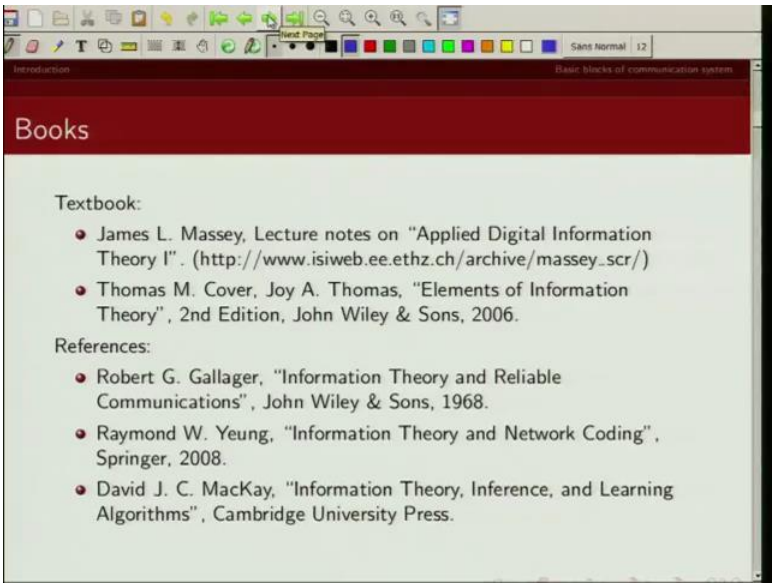
An Introduction to Information Theory
Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture - 1A
Introduction

Welcome to the course on An Introduction to Information Theory. I am Adrish Banerjee from Indian institute of technology, Kanpur and I am going to teach this course. Now in today's lecture I am going to give a brief overview of the contents of this course, what does information theory deals with and I am also going to talk about some of the books that we are going to follow in this course.

(Refer Slide Time: 00:44)

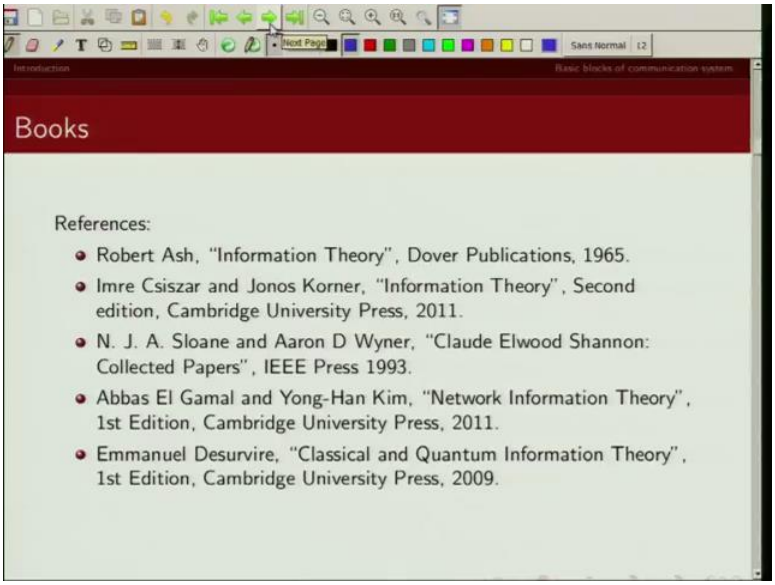


So, to first lecture notes that we are going to follow in this course is by late professor James L Massey these are Applied Digital Information Theory I and these lecture notes are available at this website which is maintained by ETH Zurich. These are very nice sets of lecture notes on information theory and very easy to read and understand. Another classic book on information theory is by late professor Thomas cover and his student Joy Thomas it is called Elements of Information Theory, second edition and this book is

available in cheap Indian edition. It is a very classic book on information theory and it covers most of the topics that we are going to cover in this particular course.

Some of the other books, very nice books - this book by professor Robert G Gallager is a very classic book on information theory you can also follow this book or there is modern book on information theory which contains some modern topics which we would not be covering in this particular course, but very nice book on information theory by professor Raymond Yeung you can also use this book. Another nice book on information theory is book by late professor David MacKay again this book is very easy to read and very nice book.

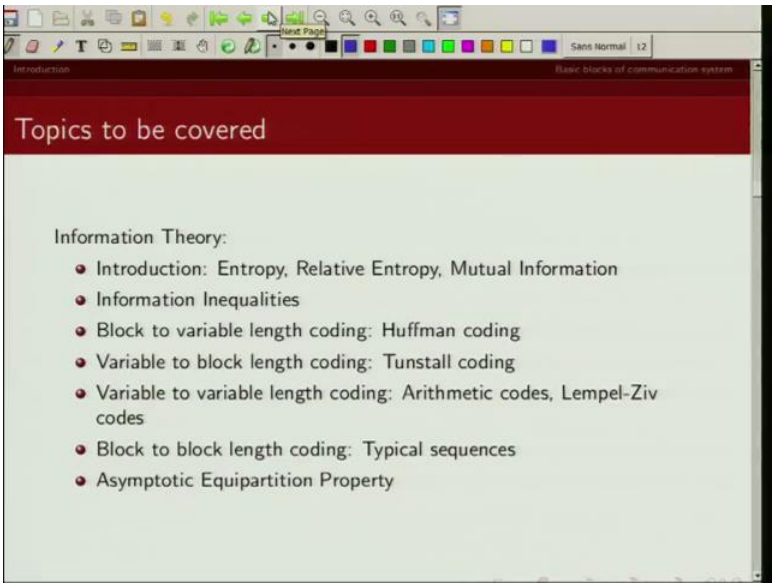
(Refer Slide Time: 02:24)



And old classic on information theory is book by professor Robert Ash this does not cover all the topics that we will be covering in this course, but whatever topics it covers is presents it in very nice fashion. Another classic book on information theory is by professor Csiszar and Korner called information theory published by Cambridge university press. I strongly suggest you read this collected papers of Shannon, it is a very nice collection of papers by Claude Elwood Shannon, the father of information theory – it is edited by Sloane and Wyner, very nice collection of papers by Claude Shannon.

Now there is a separate book on network information theory by professor Abbas El Gamal and Yong-Han Kim, they also have a set of lecture notes which are readily available on the internet on Network Information Theory that is a very nice reference on network information theory. And finally, this books covers both classical as well as quantum information theory though we will be covering in this course only classical information theory, but those of you who are interested to learn quantum information theory can refer to this book by Emmanuel Desurvire.

