Information Theory and Coding

Course Code: EC716+CS4103

By

Dr. Nagendra Kumar

Assistant Professor

Department of ECE

NIT Jamshedpur

❖Text Books:

➢ **T1.** Information Theory, Inference, and Learning Algorithms: David J.C. MacKay, Cambridge

➢**T2.** Information Theory Coding and Cryptography: Ranjan Bose, McGraw-Hill.

❖ Reference Books:

➢**R1.** Principle of Digital Communication and Coding: Andrew J. Viterbi, Jim K. Omura, McGraw-Hill, United States of America.

➢**R2.** Principle of Communication Engineering: W. Jacobs, John Wiley.

➢**R3.** Information Theory and Reliable Communication: R. Gallager, John Wiley

Objectives

The main objective is to introduce the fundamental limits of communication with practical techniques to realize the limits specified by information theory. The course emphasizes:

- To deeply understand the mathematics of Information Theory and its physical meaning.
- To understand various channel coding techniques.
- Can apply the knowledge to real problems in communication applications.

Introduction

- The purpose of communication system is to carry information bearing base band signals from one place to another placed over a communication channel.
- Information theory is concerned with the fundamental limits of communication.
- It provides some fundamental knowledge to understanding and characterizing the performance of communication systems
- What is the ultimate limit to data compression?

Contd…

- What is the ultimate limit of reliable communication over a noisy channel, e.g. how many bits can be sent in one second over a telephone line?
- Information Theory is a branch of probability theory which may be applied to the study of the communication systems that deals with the mathematical modelling and analysis of a communication system rather than with the physical sources and physical channels.
- Two important elements presented in this theory are Binary Source (BS) and the Binary Symmetric Channel (BSC).
- A binary source is a device that generates one of the two possible symbols ‘0’ and ‘1’ at a given rate ‘r’, measured in symbols per second

Contd…

- These symbols are called bits (binary digits) and are generated randomly.
- The BSC is a medium through which it is possible to transmit one symbol per time unit. However this channel is not reliable and is characterized by error probability ‘p’ $(0 \le p \le 1/2)$ that an output bit can be different from the corresponding input.
- Information theory tries to analyse communication between a transmitter and a receiver through an unreliable channel and in this approach performs an analysis of information sources, especially the amount of information produced by a given source, states the conditions for performing reliable transmission through an unreliable channel.

Contd…

- The source information measure, the channel capacity measure and the coding are all related by one of the Shannon theorems, the channel coding theorem which is stated as: ‘If the information rate of a given source does not exceed the capacity of a given channel then there exists a coding technique that makes possible transmission through this unreliable channel with an arbitrarily low error rate.

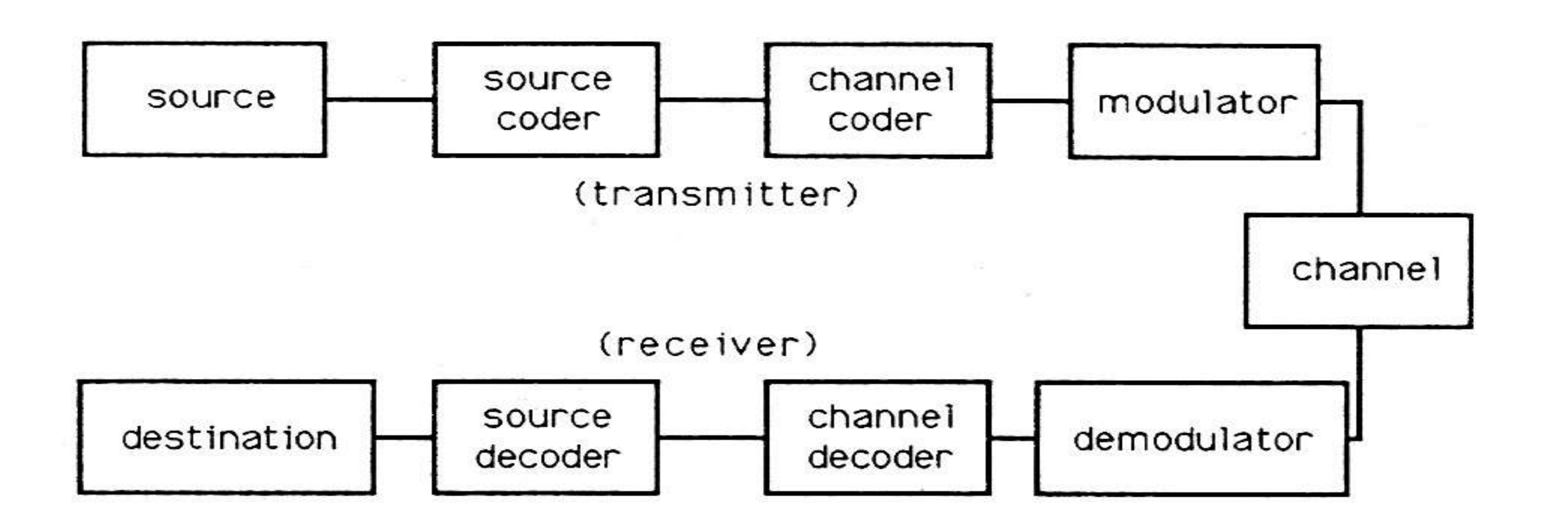
Contd…

- There are three main concepts in this theory:

1. The first is the definition of a quantity that can be a valid measurement of information which should be consistent with a physical understanding of its properties.
2. The second concept deals with the relationship between the information and the source that generates it. This concept will be referred to as the source information. Compression and encryptions are related to this concept.
3. The third concept deals with the relationship between the information and the unreliable channel through which it is going to be transmitted. This concept leads to the definition of a very important parameter called the channel capacity. Error-correction coding is closely related to this concept

Contd…

Digital Communication System



What is Information?

- Information of an event depends only on its probability of occurrence and is not dependent on its content.
- The randomness of happening of an event and the probability of its prediction as a news is known as information.
- The message associated with the leastlikelihood event contains the maximum information.

Axioms of Information:

1. Information is a non-negative quantity: $I(p) \geq 0$.
2. If an event has probability 1, we get no information from the occurrence of the event: $I(1) = 0$.
3. If two independent events occur (whose joint probability is the product of their individual probabilities), then the information we get from observing the events is the sum of the two information: $I(p_1 * p_2) = I(p_1) + I(p_2)$.
4. $I(p)$ is monotonic and continuous in p.

Information Source

- An information source may be viewed as an object which produces an event, the outcome of which is selected at random according to a probability distribution.
- The set of source symbols is called the **source alphabet** and the elements of the set are called **symbols** or **letters**
- Information source can be classified as having memory or being memory-less.
- A source with memory is one for which a current symbol depends on the previous symbols.

Information Source

- A memory-less source is one for which each symbol produced is independent of the previous symbols.
- A discrete memory-less source (DMS) can be characterized by the list of the symbol, the probability assignment of these symbols and the specification of the rate of generating these symbols by the source.

Information Content of a DMS

- ➢The amount of information contained in an event is closely related to its uncertainty.
- ➢A mathematical measure of information should be a function of the probability of the outcome and should satisfy the following axioms

a) Information should be proportional to the uncertainty of an outcome

b) Information contained in independent outcomes should add up

Information Content of a Symbol
(i.e. Logarithmic Measure of Information):

- Let us consider a DMS denoted by ‘x’and having alphabet {x1, x2, ……, xm}.
- The information content of the symbol xi, denoted by I (x_i) is defined by

$$I(x_i) = log_b\left(\frac{1}{p(x_i)}\right) = -log_b\, p(x_i)$$

where p(xi) is the probability of occurrence of symbol xi.

- For any two independent source messages xi and xj with probabilities P_i and P_j respectively and with joint probability P (x_i, x_j) = Pi Pj, the information of the messages is the addition of the information in each message. $I_{ij} = I_i + I_j$.

Contd…

Note that $I(x_i)$ satisfies the following properties.

1. $I(x_i) = 0$ for $P(x_i) = 1$
2. $I(x_i) \geq 0$
3. $I(x_i) > I(x_j)$ if $P(x_i) < P(x_j)$
4. $I(x_i, x_j) = I(x_i) + I(x_j)$ if x_i and x_j are independent

- **Unit of I (x_i):** The unit of I (x_i) is the bit (binary unit) if b = 2, Hartley or decit if b = 10 and nat (natural unit) if b = e. it is standard to use b = 2.

$$\log_2(a) = \frac{\ln(a)}{\ln(2)} = \frac{\log(a)}{\log(2)}$$

Entropy (i.e. Average Information):

- Entropy is a measure of the uncertainty in a random variable. The entropy, H, of a discrete random variable X is a measure of the amount of uncertainty associated with the value of X.
- For quantitative representation of average information per symbol we make the following assumptions:

i) The source is stationary so that the probabilities may remain constant with time.

ii) The successive symbols are statistically independent and come from the source at an average rate of ‘r’ symbols per second.

