send a symbol. So, $1/T_s$ would have some implication on the symbol rate.

Now, let a discrete memoryless channel that we are talking about have the capacity C alright; not only does it have a capacity C we are using this channel every T_c seconds. So, it is not that I am using this channel every second to send something maybe I am using it several times a second maybe $1/T_c$ times every second I use this

channel. So, that 2 independent things my source is generating a data and my channel is sending the data.

Let us quickly look at it diagrammatically.

(Refer Slide Time: 45:00)



So, we have this channel a nice fat pipe and it has a capacity C right. We do not use this channel once a second or twice a second, but once every T_C seconds, we use this channel. So, we have this T_C independent of it I have a source which has an entropy. So, this is my source and independent of this; it has this entropy H of X , but this entropy is not enough, it generates a symbol every 1 over T_S seconds; the rate is 1 over T_S . So, the point is this T_S and this T_C can be independent ok. This guy could be a person tossing your coin and it has an associated H of X and I have instructed this person to toss this coin once every T_S second.

So, please note the independence of this side and this side I can pay this guy more and ask him to toss the coin faster; independent of that I can start using this channel slowly or more frequently. So, coming back to our slide the noisy channel coding theorem relates the source rate what is the source rate? It is H of X divided by T of S . So, the left hand side of the dotted line is related to the right hand side which is the ratio C divided by T_C and noisy channel coding theorem says that if this left hand side the source rate is less than the C over T_C .

Then there exists a coding scheme for which the source output can be transmitted over this noisy channel and reconstructed with an arbitrarily low probability of error. Now this is the real punch line arbitrarily low how low do you want it to be? 10^{-5} no problem, 10^{-8} no problem you just make sure that this condition is met.

Not so happy, not satisfied 10^{-10} no problem 10^{-15} and nobody wants to go that low, but yes I can do that. There exists a coding scheme that is what the noisy channel coding theorem says. But conversely, if your friend $H(X)$ divided by $T S$ the source rate happens to be greater than the rate at which the capacity over $T C$ ok. We will we will put a name to this if this exceeds then I am sorry it is not possible to transmit information reliably and I cannot put a limit the error rate.

So, what are we saying? We are saying that this parameter C over $T C$ is somehow playing a very critical role and therefore, this is called the critical rate. So, if we restate this noisy channel coding theorem; it says that if the source rate is less than the critical rate, we are in business. There exists a coding scheme which will guarantee arbitrarily low probability of error.


And conversely if your source rate exceeds the critical rate then very hard because you will not be able to limit the probability of error. So, this C over $T C$ happens to be the key fortunately we know how to calculate C , we know how to calculate $H(X)$. So, we can figure out how to choose. So, we have a design problem at hand we have a luxury to choose this $T S$ and $T C$ such that I can hopefully do a reliable communication.

(Refer Slide Time: 49:50)

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* 21 *Ranjan Bose
Department of Electrical Engineering*

So, this channel coding theorem is a very very important result information theory. In fact, it specifies a fundamental limit on the rate at which the reliable communication be carried out over an unreliable channel. What I will be talking about? Reliable communication what is reliable communication? You define of course it is in terms of bit error rate. So, for you reliable is 10^{-7} , you got it. For you reliable is 10^{-12} ; be my guest no problem; we will give you reliable communication you choose the level over an unreliable channel just to meet the condition ok.

But this noisy channel coding theorem is good and bad; it tells us about the existence of some codes ok. It will tell you, it will show you the carrot it exists it is an existence proof, but it will never it does not give you the recipe of finding that code ok. It does not tell you how to come up with those codes this channel codes which will ensure or give you this wonderful reliable communication, but this is not a bad news for those who make a living out of finding better and better channel codes because they are still in business. So, this is the reason why it is a fundamental limit.

(Refer Slide Time: 51:33)

Information Theory, Coding and Cryptography

Summary

- Channel Models
- Channel Capacity
- Symmetric Channels
- Noisy Channel Coding Theorem
- Examples

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

So, let us look at what we have talked about in today's lecture. We started off with a few channel models and then we introduced this very very important notion of channel capacity. We looked at symmetric channels and how weakly symmetric channels can easily be used to determine the capacity of a channel. And finally, we scratched the surface regarding the noisy channel coding theorem. And of course, we looked at a few examples like the binary symmetric channel and the capacity of the binary erasure channel.

With that, we come to the end of this module.