(Refer Slide Time: 04:01)



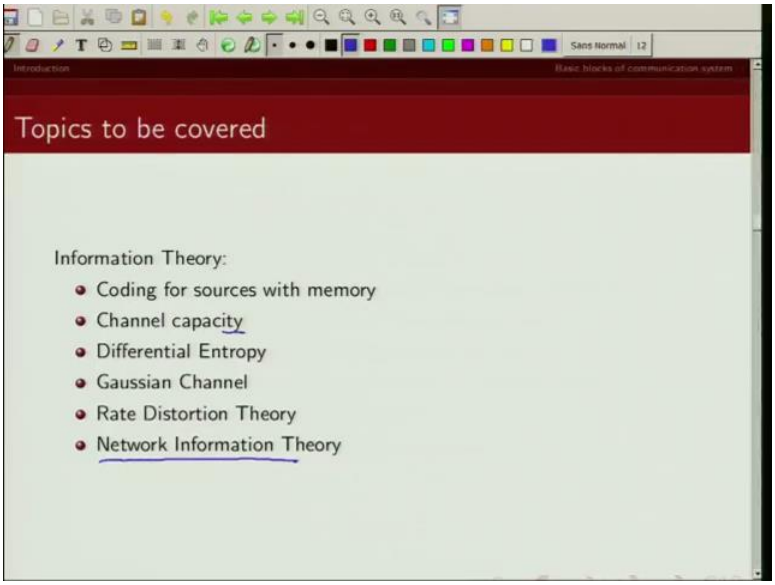
So let us now talk about what are the course content for this particular course. So, we will first start of with quantifying information what is information and how do we quantify information. So, we will talk about heartless measure of information and Shannon measure of information and then we will talk about entropy, relative entropy, mutual information and their properties. After that we will talk about some information inequalities, Jensen's inequalities, log sum inequality and some other results such as fano's lemma which we will be using subsequently later on to prove some other results. Then we will move into the problem of source compression source coding. So, source coding can be from block to variable length coding, variable to block length coding, block to block length coding and variable length coding.

So, that depends on whether the input sequences of fixed length if you are calling a block or input length is variable that is this block and this block to variable whether the output is of variable length or fixed length. So, we can get compression if we take block of data, we can get compression by representing blocks which are happening more frequently by lesser number of bits and blocks which are happening less frequently by larger number of bits. So, in that way we can reduce the number of bits required to represent the source.

So, we will talk about optimal blocks to variable length coding which is your Huffman coding, we will prove the conditions for optimality what is the minimum number of bits to required to represent the code. We will also talk about variable to block length coding and for a specific instance where the message can be passed in a particular fashion, we will talk about an optimal variable block length coding in that particular instance which is what is known as Tunstall coding. We will also talk about variable to variable length coding which are very prominently used in lot of source compression algorithms. So, in particular we will talk about arithmetic coding and Lempel-Ziv algorithm which comes under class of universal source compression algorithm, you do not require any knowledge about the prior source distribution of the source that you are trying to compress. We will also talk about block to block length coding, now if we are doing block to block length coding to get compression; obviously, we are encountering a lossy compression.

Now we will show that there are some sequences which are more likely to happen, given a source distribution there are some sequences which are more likely to happen from that particular source and some sequences are less likely to happen. So, when we do block to block length coding we will try to encode, assign individual code words to all those sequences which are more likely to happen which we call as typical sequences and of course for all other non typical sequences we will assign one particular code word. So, when we transmit a bit of sequence the decoder will be able to decode that sequence where as for the non typical sequence we would not be able to distinguish what sequences was transmitted and hence this is the example of a lossy source compression. We will also talk about what is collectively known as Asymptotic Equipartition Property, this is information theory analog of loss large numbers; we will talk about what are these properties and what is the consequence of these properties.

(Refer Slide Time: 08:15)



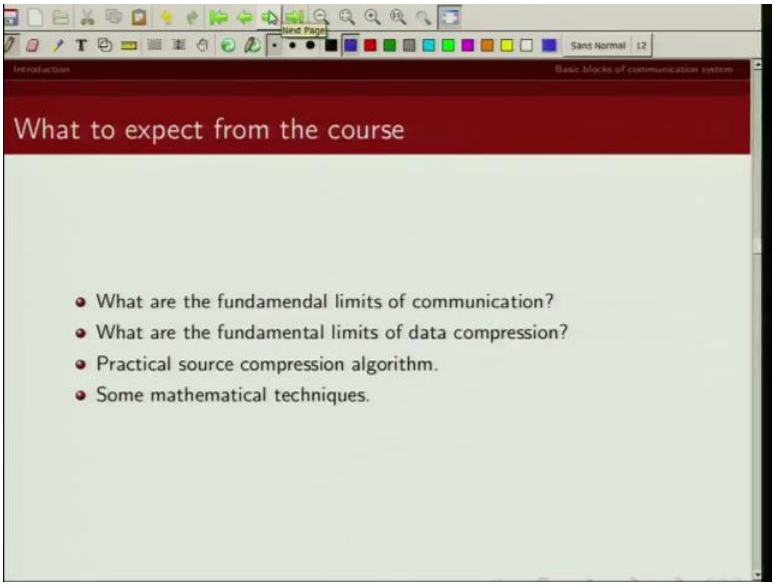
Then we will move to source compression for sources which have memory. So, we will exploit basically temporal correlation between bits that are coming out of source we are going to exploit that correlation to design ours source encoder. So, this will be our coding for sources with memory and we take a very simple example to illustrate that. After this few lectures on source compression we will now move to channel capacity computation. So, channel is the medium over which we are communicating, how many bits we can transmit over communication link that basically channel capacity. So, we will talk about very simple channel models then we will talk about how to compute the channel capacity. Then we will move from discreet random variable to continuous random variable and we will define entropy for continuous random variable, that is basically what is known as differential entropy and we will consider an example of very commonly channel which is addictive white Gaussian noise channel, we will compute the capacity of addictive white Gaussian noise channel.

Next topic deals with what is known as rate distortion theory. Now if we have a real number and you try to represent that real number, if you want to represent it you require infinite number of bits right, but if you try to represent that real number with finite number of bits then essentially you are introducing some sort of distortion. So, how do you find out that your representation using fixed number of bits is a good representation

of a real number? So, you need to describe a goodness measure between your original real number and its representation that is basically known as distortion measure. Now rate distortion theory deals with if you have a source and a given distortion measure you are interested in knowing for example, what is the minimum average distortion that can be achieved for a given rate.