Entropy (i.e. Average Information):

- The quantity H(X) is called the entropy of source X. it is a measure of the average information content per source symbol.
- The source entropy H(X) can be considered as the average amount of uncertainty within the source X that is resolved by the use of the alphabet.
- $H(X) = E\,[I(x_i)] = -\,\Sigma P(x_i)\,I(x_i) = -\,\Sigma P(x_i)\log_2 P(x_i)$ b/symbol.
- **Entropy for Binary Source:**

$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1 \text{ bit/symbol}$$

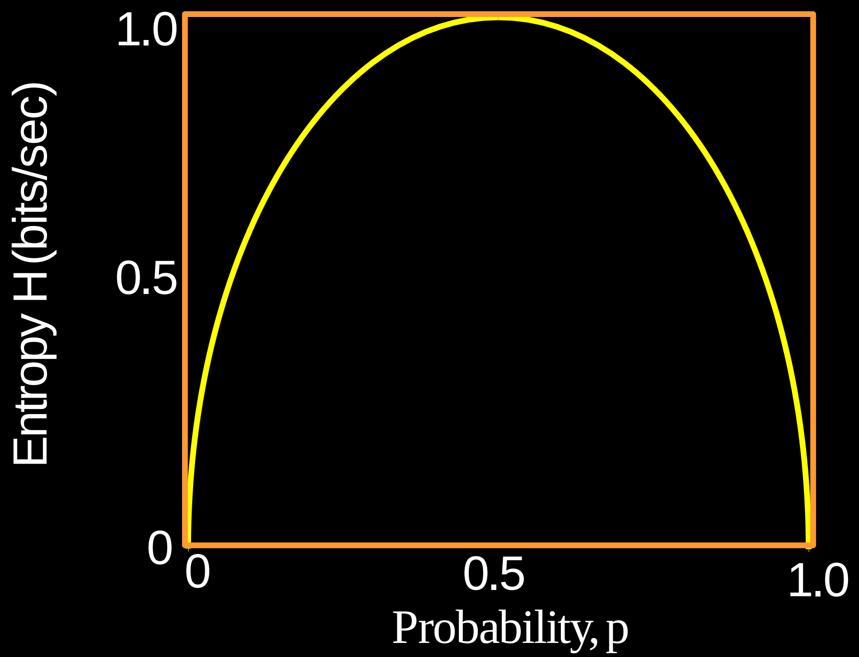
Entropy (i.e. Average Information):

- The source entropy H(X) satisfies the relation: **$0 \leq H(X) \leq \log_2 m$**, where m is the size of the alphabet source X.
- **Properties of Entropy:**

1) $0 \leq H(X) \leq \log_2 m$; m = no. of symbols of the alphabet of source X.
2) When all the events are equally likely, the average uncertainty must have the largest value i.e. $\log_2 m \geq H(X)$
3) H (X) = 0, if all the $P(x_i)$ are zero except for one symbol with P = 1.

Information Rate:

- If the time rate at which X emits symbols is ‘r’ (symbols s), the information rate R of the source is given by
- **R = r H(X) b/s** [(symbols / second) X (information bits/ symbol)].
- R is the information rate. H(X) = Entropy or average information.



Conditional and Joint Entropies

- Using the input probabilities P (x_i), output probabilities P (y_j), transition probabilities P (y_j/x_i) and joint probabilities P (x_i, y_j), various entropy functions for a channel with m inputs and n outputs are defined

$$\mathrm{H(X)} = -\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{P(x_i)log_2p(x_i)}$$

$$\mathrm{H(Y)} = -\sum_{\mathrm{j=1}}^{n} \mathrm{P(y}_j\mathrm{)log_2p(}y_j\mathrm{)}$$

Contd…

$$H(X|Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i|y_j)$$

$$H(Y|X) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(y_i|x_j)$$

$$H(X, |Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i, y_j)$$

Contd…

- H (X) is the average uncertainty of the channel input and H (Y) is the average uncertainty of the channel output.

- The conditional entropy H (X/Y) is a measure of the average uncertainty remaining about the channel input after the channel output has been observed. H (X/Y) is also called equivocation of X w.r.t. Y.

- The conditional entropy H (Y/X) is the average uncertainty of the channel output given that X was transmitted.

Contd…

- The joint entropy H (X, Y) is the average uncertainty of the communication channel as a whole. Few useful relationships among the above various entropies are as under:

a. $H(X, Y) = H(X/Y) + H(Y)$

b. $H(X, Y) = H(Y/X) + H(X)$

c. $H(X, Y) = H(X) + H(Y)$

d. $H(X/Y) = H(X, Y) – H(Y)$

- X and Y are statistically independent

Contd…

- The conditional entropy or conditional uncertainty of X given random variable Y is the average conditional entropy over Y.
- The joint entropy of two discrete random variables X and Y is merely the entropy of their pairing: (X, Y), this implies that if X and Y are independent, then their joint entropy is the sum of their individual entropies.

The Mutual Information:

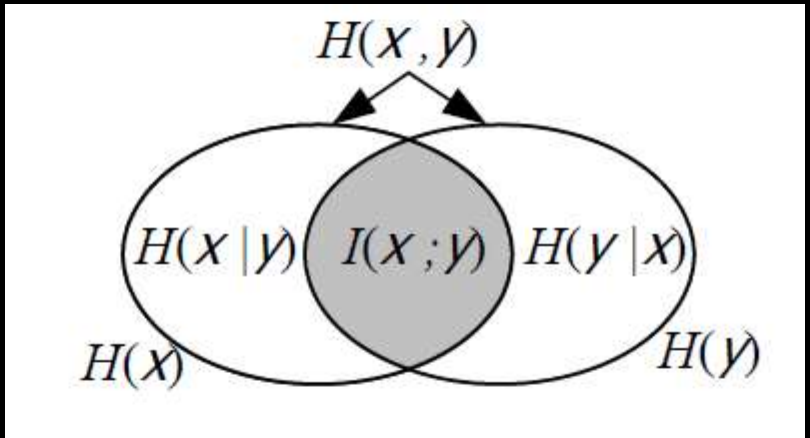
- Mutual information measures the amount of information that can be obtained about one random variable by observing another.
- It is important in communication where it can be used to maximize the amount of information shared between sent and received signals.
- The mutual information denoted by I (X, Y) of a channel is defined by:

$$I(X;Y) = H(X) - H(X|Y) \text{ bits/symbol}$$

- Since H (X) represents the uncertainty about the channel input before the channel output is observed and H (X/Y) represents the uncertainty about the channel input after the channel output is observed, the mutual information I (X; Y) represents the uncertainty about the channel input that is resolved by observing the channel output.

Properties of Mutual Information I (X; Y)

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y/X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- The Entropy corresponding to mutual information [i.e. I (X, Y)] indicates a measure of the information transmitted through a channel. Hence, it is called **'Transferred information'.**



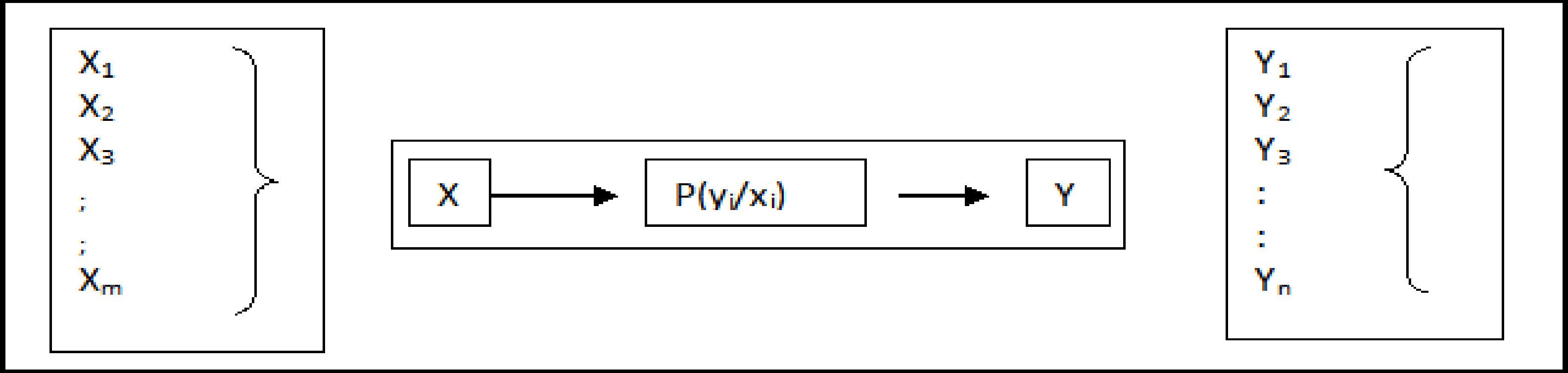
The Discrete Memoryless Channels (DMC):

- **Channel Representation:** A communication channel may be defined as the path or
  medium through which the symbols flow to the receiver end.