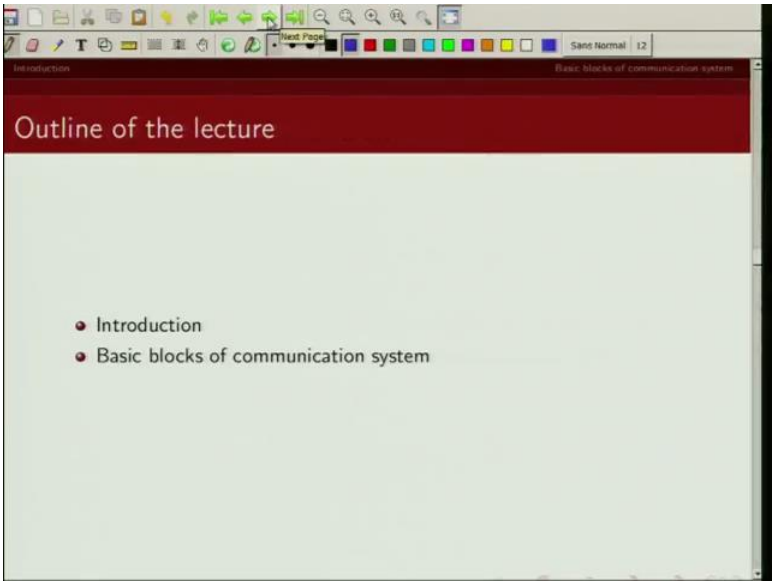
So, given distortion measure we are interested basically in finding out or we can ultimately say that given a distortion average distortion measure you want to find out what is the minimum rate that you can achieve. And finally, if time permits we are going to talk about network information theory. Now earlier here when I talked about channel capacity I am going to talk about point to point channel. So, there is 1 transmitter and 1 receiver. So, we are going to characterize a capacity of such channels. Now think of scenarios when you have multiple senders and multiple receivers, how do you find out capacity of such channels or let us say you have multiple senders multiple receivers and you want to do distributed source compression. So, all these problems come under this network information theory. So, this is roughly the syllabus that we plan to cover in this 20 hour lectures, spanned over 8 weeks.

(Refer Slide Time: 11:53)



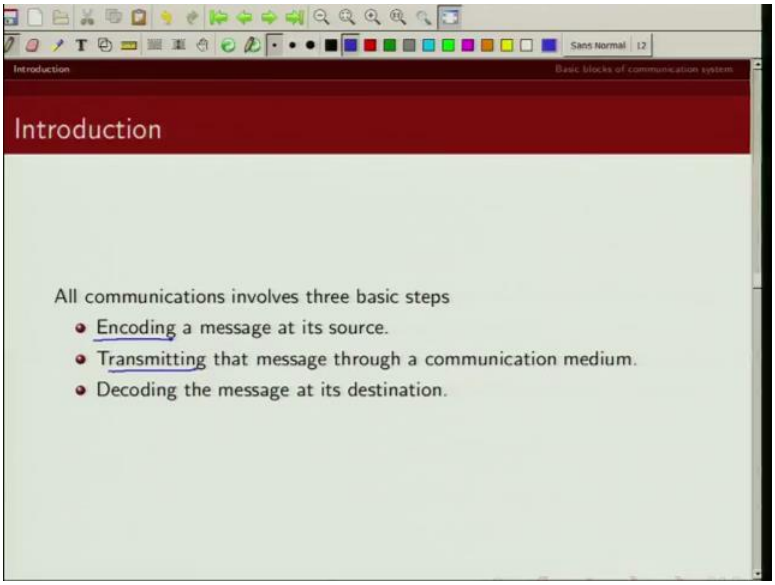
So, what can you expect after going through this course, you should be able to understand or find out what are the fundamental limits of communication, at what rate we can reliably communicate over a communication channel. You should also be able to find out what are the fundamental limits of data compression. So, example if you want to do loss less compression what is the minimum number of bits required to represent the source or you want to do a lossy compression given average distortion measure basically what is the minimum number of bits you required to represent the particular source. In the process you will also come across some practical algorithms for source compression and you will also learn few mathematical techniques how to prove theorems and thinks like that.

(Refer Slide Time: 12:49)



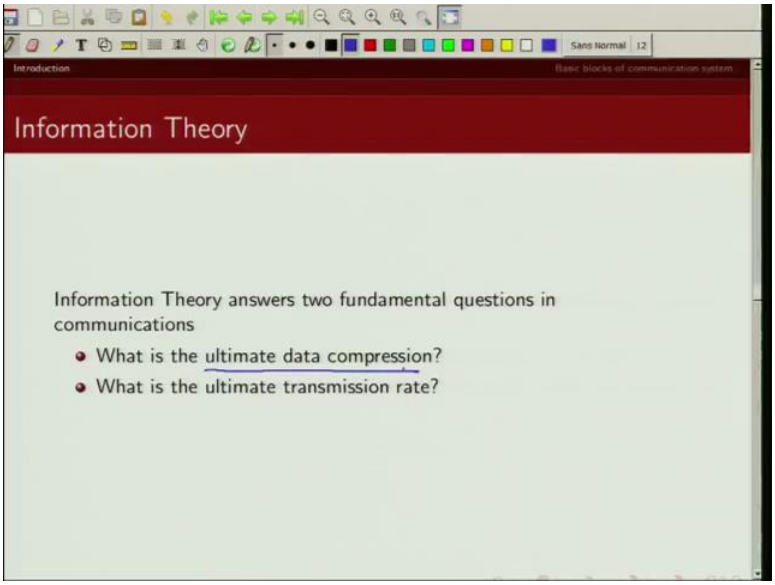
So, let us look at a block diagram of a communication system and see where does information theory what are the problems that information theory deals with. So, what are these various steps involved in communication?

(Refer Slide Time: 13:04)



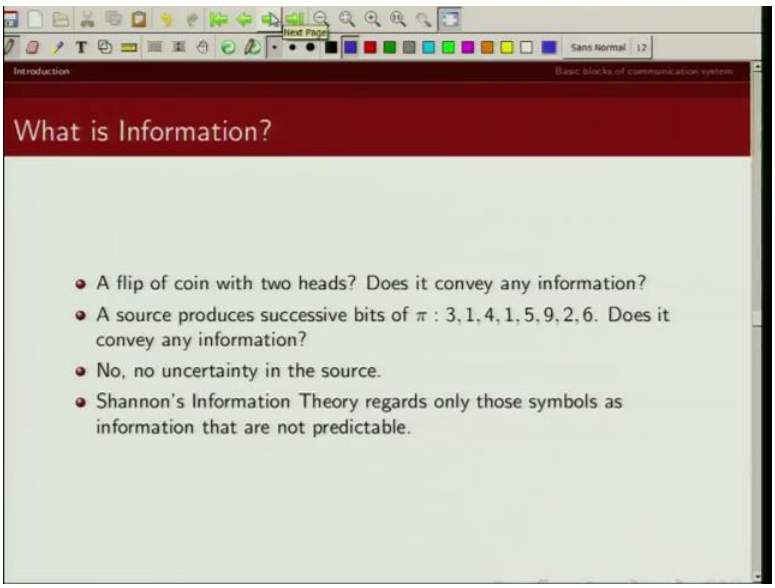
So, we can roughly say that communication involves three processes - the first is encoding a message. So, given source you want to efficiently represent that source using a fixed number of bits that is encoding and then once the message is encoded you want to transmit it over a communication link. That is basically this point transmitting the message over a communication link. And finally, the bit that is being transmitted over a communication channel that will get corrupted due to noise, so at the receiver you would like to decode and estimate what was the message that was transmitted. So, roughly in any communications these are the three basic steps involved.

(Refer Slide Time: 13:58)



So, information theory answers two fundamental questions in communication number one - what is the ultimate data compression that you can achieve. So, what is the minimum number of bits required to represent the source and the other question that is it answers is what is the maximum rate at which we can transmit over a communication channel, and still at the receiver you should be able to reliably decode that message.

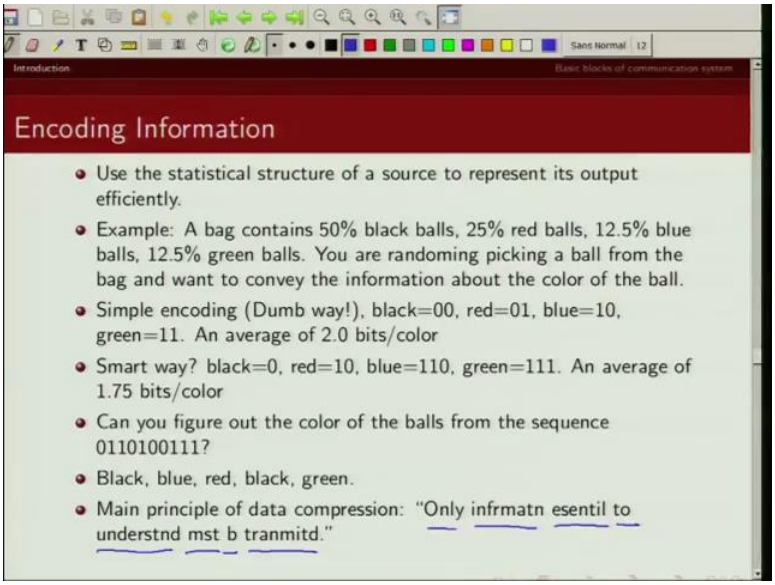
(Refer Slide Time: 14:34)



So, before we go to the block diagram for digital communication system let us first intuitively try to understand what is information. So, I am giving you an example - let us say you have a coin which is a biased coin. So, this is the coin which has both the sides head and I flip this coin and ask you to guess what is the outcome does this flipping of coin convey any information to you, the answer is no, why? Because it is the biased coin you know both sides are head. So, no matter what I flip I will always get a head. So, this my flipping of coin does not convey any information because you know beforehand that you will get a head right. If this would have been an unbiased coin which has head on 1 side and tail on 1 side then flipping of coin would have conveyed information because you do not know whether head is going to come or tail is going to come.

Let us look at a source which produces these bits 3 1 4 1 5 9 2 6, does this convey information? Again in this case these are basically representation of this number pie 3 1 4 1 5 9 2 6. So, if I give you this numbers you know this I am giving you basically representation of pie. So, this again does not convey any information to you. So, why does not it convey any information? Because there is no uncertainty in the source, the outcome of this experiment is known. So, Shannon's information theory regards only those symbols as containing information that are not predictable. So, as I gave an example if you have an unbiased coin or let us say you have an unbiased dice if you roll, the dice rolling of the dice when I ask you to guess what is the number is that event conveys information because you do not know when I roll an unbiased dice whether I am going to get 1 2 3 4 or 5 or 6.

(Refer Slide Time: 17:08)



So, let us pick a very simple example to illustrate how we can do encoding of information. So, as I have written here we are going to use the statistical structure of the source to represent the output efficiently or what do I mean by statistical structure the source. So, clearly if we are interested in source compression we should try to represent sources which is occurring more frequently lesser number of bits and outcome which are happening with less probability we should use more number of bits to represent that.

So, let us take an example I have a bag that contains black balls which are 50 percent, red balls which have 25 percent, blue balls which is 12.5 percent and green ball which is 12.5 percent. So, half the balls are black balls, one-fourth of them are red balls, one-eighth are blue balls and one-eighth are green balls. Now from this bag of balls which has black color ball, red color ball, blue color ball, and green color ball and randomly picking up the ball and showing you the color of the ball. Now I am asking you to convey that information convey the color of that ball I want to encode that information. So, I have a bag of balls I am picking up the ball randomly and showing you the color and I want to encode the color of the ball that information I want to encode using bits. So, how can I do that?

So, I said a simple encoding which I call a dumb way I will explain why this is dumb way, a simple encoding would be as follows. Now how many different colors of balls you have, you have black ball, red ball, blue ball and green balls; you have 4 different types of balls. So, very simple encoding could be I could use 2 bits to represent the color of the ball because I have 4 possibilities I can use 2 bits to represent. So, for example, I can use 00 to represent, black ball I can use 01, to represent red ball I can use 10, to represent a blue ball and I can use 11 to represent a green ball.

Now if I do this on an average I am requiring two bits to represent a color. Now why this is dumb way, please note that 50 percent of my balls are black. So, when I pick up a ball I every second ball is like to be a black ball. Now this encoding which I did here does not take into account the frequency of the color which comes out. So, clearly you can see if I pick up a ball and more likely to get a black ball then it is a green ball right, but I am using two bits to represent a green ball and using two bits to represent black ball also. So, that is why I said this is a simple encoding, but a dumb way of encoding because I am not exploiting the fact that 50 percent of my balls are of black color. So, ideally what should I be doing? I should be representing black color with less number of bits and green color with more number of bits because black is occurring with more frequently. So, a smart way of doing could be. So, I am using 0 to represent a black ball I am using 10 to represent a red ball I am using 110 to represent a blue ball and I am using 111 to represent a green ball. Please note none of these bits which I am using to represent a number is a prefix of any other code for example, if I get a 0 that is a black ball. Now note none of these are starting with 0 - red does not start with 0, blue does not start with 0, green does not start with 0.