- A DMC is a statistical model with an input X and output Y. Each possible input to output path is indicated along with a conditional probability P ($y_j|x_i$), where P ($y_j|x_i$) is the conditional probability of obtaining output y_j given that the input is x_1 and is called a **channel transition probability.**

- A channel is completely specified by the complete set of transition probabilities. The channel is specified by the matrix of transition probabilities [P(Y|X)]. This matrix is known as **Channel Matrix**.

Contd…



$$P(Y|X) = \begin{bmatrix} P(y_1/x_1) & \cdots & P(y_n/x_1) \\ \vdots & \ddots & \vdots \\ P(y_1/x_m) & \cdots & P(y_n/x_m) \end{bmatrix}$$

Contd…

- Since each input to the channel results in some output, each row of the column matrix must sum to unity. This means that

$$\sum_{j=1}^{n} P\left(y_j \middle| x_i\right) = 1 \text{ for all } i$$

- Now, if the input probabilities P(X) are represented by the row matrix, we have

$$[P(X)] = [P(x_1)P(x_2) \ldots P(x_m)]$$

- Also the output probabilities P(Y) are represented by the row matrix, we have

$$P(Y)=[P(y_1)\ P(y_2)\ \ldots.P(y_n)]$$

Contd…

Then

$$[P(Y)] = [P(X)][P(Y|X)]$$

Now if P(X) is represented as a diagonal matrix, we have

$$[P(X)]_d = \begin{bmatrix} P(x_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P(x_m) \end{bmatrix}$$

Then

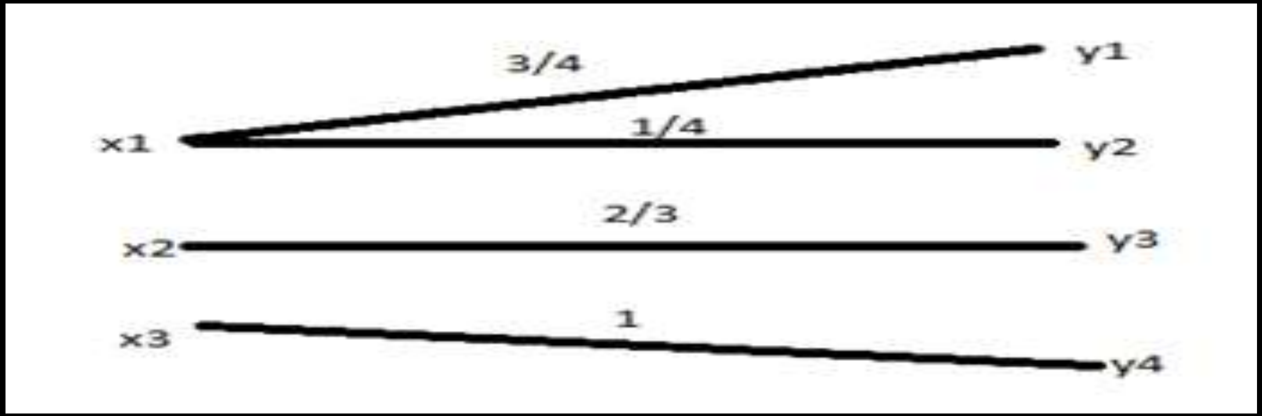
$$[P(X,Y)] = [P(X)]_d[P(Y|X)]$$

- Where the (i, j) element of matrix [P(X,Y)] has the form P(x_i, y_j).
- The matrix [P(X, Y)] is known as the **joint probability matrix** and the element P(x_i, y_j) is the joint probability of transmitting x_i and receiving y_j.

Types of Channels:

- ➢Other than discrete and continuous channels, there are some special types of channels with their own channel matrices. They are as follows:
- ➢**Lossless Channel:** A channel described by a channel matrix with only one non–zero element in each column is called a lossless channel.

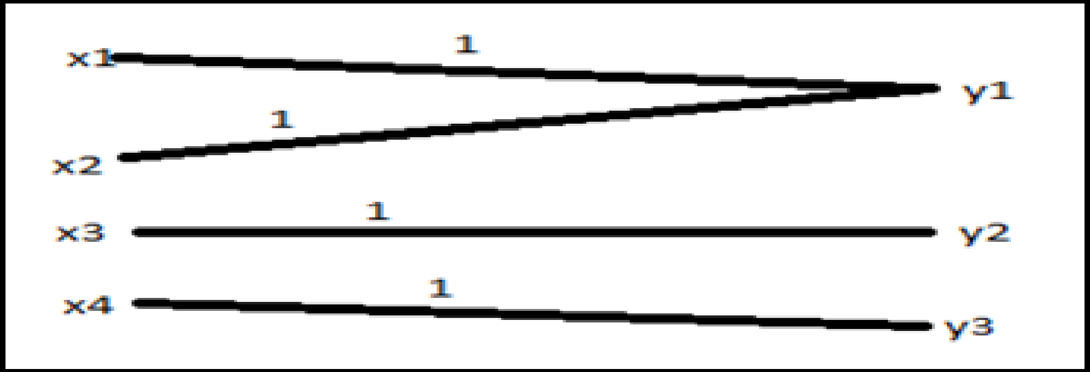
$$P(Y|X) = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 \\ 0 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Contd…

- **Deterministic Channel:** A channel described by a channel matrix with only one non – zero element in each row is called a deterministic channel.

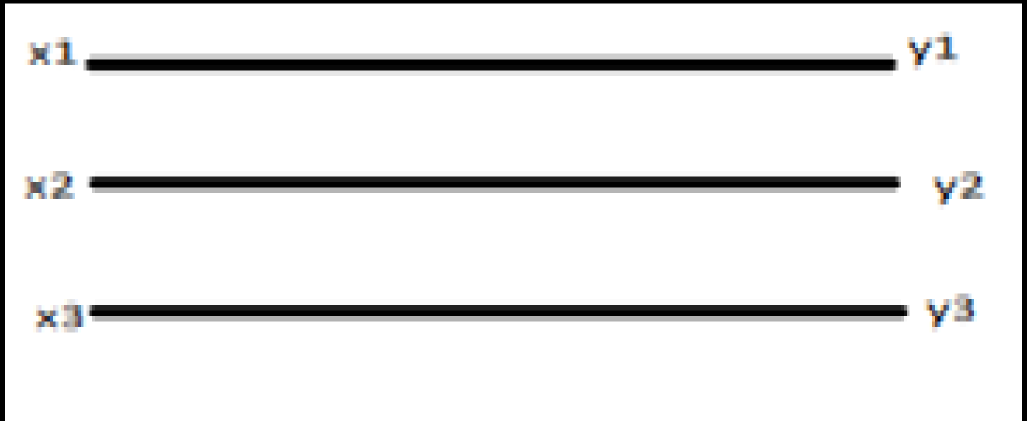
$$P(Y|X) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Contd…

- **Noiseless Channel:** A channel is called noiseless if it is both lossless and deterministic. For a lossless channel, m = n

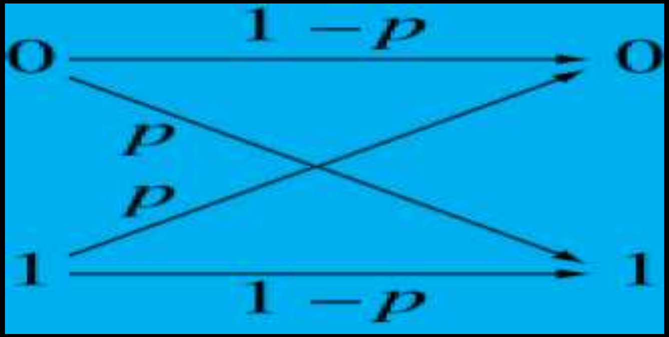
$$P(Y|X) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Contd…

- **Binary Symmetric Channel:** BSC has two inputs ($x_1 = 0$ and $x_2 = 1$) and two outputs ($y_1 = 0$ and $y_2 = 1$).
- This channel is symmetric because the probability of receiving a 1 if a 0 is sent is the same as the probability of receiving a 0 if a 1 is sent.

$$P(Y|X) = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$



The Channel Capacity

- The channel capacity represents the maximum amount of information that can be transmitted by a channel per second.
- To achieve this rate of transmission, the information has to be processed properly or coded in the most efficient manner.
- **Channel Capacity per Symbol C_S:** The channel capacity per symbol of a discrete memory-less channel (DMC) is defined as

$$C_s = \max_{\{P(x_i)\}} I(X;Y) \text{ bits/symbol}$$

Where the maximization is over all possible input probability distributions $\{P(x_i)\}$ on X.

Contd…

- **Channel Capacity per Second C:** I f ‘r’ symbols are being transmitted per second, then the maximum rate or transmission of information per second is ‘r C_S’. this is the channel capacity per second and is denoted by C (b/s) i.e.

$$C = rC_S \, b/s$$

Capacities of Special Channels:

- **Lossless Channel:** For a lossless channel, H (X/Y) = 0 and I (X; Y) = H (X).
- Thus the mutual information is equal to the input entropy and no source information is lost in transmission.

$$C_S = \max_{\{P(x_i)\}} H(X) = \log_2 m$$

Where m is the number of symbols in X.

Contd…

- **Deterministic Channel:** For a deterministic channel,

  $H(Y/X) = 0$ for all input distributions $P(x_i)$ and $I(X;Y) = H(Y)$.

- Thus the information transfer is equal to the output entropy. The channel capacity per symbol will be

$$C_S = \max_{\{P(x_i)\}} H(Y) = \log_2 n$$

where n is the number of symbols in Y.

Contd…

- **Noiseless Channel:** since a noiseless channel is both lossless and deterministic, we have I (X; Y) = H (X) = H (Y) and the channel capacity per symbol is

$$C_S = \log_2 m = \log_2 n$$

- **Binary Symmetric Channel:** For the BSC, the mutual information is

$$I(X;Y) = H(Y) + p\log_2 p + (1-p)\log_2(1-p)$$

- And the channel capacity per symbol will be

$$C_S = 1 + p\log_2 p + (1-p)\log_2(1-p)$$

Capacity of an Additive Gaussian Noise (AWGN) Channel - Shannon – Hartley Law

- The Shannon – Hartley law underscores the fundamental role of bandwidth and signal – to – noise ration in communication channel. It also shows that we can exchange increased bandwidth for decreased signal power for a system with given capacity C.

- In an additive white Gaussian noise (AWGN) channel, the channel output Y is given by

$$Y = X + n$$

- Where X is the channel input and n is an additive bandlimited white Gaussian noise for zero mean and variance σ^2.

Contd…

- The capacity C of an AWGN channel is given by

$$C_s = \max_{\{P(x_i)\}} I(X;\ Y) = 1/2\log_2(1+\mathrm{S/N})\ \text{bit/sample}$$

- Where S/N is the signal – to – noise ratio at the channel output.
- If the channel bandwidth B Hz is fixed, then the output y(t) is also a bandlimited signal completely characterized by its periodic sample values taken at the Nyquist rate 2B samples/s.
- Then the capacity C (b/s) of the AWGN channel is limited by

$$C = 2B * C_S = B\log_2(1+\mathrm{S/N})\, b/s$$

- This above equation is known as the **Shannon – Hartley Law.**

Channel Capacity: Shannon – Hartley Law Proof.

- The bandwidth and the noise power place a restriction upon the rate of information that can be transmitted by a channel. Channel capacity C for an AWGN channel is expressed as

$$C = Blog_2(1 + S/N)$$

Where B = channel bandwidth in Hz; S = signal power; N = noise power;

Contd…

- **Proof:** Assuming signal mixed with noise, the signal amplitude can be recognized only within the root mean square noise voltage.
- Assuming average signal power and noise power to be S watts and N watts, respectively, the RMS value of the received signal is $\sqrt{S+N}$ and that of noise is $\sqrt{N}$
- Therefore the number of distinct levels that can be distinguished without error is expressed as

$$M = \frac{\sqrt{S+N}}{\sqrt{N}} = \sqrt{1 + \mathrm{S/N}}$$

Contd…

- The maximum amount of information carried by each pulse having $\sqrt{1+\mathrm{S}/\mathrm{N}}$ distinct levels is given by

$$I = \log_2\left(\sqrt{1+\frac{\mathrm{S}}{\mathrm{N}}}\right) = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \text{ bits}$$

- The channel capacity is the maximum amount of information that can be transmitted per second by a channel. If a channel can transmit a maximum of K pulses per second, then the channel capacity C is given by

$$\mathrm{C} = \frac{K}{2}\log_2\left(1+\frac{S}{N}\right) \text{ bits/Second}$$

Contd…

- A system of bandwidth nf_m Hz can transmit $2nf_m$ independent pulses per second. It is concluded that a system with bandwidth B Hz can transmit a maximum of 2B pulses per second. Replacing K with 2B, we eventually get

$$\mathrm{C} = B\log_2\left(1+\frac{S}{N}\right)\ \text{bits/Second}$$

- The bandwidth and the signal power can be exchanged for one another.

Differential Entropy

- The differential entropy h(X) of a continuous random variable X with probability density function $f_X(x)$ is defined as

$$h(X) = \int_{-\infty}^{\infty} f_X(x) log_2\left(\frac{1}{f_X(x)}\right) dx$$

- Differential entropy for a random variable with uniform probability density function

$$f_X(x) = \begin{cases} \frac{1}{a}, 0 \leq x \leq a \\ 0, \text{otherwise} \end{cases}$$

Contd…

- Further,

$$h(X) = \int_{0}^{a} \frac{1}{a} log_2(a) dx = log_2(a)$$

Unlike entropy differential entropy can be negative

Contd…

- Differential entropy of Gaussian Source with mean μ and variance σ^2: One of most commonly occurring and practically relevant sources

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

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

- Differential entropy increases with variance σ^2: As uncertainty increases
- It does not depend on mean

Joint/Conditional Differential Entropy:

- $\boldsymbol{f_Y(y)}$and $\boldsymbol{f_X(x)}$ are marginal PDF of X and Y, then for the joint PDF $f_{XY}(x,y)$, we can write joint entropy and conditional entropy, respectively, as

$$h(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) log_2\left(\frac{1}{f_{XY}(x,y)}\right) dxdy$$

$$h(x|y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) log_2\left(\frac{1}{f_{X|Y}(x|y)}\right) dxdy$$

Source Coding

Source Coding:

- **Definition:** A conversion of the output of a discrete memory less source (DMS) into a sequence of binary symbols i.e. binary code word, is called **Source Coding**.
- The device that performs this conversion is called the **Source Encoder**.
- **Objective of Source Coding:** An objective of source coding is to minimize the average bit rate required for representation of the source by reducing the redundancy of the information source

Few Terms Related to Source Coding Process:

1. Code word Length:

- Let X be a DMS with finite entropy H (X) and an alphabet $\{x_1 \ldots \ldots . . x_m\}$ with corresponding probabilities of occurrence P(x_i) (i = 0, …. , M-1). Let the binary code word assigned to symbol x_i by the encoder have length n_i, measured in bits. The length of the code word is the number of binary digits in the code word.

2. Average Code word Length:

- The average code word length L, per source symbol is given by

$$\bar{L} = \sum_{\mathrm{i}=0}^{\mathrm{M}-1} \mathrm{p}(\mathrm{x_i}) n_i$$

- The parameter $\bar{L}$ represents the average number of bits per source symbol used in the

  source coding process.

Contd…

1. **Code Efficiency:**

- ➢ The code efficiency η is defined as

$$\eta = \frac{L_{min}}{\bar{L}}$$

2. Code Redundancy:

- ➢ The code redundancy γ is defined as

$$\boldsymbol{\gamma = 1 - \eta}$$

The Source Coding Theorem

- The source coding theorem states that for a DMS X, with entropy H (X), the average code word length $\bar{L}$ per symbol is bounded as $\bar{L} \geq H(X)$
- And further, $\bar{L}$ can be made as close to H (X) as desired for some suitable chosen code
- Thus,

$$\bar{L}_{min} = H(X)$$

- The code efficiency can be rewritten as

$$\eta = \frac{H(X)}{\bar{L}}$$

Classification of Code

1. Fixed – Length Codes
2. Variable – Length Codes
3. Distinct Codes
4. Prefix – Free Codes
5. Uniquely Decodable Codes
6. Instantaneous Codes
7. Optimal Codes

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

Contd…

1. **Fixed – Length Codes:**

   A fixed – length code is one whose code word length is fixed. Code 1 and Code 2 of above table are fixed – length code words with length 2.