Similarly, look at the coding for red ball I am using 10. Now note, blue does not start with 10 green does not start with 10 and I am using 110 clearly green does not have 110 as a prefix. So, there will be no confusion in decoding this particular encoding scheme. Now on an average here I require 1.75 bits per color, how did I compute that? Now for black I require only 1 bit and black is happening with probability half, for red I require two bits and red is occurring with probability 1 by 4, blue I am using 3 bits and it occurs with probability 1 by 8 that is this and similarly green balls I am using 3 bits to represent green balls and they occur with probability 1 by 8. So, this is my average number of bits

required to represent this color and this comes out to be 1.75 and please note, here I am using 1.7bits per color whereas, in my down way of encoding I was using 2 bits per color. So, clearly I am doing better. Now let us take an example I will give you some string of code words and you will have to guess what color of wall is this.

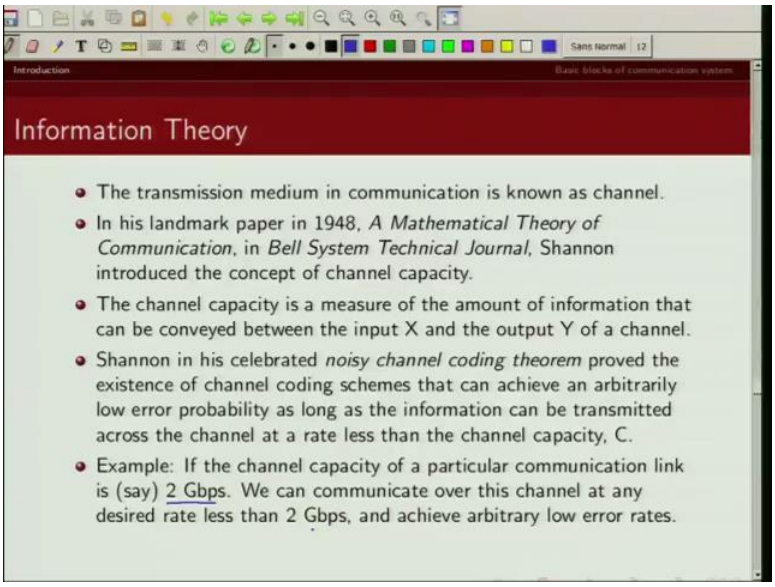
So let us look at this sequence of number 0110100111. Now note that for black I am using code 0 and for red I am using 10, for blue I am using 110 and for green I am using 111, let us look at this. The first bit is 0, what color is this? This is black because black I am representing by 0. So, this is black. Next I have 1, am I using any 1 to represent any quote. So, 1 could be red, could be blue, could be green, and let us look at next bit. So, next bit is 11 now which color has code word starting with 11? Blue has 110 and green has 111. So, let us look at another bit.

So, this is 110. So, what is 110? That is blue; now let us look at next step that is 1. So, now, again this could be red, this could blue, or this could be green. So, let us look at other bit 10 what is 10? 10 is red. So, next ball is red. Next we have 0 here, what is 0? 0 is black color ball and next we have 1. So, 1 could be red, could be blue, could be green. Let us looks at next two bits 11, now 11 could be blue or could be green let us look at next bit which is 111 and what is 111? That is green. So, this is, you can see this is the sequence of colors which have been encoded using bit this frequency.

Another interesting thing you note here is I am using a variable length code for black I am using 1 bits, for red I am using 2 bits, for blue I am using 3 bits, for green I am using 3 bits. So, I am using a variable length code to describe the color of the ball, but because none of these code words are prefix another code word I am able to make out the boundaries or I am able to make out when particular code word ending. So, you can see using this is an example of a available length coding where we use variable length code to describe the color of the ball and you can see that we are getting advantage in terms of number of bits we are using to describe the color compare to this dumb way of describing where we are not taking into account the fact that black balls are more likely to happen then let us say blue balls, green balls or red balls.

So, the message is like this. So, the main principle of data compression is can you read this out, only information essential to understand must be transmitted. Note that I have deleted some letters, but still you are able to understand what is written. So, that is the basic idea behind compression that we would like to represent our source with minimum number of bits.

(Refer Slide Time: 28:08)

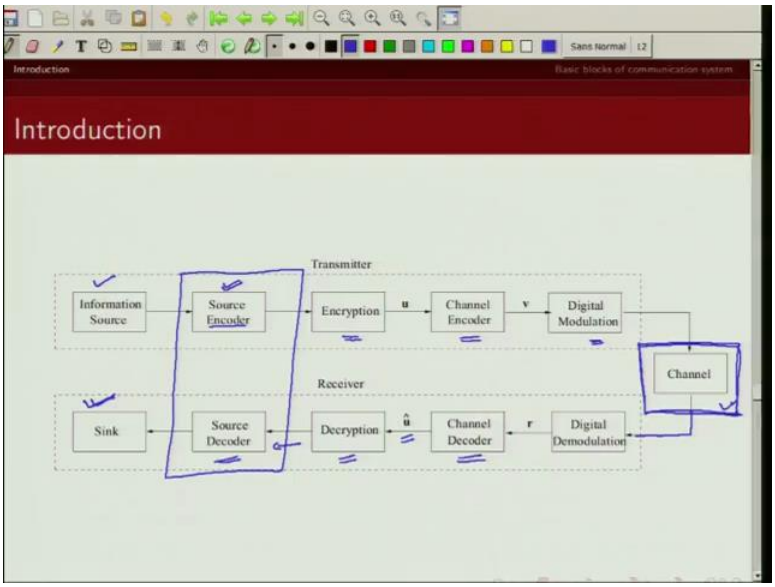


So, this is some examples to illustrate how source compression is done. Now let us move into the next part of information theory which will deal with describing or calculating the channel capacity. So, let us explain what is channel and what is channel capacity. So, the transmission medium over which we are communicating is known as channel. So, channel is the medium over which we are communicating. Now Shannon in his landmark paper in 1948 described this concept of channel capacity and you showed that if you try to come transmit at a rate below channel capacity then you are you can get arbitrarily low batterer performance as long as you consider very large block sizes.