2. **Variable – Length Codes:**

   A variable – length code is one whose code word length is not fixed. All codes of above table except Code 1 and Code 2 are variable – length codes.

3. **Distinct Codes:**

   A code is distinct if each code word is distinguishable from each other. All codes of above table except Code 1 are distinct codes.

Contd…

4. **Prefix – Free Codes:**

A code in which no code word can be formed by adding code symbols to another code word is called a prefix- free code. In a prefix – free code, no code word is prefix of another. Codes 2, 4 and 6 of above table are prefix – free codes.

5. **Uniquely Decodable Codes:**

A distinct code is uniquely decodable if the original source sequence can be reconstructed perfectly from the encoded binary sequence. A sufficient condition to ensure that a code is uniquely decodable is that no code word is a prefix of another. Thus the prefix – free codes 2, 4 and 6 are uniquely decodable codes. Prefix – free condition is not a necessary condition for uniquely decidability. Code 5 albeit does not satisfy the prefix – free condition and yet it is a uniquely decodable code since the bit 0 indicates the beginning of each code word of the code.

Contd…

6. Instantaneous Codes:

A uniquely decodable code is called an instantaneous code if the end of any code word is recognizable without examining subsequent code symbols. The instantaneous codes have the property previously mentioned that no code word is a prefix of another code word. Prefix – free codes are sometimes known as instantaneous codes.

7. **Optimal Codes:**

A code is said to be optimal if it is instantaneous and has the minimum average L for a given source with a given probability assignment for the source symbols

Kraft Inequality

- Let X be a DMS with alphabet $\{x_i\}(i=0,2,\ldots,M\text{-}1)$. Assume that the length of the assigned binary code word corresponding to x_i is n_i.
- A necessary and sufficient condition for the existence of an instantaneous binary code is

$$K=\sum_{i=0}^{M-1} 2^{-n_i} \leq 1$$

- This is known as the **Kraft Inequality**
- It may be noted that Kraft inequality assures us of the existence of an instantaneously decodable code with code word lengths that satisfy the inequality.
- But it does not show us how to obtain those code words, nor does it say any code satisfies the inequality is automatically uniquely decodable.

Entropy Coding

- The design of a variable – length code such that its average code word length approaches the entropy of DMS is often referred to as **Entropy Coding**.

- There are basically two types of entropy coding, viz.

1) **Shannon – Fano Coding**
2) **Huffman Coding**

Shannon – Fano Coding:

- An efficient code can be obtained by the following simple procedure, known as **Shannon–Fano algorithm.**

1) List the source symbols in order of decreasing probability.
2) Partition the set into two sets that are as close to equi-probables as possible and assign 0 to the upper set and 1 to the lower set.
3) Continue this process, each time partitioning the sets with as nearly equal probabilities as possible until further partitioning is not possible
4) Assign code word by appending the 0s and 1s from left to right

Shannon – Fano Coding - Example

- Let there be six (6) source symbols having probabilities as $x_1 = 0.30$, $x_2 = 0.25$, $x_3 = 0.20$, $x_4 = 0.12$, $x_5 = 0.08$ $x_6 = 0.05$. Obtain the Shannon – Fano Coding for the given source symbols.
- Shannon Fano Code words

x_i	P(x_i)	Step 1	Step 2	Step 3	Step 4	Code
x_1	0.30	0	0			00
x_2	0.25	0	1			01
x_3	0.20	1	0			10
x_4	0.12	1	1	0		110
x_5	0.08	1	1	1	0	1110
x_6	0.05	1	1	1	1	1111

- **H (X) = 2.36 b/symbol**
- **$\overline{L}$ = 2.38 b/symbol**
- **η = H (X)/ $\overline{L}$ = 0.99**

Huffman Coding:

- Huffman coding results in an optimal code. It is the code that has the highest efficiency.
- The Huffman coding procedure is as follows:

1) List the source symbols in order of decreasing probability.
2) Combine the probabilities of the two symbols having the lowest probabilities and reorder the resultant probabilities, this step is called reduction 1. The same procedure is repeated until there are two ordered probabilities remaining.

Contd…

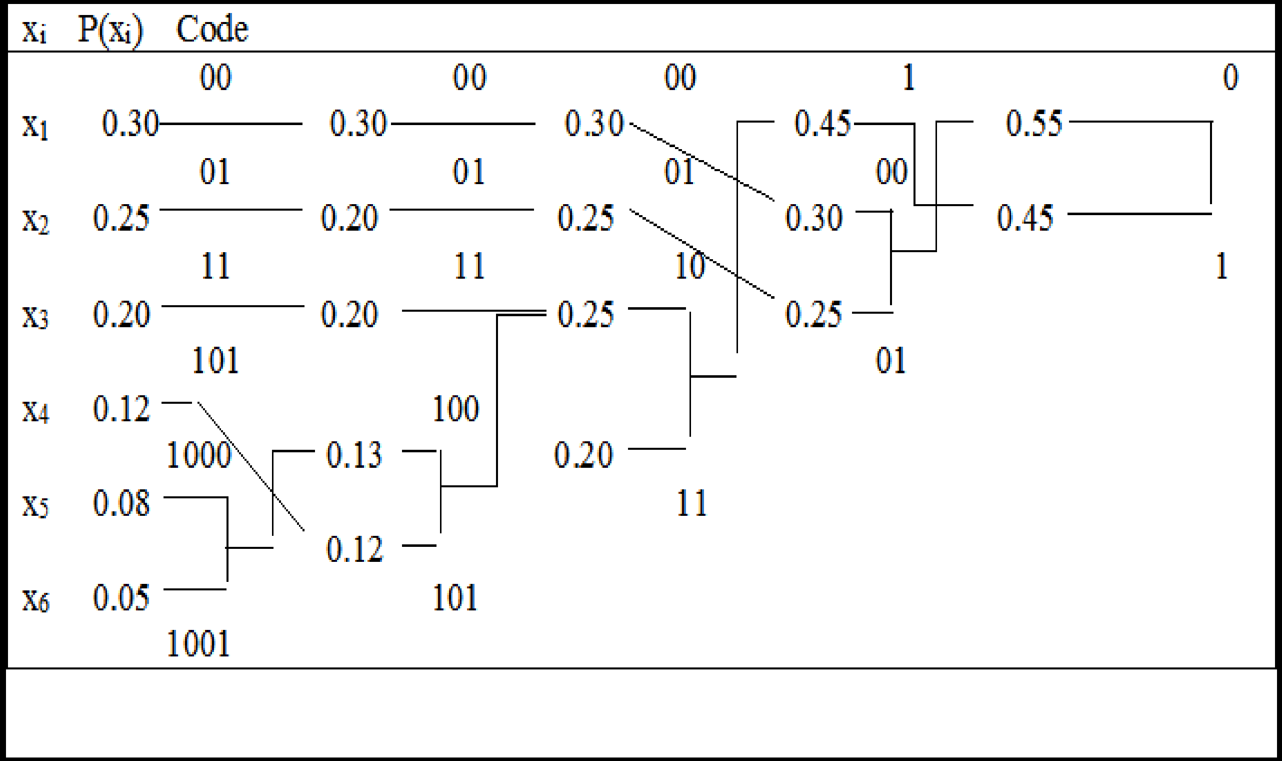
3) Start encoding with the last reduction, which consists of exactly two ordered probabilities. Assign 0 as the first digit in the code word for all the source symbols associated with the first probability; assign 1 to the second probability.

4) Now go back and assign 0 and 1 to the second digit for the two probabilities that were combined in the previous reduction step, retaining all the source symbols associated with the first probability; assign 1 to the second probability.

5) Keep regressing this way until the first column is reached.

6) The code word is obtained tracing back from right to left.

Huffman Encoding - Example



Source Sample x_i	P (x_i)	Codeword
X_1	0.30	00
X_2	0.25	01
X_3	0.20	11
X_4	0.12	101
x_5	0.08	1000
x_6	0.05	1001

H (X) = 2.36 b/symbol

$\bar{L}$ = 2.38 b/symbol

η = H (X)/ $\bar{L}$ = 0.99

Redundancy:

- Redundancy in information theory refers to the reduction in information content of a message from its maximum value
- For example, consider English having 26 alphabets. Assuming all alphabets are equally likely to occur, P (x_i) = 1/26. For all the 26 letters, the information contained is therefore

$$\log_2 26 = 4.7 \ \textit{bits/letter}$$

- Assuming that each letter to occur with equal probability is not correct, if we assume that some letters are more likely to occur than others, it actually reduces the information content in English from its maximum value of 4.7 bits/symbol.
- We define relative entropy on the ratio of H (Y/X) to H (X) which gives the maximum compression value and Redundancy is then expressed as

$$\text{Redundancy} = \mathrm{H}(\mathrm{Y}|\mathrm{X})/\mathrm{H}(\mathrm{X})$$

Channel Coding

Why channel coding?

- The challenge in digital communication system is that of providing a cost effective facility for transmitting Information , at a rate and a level of reliability and quality that are acceptable to a user at the receiver.
- The two key system parameters :

1) Transmitted signal power.
2) Channel bandwidth.

- Power spectral density of receiver noise (Important parameter)
- These parameters determine the signal energy per bit-to-noise power spectral density ratio Eb/No.

Contd…

- In practical , there are a limit on the value that we can assign to Eb /No.

  So, it is impossible to provide acceptable data quality (i.e. , low enough error performance).

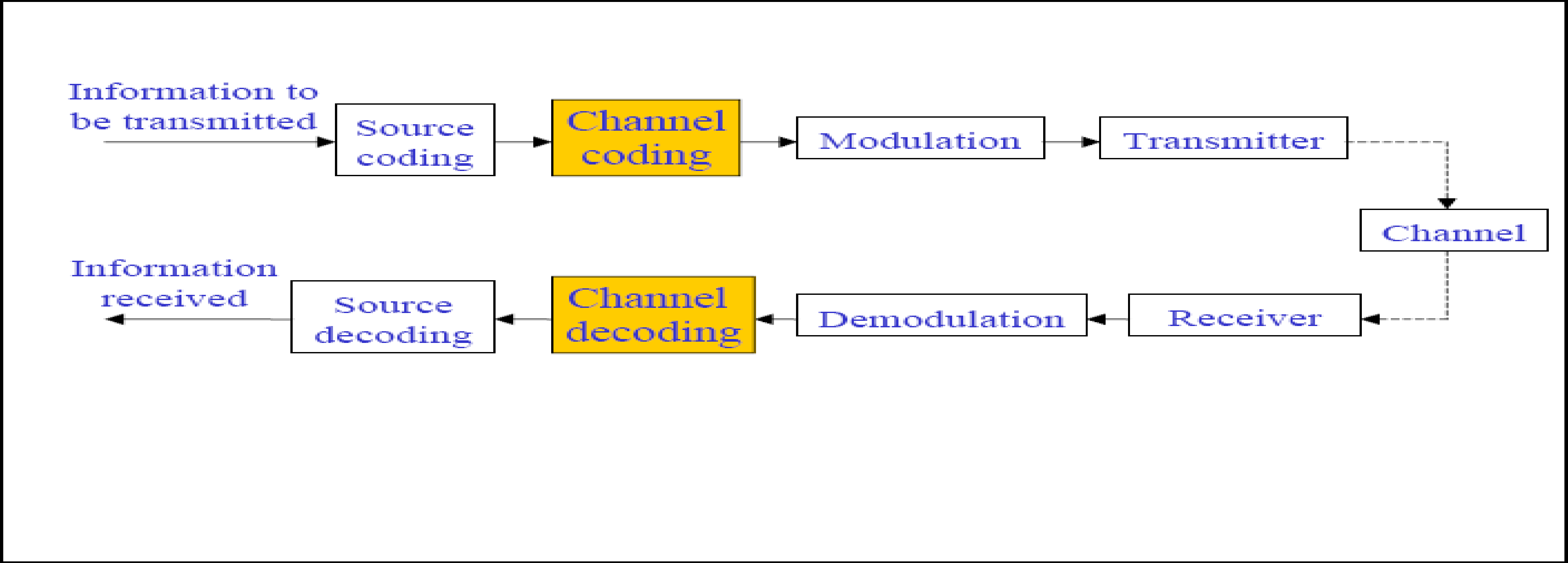
- For a fixed Eb/No , the only practical option available for changing data quality is to use ERROR-CONTROL CODING.

- **The two main methods of error control are:**

i. Forward Error Correction (FEC).

ii. Automatic Repeat request (ARQ).

CHANNEL CODING Block Diagram



Forward Error Correction (FEC)

- The key idea of FEC is to transmit enough redundant data to allow receiver to recover from errors all by itself. No sender retransmission required.
- The major categories of FEC codes are

i. Block codes,

ii. Cyclic codes

iii. BCH codes

iv. Reed-Solomon codes .

v. Convolutional codes

Forward Error Correction (FEC)

➢FEC require only a one-way link between the transmitter and receiver.

➢In the use of error-control coding there are trade offs between:

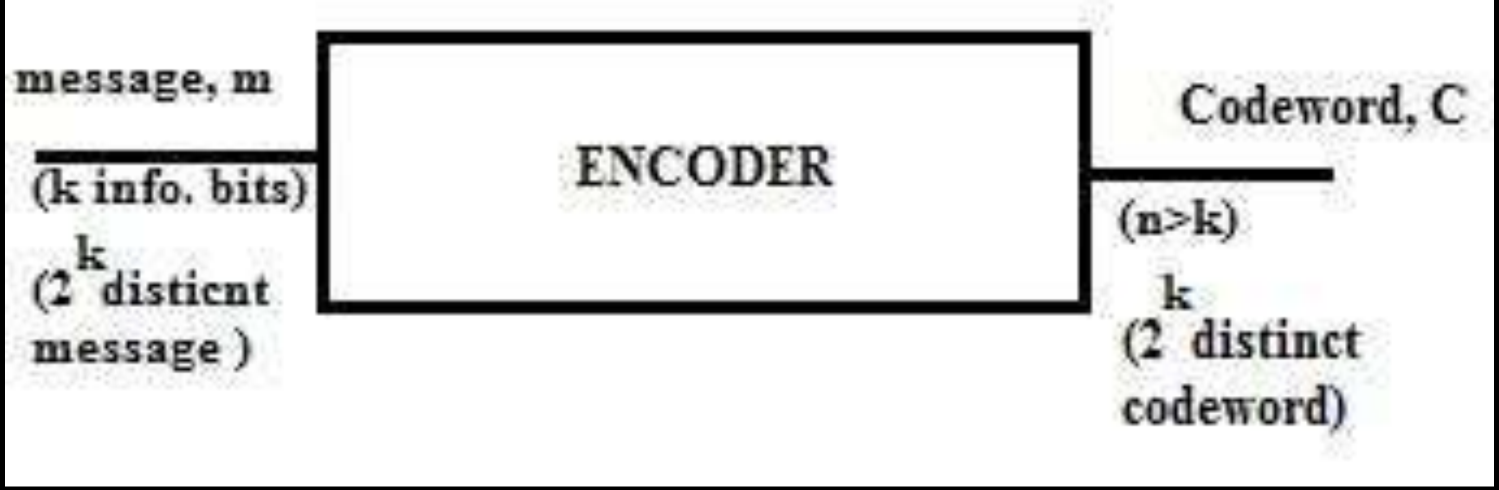
i. Efficiency & Reliability.

ii.Encoding/Decoding complexity & Bandwidth .

Channel Coding Theorem

- The channel coding theorem states that if a discrete memoryless channel has capacity C and the source generate info at rate less than C ,then there exists a coding technique that the output of the source may be transmitted over the channel with an arbitrarily low probability of symbol error.
- For the special case of BSC the theorem tell us that it is possible to find a code that achieves error free transmission over the channel.
- The issue that mater not the signal to noise ratio put how the channel input is encoded.
- The theorem asserts the existence of good codes but dose not tell us how to find them.
- By good codes we mean families of channel codes that are capable of providing reliable (error-free) transmission of info over a noisy channel of interest at bit rate up to a max value less than the capacity of the channel.