So, channel capacity is a measure of how much information that we can transform from the input of the channel to the output of the channel. So, this is a measure of amount of information that can be conveyed between the input of the channel and the output of the channel. Shannon in his landmark paper has mentioned that that there exist channel

coding schemes which can achieve arbitrarily low error probability provided the transmission rate is less then channel capacity. So, channel capacity is the fundamental limit of data transmission we cannot communicate reliably over a communication channel if the transmission rate is above channel capacity. So, for example, if you say channel capacity of particular communication link is 2 Giga bit per second then we should be able to send at any rate less than 2 Giga bits per second and it should be able to reliably communicate it.

(Refer Slide Time: 30:20)



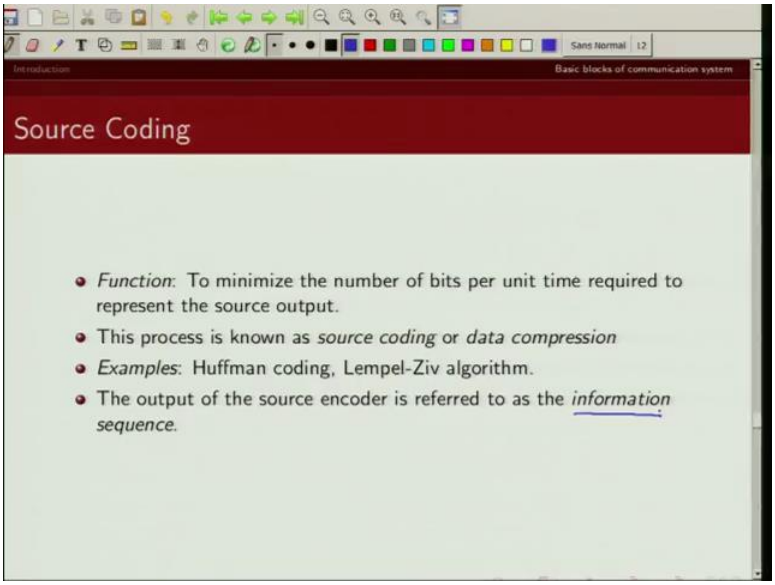
So, let us look at the block diagram of communication system. So, this is information source that you want to transmit to a receiver here. So, what is the first step involved. So, this source encoder what is does is, it tries to represent the information source in a compact fashion. For example, in an English language the letter q is always followed by u is no word starting with q which does not have u as second letter. So, if you want to start transmit let us say a word starting with q is redundant to send u because q will be always followed u right. So, the whole idea of source encoder is to basically represent your source efficiently in minimum number of possible bits and in this course we are going to spend quite a bit of time in talking about this source encoders. Will talk about practical source encoding algorithms, will talk about what is a fundamental limit if you try to compress beyond that, you cannot get lost less compression. Once we have the

compress bit the idea of this encryption is to secure basically. So, you do not want those bits to be deciphered by unwanted user.

So, this is for security you do some encryption the idea of channel encoder is to introduce some redundancy in the source in a control fashion, so that with help of those control redundant bits you should be able to detect or correct errors. So, after this channel encoder you have what is your modulation. So, you want pass it those bits over a communication channel you do modulation, now once you modulate this signal the signal passes through your transmission medium which is channel. Again we will spend some time in characterizing what is the capacity of this transmission medium in this particular course. So, the output of this channel will be a noisy version of the transmitted way form.

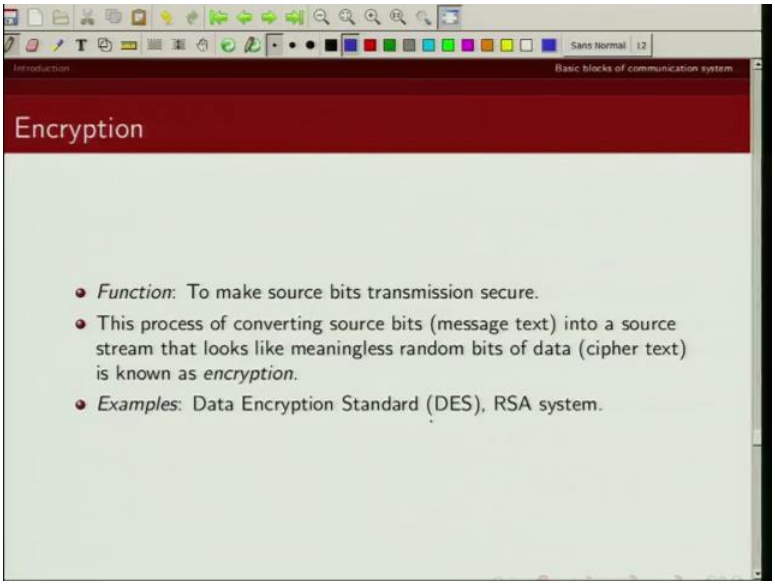
So, after demodulation you would try to correct and detect errors in your message source. So, once you do that give an estimate of your information which now decryption is opposite action of this encryption. So, these decrypted bits comes to the source decoder which will then try to recover back your original sequence from the compressed bits and that is basically delivered to your receiver. So, this is broadly the block diagram of your communication system. In this particular course we are going to spend good amount of time in this source encoder and decoder, and as I said we are going to spend some time in characterizing the capacity of your communication transmission medium which is basically channel.

(Refer Slide Time: 34:10)



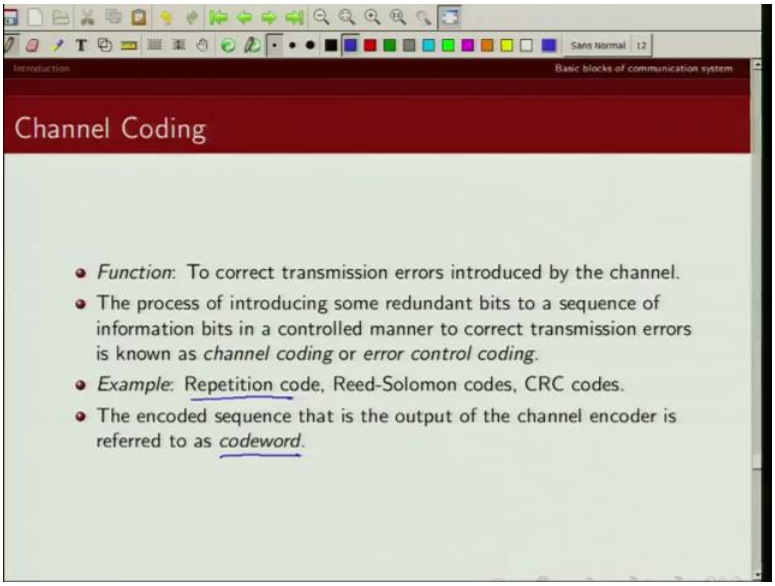
So, let us quickly run through the various functionality of this block of digital communication. So, as I said the idea of source compression is to minimize the number of bits required to represent a source and this process is known as data compression or source coding, and we are going to talk about various examples of source coding in this particular code for example, Huffman code, Lempel-Ziv algorithm the output of the source encoder we normally refer them as information sequence. Encryption as I said, the idea of encryption is to meet the source bit secure and that is idea of this encryption.