Linear Block Codes

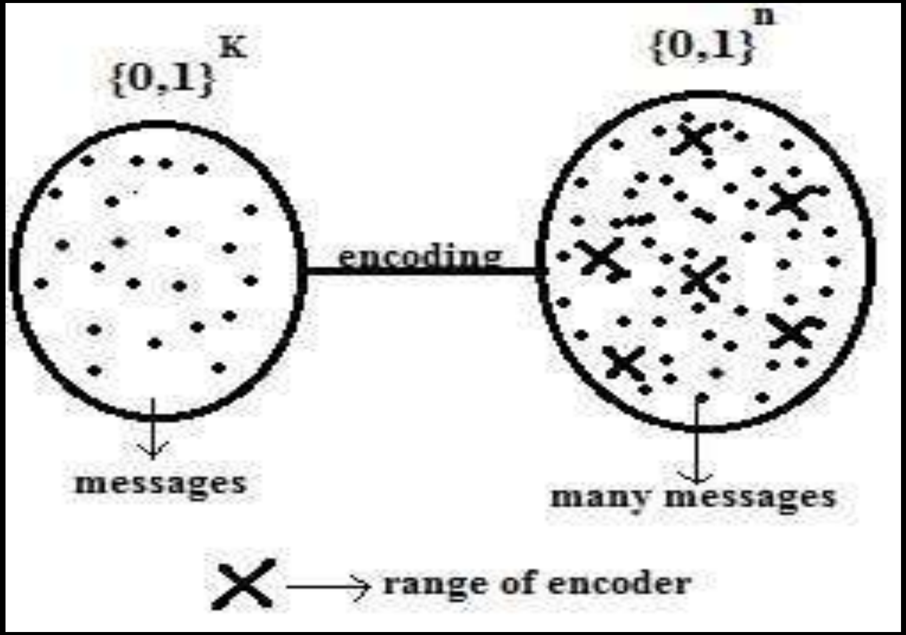


- The encoder generates a block of n coded bits from k information bits and we call this as (n, k) block codes. The coded bits are also called as code word symbols.

Why linear?

- A code is linear if the modulo-2 sum of two code words is also a code word.

Contd…



- ‘n’ code word symbols can take 2^n possible values. From that we select 2^k code words to form the code.
- A block code is said to be useful when there is one to one mapping between message m and its code word c as shown above.

Generator Matrix

- All code words can be obtained as linear combination of basis vectors.
- The basis vectors can be designated as $\{g_1, g_2, g_3, \ldots.., g_k\}$
- For a linear code, there exists a k by n generator matrix such that

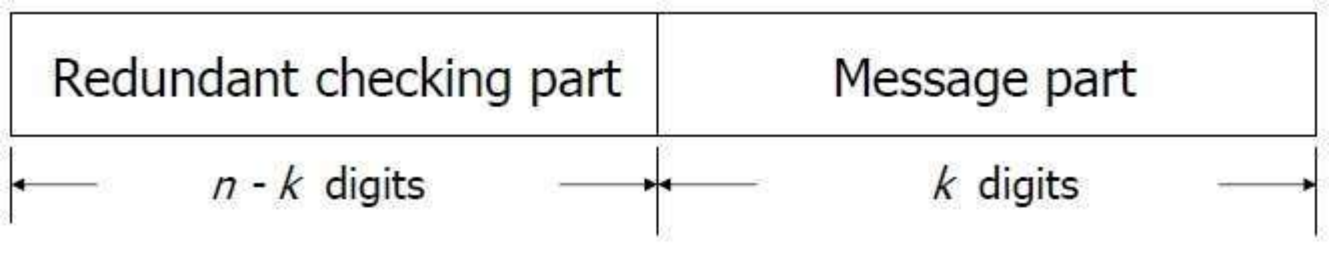
$$c_{1*n} = m_{1*k}\,.\,G_{k*n}$$

where c=$\{c_1, c_2, \ldots.., c_n\}$ and m=$\{m_1, m_2, \ldots\ldots., m_k\}$

$$G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & & \vdots \\ g_{k1} & g_{k2} & \cdots & g_{kn} \end{bmatrix}$$

Block Codes in Systematic Form

- In this form, the code word consists of (n-k) parity check bits followed by k bits of the message.
- The structure of the code word in systematic form is:



- The rate or efficiency for this code R= k/n

Contd…

$$G = [\, I_k \;\; P]$$

$$C = m.G = [\underbrace{m}_{\text{Message part}}\ \underbrace{mP}_{\text{Parity part}}]$$

Example:

- ➢ Let us consider (7, 4) linear code where k=4 and n=7

$$m=(1110) \text{ and } G = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} 1\,1\,0\,1\,0\,0\,0 \\ 0\,1\,1\,0\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,0 \\ 1\,0\,1\,0\,0\,0\,1 \end{bmatrix}$$

Contd…

C= m.G = $\boldsymbol{m_1 g_1} + \boldsymbol{m_2 g_2} + \boldsymbol{m_3 g_3} + \boldsymbol{m_4 g_4}$

= 1.$\boldsymbol{g_1}$ + $\mathbf{1}$. $\boldsymbol{g_2}$ + $\mathbf{1}$. $\boldsymbol{g_3}$ + $\mathbf{0}$. g_4

C = (1101000) +(0110100) + (1110010)

= (0101110)

Contd…

- **Another method**:

Let m=(m_1, m_2, m_3, m_4) and c= $(c_1, c_2, c_3, c_4, c_5, c_6, c_7)$

$$c=m.G= (m_1, m_2, m_3, m_4) \begin{bmatrix} 1\,1\,0\,1\,0\,0\,0 \\ 0\,1\,1\,0\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,0 \\ 1\,0\,1\,0\,0\,0\,1 \end{bmatrix}$$

- By matrix multiplication we obtain :

$c_1 = m_1 + m_3 + m_4$, $c_2 = m_1 + m_2 + m_3$, $c_3 = m_2 + m_3 + m_4$,
$c_4 = m_1$,
$c_5 = m_2$, $c_6 = m_3$, $c_7 = m_4$

The code word corresponding to the message (1110) is (0101110) .

Parity Check Matrix (H)

- When G is systematic, it is easy to determine the parity check matrix H as

$$H = [I_{n-k} \quad P^T]$$

- The parity check matrix H of a generator matrix is an (n-k)-by-n matrix satisfying

$$H_{(n-k)*n} G^T_{n*k} = 0$$

- Then the code words should satisfy (n-k) parity check equations

$$C_{1*n} H_{n*(n-k)} = m_{1*k} G_{k*n} H^T_{n*(n-k)} = 0$$

Example:

Consider generator matrix of (7, 4) linear block code

H = [$I_{n-k}\,P^T$]   and   G = [$P \quad I_k$]

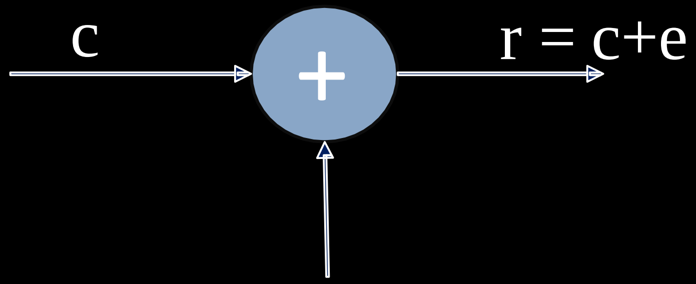
- ➢The corresponding parity check matrix is

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$G.H^T == \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = 0$$

Syndrome and Error Detection

- For a code word c, transmitted over a noisy channel, let r be the received vector at the output of the channel with error e



$$e_i = \begin{cases} 1, \text{ if } r \neq c \\ 0, \text{ if } r = c \end{cases}$$

Syndrome of received vector r is given by:

$$s = r.H^T = (s_1, s_2, s_3, \ldots \ldots \ldots, s_{n-k})$$

Properties of syndrome:

- The syndrome depends only on the error pattern and not on the transmitted word.

$$s = (c+e).H^T = c.H^T + e.H^T = e.H^T$$

- All the error pattern differ by at least a code word have the same syndrome ‘s’.

Example:

Let C=(0101110) be the transmitted code and r=(0001110) be the received vector.

s=r. H^T=$[s_1, s_2, s_3]$ =$[r_1, r_2, r_3, r_4, r_5, r_6, r_7]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

- ➢The syndrome digits are:

$$s_1 = r_1 + r_4 + r_6 + r_7 = 0$$

$$s_2 = r_2 + r_4 + r_5 + r_6 = 1$$

$$s_3 = r_3 + r_5 + r_6 + r_7 = 0$$

Contd…

The error vector, e=$(e_1, e_2, e_3, e_4, e_5, e_6, e_7)$=(0100000)

C^*=r + e

= (0001110)+(0100000)

= (0101110)

where C^* is the actual transmitted code word

Minimum Distance of a Block Code

- **Hamming weight w(c) :** It is defined as the number of non-zero components of c.