(Refer Slide Time: 34:51)



So, we convert this source bits into what looks like a meaningless data which we call as cipher text. So, the process of converting a message text into some random looking text which we call as cipher text is known as encryption. We would not talk about encryption in this particular course some examples of encryption is data encryption standard and RSA system.

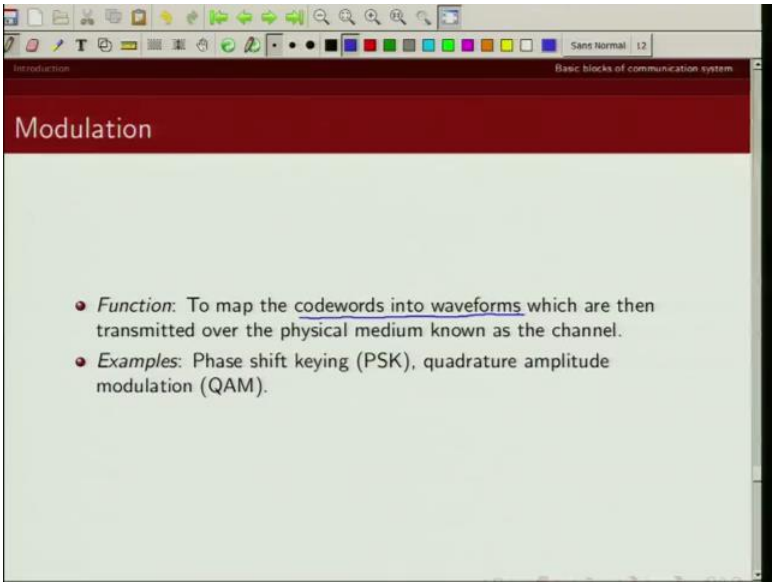
(Refer Slide Time: 35:32)



Channel coding as we said its idea is to introduce redundancy in the message bits and use that redundancy to correct errors. So, whatever transmission errors happen, the idea of this channel encoding is to introduce some additional bits into your message bits and use those redundant bits to correct or detect error. This is also known as channel coding or error control coding. So, as I said we introduce some redundant bits to a sequence of information bits in a controlled manner, this is important because we will have to use those redundant bits to actually detect and correct error.

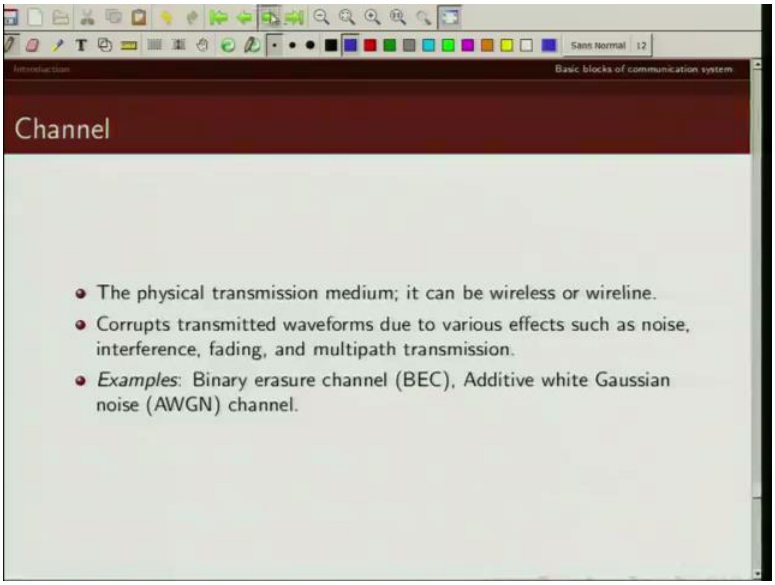
So, when we add these redundant bits we have control over what redundant bits we are adding. Some examples of channel coding schemes are for example, repetition code where in repetition code we repeat what we are sending Reed-Solomon code and cyclic redundancy codes. The output of the channel encoder is referred to as codeword.

(Refer Slide Time: 36:49)



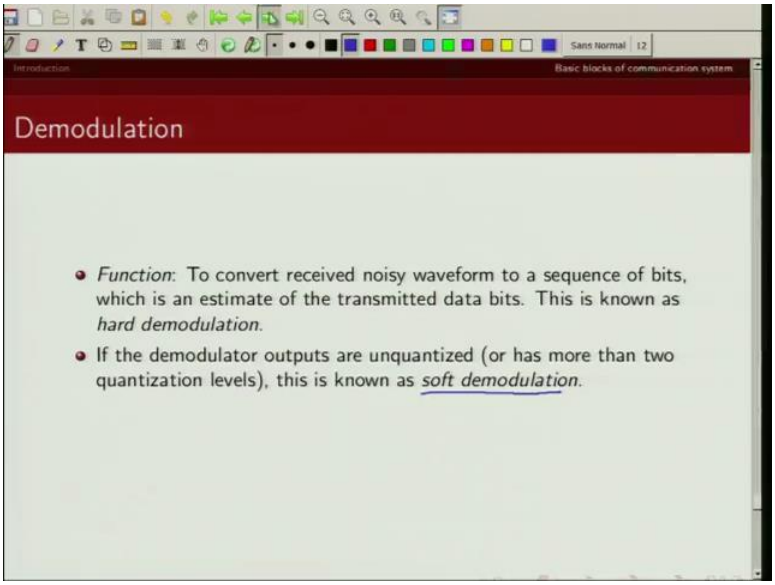
Modulation as a said you want to send these coded sequence over the channel, so we map these code words into waveforms. Some examples of digital modulation schemes are phase shift keying and quadrature amplitude modulation, this you must have studied in a digital communication course.

(Refer Slide Time: 37:03)



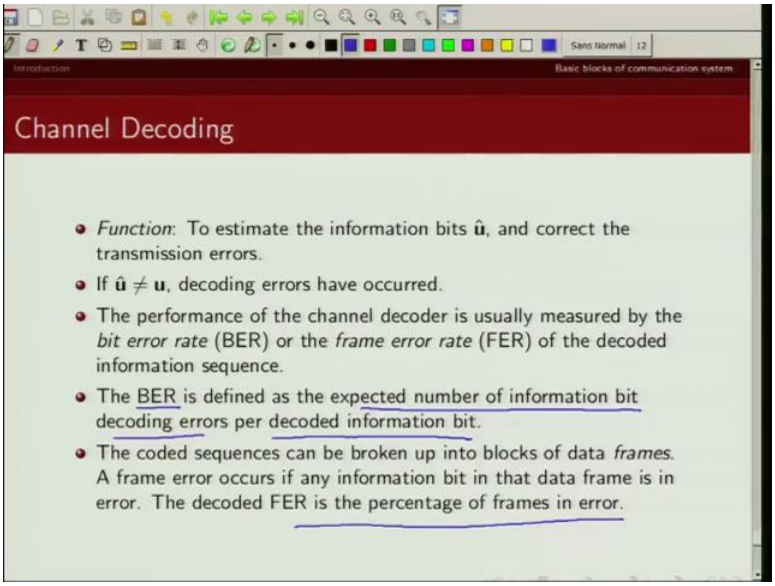
Channel as I said is a medium over which we are sending our waveforms and this medium could be wireless or it could be a wireline channel. Now what does channel does? Channel introduces, channel essentially corrupts the waveforms that we are sending because it add noise their effect of interference from other signals there is effect fading. So, our transmitted waveform will be corrupted by the channel. Some very simple models of channels are these binary erasure channel addictive white Gaussian noise channel, we will talk about them more when we talk about channels and channel capacity - how to computed channels capacity.

(Refer Slide Time: 37:56)



Now, demodulation is basically opposite action of modulation. So, you get noisy waveforms, you want to convert them back into sequence of bits that is basically demodulation. Now, if you convert those noisy bits correctly into 0s and 1s that is known as hard demodulation and instead if you quantize a demodulation output into multiple levels that is basically known as soft demodulation.

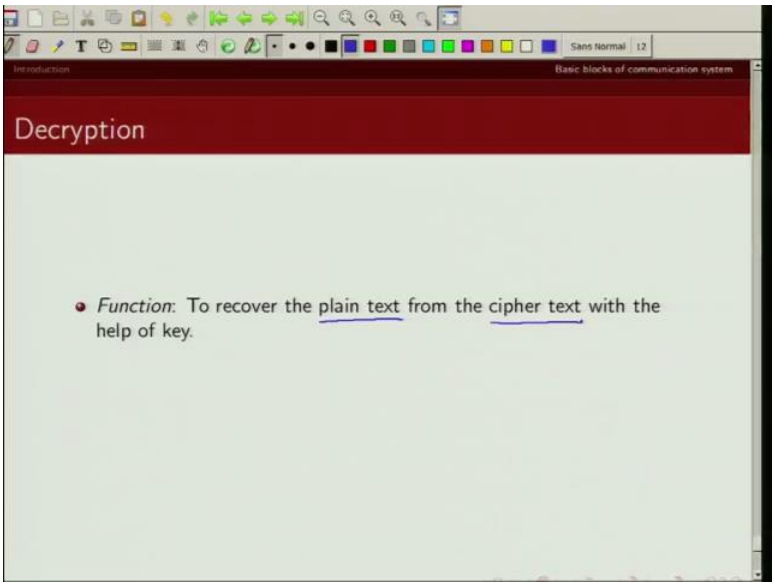
(Refer Slide Time: 38:29)



The objective of this channel decoding is to correct and detect errors which are introduced by the channel. So, that is the objective. So, when does error occur? If you are estimated code sequence is not same as the transmitted code information sequence then errors have occurred and we can find out error using these two matrix bit error rate or frame error rate. So, bit error rate is essentially fraction of transmitted bits that are in error is known as is basically. So, number of error divide by number of bits that we have transmitted is basically your bit error rate and frame is a block of data. So, if there is any error in a block of data we say that the frame is in error.

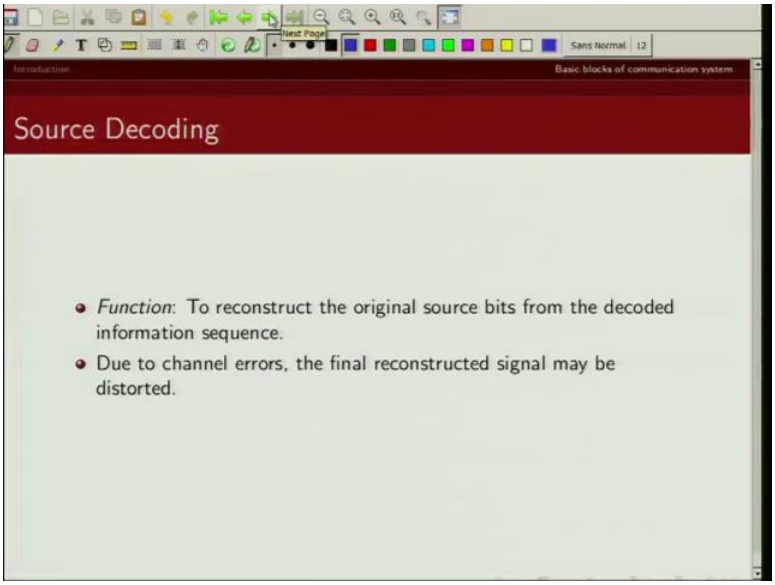
So, number of frames or number blocks which are in error divided by number of blocks that you are transmitting that is basically will be few frame error rate. So, as I have already defined the bit error rate is the expected number of information bits decoding error or decoded information bit and frame error is basically if is a percentage of frames which are in error and frame is a block of data.

(Refer Slide Time: 39:59)



So, this decryption is the opposite action of encryption. So, once you get this cipher text you want to get back your original text that is the objective of the decryption. So, you want to recover the plain text from the cipher text with help of key that is basically your decryption.

(Refer Slide Time: 40:20)



And finally, source decoding, so what you get at the input of the source decoder is compress bits, now you want to recover that original sequence that is done by source decoder. So, you want reconstruct the original source bit from the decoded information sequence that is done by source decoder. Now due to error introduces by the channel there might be some errors introduce in the final reconstructed signals. So, it is possible that your signal may get distorted as a result of that.

So, to conclude basically in this course we are going to concentrate on the source encoder decoder part of digital communication system and in the transmission medium to characterize what sort of rates, what sort of data rates we can achieve over a communication link.

Thank you.