  For ex: The hamming weight of c=(11000110) is 4
- **Hamming distance d(c, x):** It is defined as the number of places where they differ .
- The hamming distance between c=(11000110) and x=(00100100) is 4
- The hamming distance satisfies the triangle inequality $d(c, x)+d(x, y) \geq d(c, y)$
- The hamming distance between two n-tuple c and x is equal to the hamming weight of the sum of c and x

$$d(c, x) = w(c+ x)$$

- For ex: The hamming distance between c=(11000110) and x=(00100100) is 4 and the weight of c + x = (11100010) is 4.

Contd…

- **Minimum hamming distance d_{min}:** It is defined as the smallest distance between any pair of code vectors in the code.

  For a given block code C, d_{min} is defined as:

  $$d_{min}=\min\{\, d(c, x): \ c, x€C, \ c \neq x\}$$

- The Hamming distance between two code vectors in C is equal to the Hamming weight of a third code vector in C.

  $$d_{min} = \min\{w(c+x):c, x €C, c\neq x\}$$

  $$= \min\{w(y):y €C, y\neq 0\}$$

  $$= w_{min}$$

APPLICATIONS

- Communications:
  - Satellite and deep space communications.
  - Digital audio and video transmissions.

- Storage:
  - Computer memory (RAM).
  - Single error correcting and double error detecting code.

ADVANTAGES

- It is the easiest and simplest technique to detect and correct errors.
- Error probability is reduced.

DISADVANTAGES

- Transmission bandwidth requirement is more.
- Extra bits reduces bit rate of transmitter and also reduces its power.

Cyclic Codes

- ➢Definition: A code *C* is cyclic if

1) C is a linear code: c1+c2=m1G+m2G=(m1+m2)G → a new code word in the code.
2) Any cyclic shift of a codeword is also a codeword, i.e. If c_0 c_1 c_2 …. c_{n-2} c_{n-1} is a codeword, then c_{n-1} c_0 c_1 …. c_{n-3} c_{n-2} c_1 c_2 c_3 …. c_{n-1} c_0 are all codewords

- ➢Example

(i) Code *C* = {000, 101, 011, 110} is cyclic.

(ii) Given the following code :

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

- ➢Thus, it is not a cyclic code because, the cyclic shift of [10111] is [11011] .

Contd…

- The code with the generator matrix

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

has codewords

$c_1 = 1011100$ $\quad c_2 = 0101110$ $\quad c_3 = 0010111$

$c_1 + c_2 = 1110010$ $c_1 + c_3 = 1001011$ $c_2 + c_3 = 0111001$

$c_1 + c_2 + c_3 = 1100101$

and it is cyclic because the right shifts have the following impacts

$$c_1 \to c_2,\ c_2 \to c_3,\ c_3 \to c_1 + c_3$$

$$c_1 + c_2 \to c_2 + c_3,\ c_1 + c_3 \to c_1 + c_2 + c_3,\ c_2 + c_3 \to c_1$$

Contd…

- ➢A (n, k) cyclic code can be generated by a polynomial g(x) which has
  - ✓degree n-k and
  - ✓is a factor of x^n - 1.

Which is known as **generator polynomial**.

- ➢Given message bits, ($m_{k-1} \ldots m_1 m_0$), the code is generated simply as:

$$C(x) = \sum_{j=0}^{k-1} m_j x^j g(x),$$

- ➢In other words, *C(x)* can be considered as the product of *m(x)* and *g(x)*.
- ➢A (7,4) cyclic code with g(x) = x^3 + x + 1. If m(x) = x^3 + 1, C(x) = x^6 + x^4 + x + 1.

Generator Polynomials:

- It is that polynomial which is able to generate all the codeword polynomials.
- For (n, k) cyclic code

$$\mathrm{C} = (\mathrm{C}_{n-1}, \mathrm{C}_{n-2}, \dots, \mathrm{C}_2, \mathrm{C}_1, \mathrm{C}_0)$$

- In polynomial form

$$\mathrm{C}(x) = \mathrm{C}_{n-1}x^{n-1} + \mathrm{C}_{n-2}x^{n-2} + \cdots + \mathrm{C}_2x^2 + \mathrm{C}_1x + \mathrm{C}_0$$

- Theorem: Let C be an (n, k) cyclic code then: There exists only one polynomial g(x) of the minimum degree (n-k).

$$g(x) = g_{n-k}x^{n-k} + \mathrm{C}_{n-k-1}x^{n-k-1} + \cdots + g_1x + g_0)$$

Properties:

(i) $g_{n-k} = g_0$ **(Always)**

Contd…

How can we find different codewords by using generator polynomial

- A (7,4) cyclic code with $g(x) = x^3 + x + 1$.
- Generator polynomial is also a codeword, hence
- $C_1(x) = x^3 + x + 1$
- Other codeword can be obtained as
- $x\, g(x) = x^4 + x^2 + x = C_2(x)$
- $x^2\, g(x) = x^5 + x^3 + x^2 = C_3(x)$
- $x^{k-1}\, g(x) = x^5 + x^3 + x^2 = C_k(x)$

Encoding of cyclic codes

- C(x) = m(x)g(x) gives non-systematic code.

- Systematic Code
- The codeword can be expressed by the data polynomial *m*(*x*) and the check polynomial $c_p(x)$ as

$$c(x) = m(x)\ x^{n-k} + c_p(x)$$

where $c_p(x)$ = remainder from dividing $m(x)\ x^{n-k}$ by generator g(x)

Decoding

- Let $m(x)$ be the data block and $g(x)$ be the polynomial divisor, we have

$$x^{n-k}\, m(x)/g(x) = q(x) + c_p(x)\, /g(x)$$

- The transmitted block is $c(x) = x^{n-k}\, m(x) + c_p(x)$
- If there are no errors the division of $c(x)$ by $g(x)$ produces no remainder.

$$c(x)\ /\ g(x) = q(x)$$

- If one or more bit errors, then the received block $c'(x)$ will be of the form $c'(x) = c(x) + e(x)$ and the error pattern is detected from known error syndromes $s(x) = e(x)/g(x)$
- The syndrome value $s(x)$ only depends on the error bits

Cyclic Code: Example

- **Example :** Find the codeword $c(x)$ if $m(x) = 1 + x + x^2$ and $g(x) = 1 + x + x^3$, for (7, 4) cyclic code
- We have n = total number of bits = 7, k = number of information bits = 4, r = number of parity bits = $n - k = 3$

$\because$

$$c_p(x) = rem\left[\frac{m(x)x^{n-k}}{g(x)}\right]$$

$$= rem\left[\frac{x^5 + x^4 + x^3}{x^3 + x + 1}\right] = x$$

Then,

$$c(x) = m(x)x^{n-k} + c_p(x) = x + x^3 + x^4 + x^5$$

= 0111010

Contd…

- ➢Example : Let $m(x) = 1 + x + x^2$ and $g(x) = 1 + x + x^3$, for (7, 4) cyclic code
- ➢Assume $e(x)$ = 1000000. The received block $c'(x)$ = 1111010
- ➢We have $s(x) = e(x)/g(x) = x^2 + 1$. Therefore, s = 101. According to Table (b), we have the error pattern 1000000
- ➢Now, supposed the received block is 0111011, or

  $c'(x) = x^5 + x^4 + x^3 + x + 1$. Find $s(x)$ and the error pattern.

A Single-Error-Correcting (7,4) Cyclic Code

(a) Table of valid codewords

Data Block	Codeword
0000	0000000
0001	0001011
0010	0010110
0011	0011101
0100	0100111
0101	0101100
0110	0110001
0111	0111010
1000	1000101
1001	1001110
1010	1010011
1011	1011000
1100	1100010
1101	1101001
1110	1110100
1111	1111111

(b) Table of syndromes for single-bit errors

Error pattern E	Syndrome S
0000001	001
0000010	010
0000100	100
0001000	011
0010000	110
0100000	111
1000000	101

Bose_chaudhuri_hocquenghem codes (BCH)

- For positive pair of integers $m \geq 3$ and t, a (n, k) BCH code has parameters:
- Block length: $n = 2^m - 1$
- Number of check bits: $n - k \leq mt$
- Minimum distance: $d_{\min} \geq 2t + 1$
- t<(2^m – 1)/2 random errors detected and corrected.
- So also called ‘t-error correcting BCH code’.
- Major advantage is flexibility for block length and code rate.

Contd…

- Generatorpolynomial → specified in terms of its roots from Galois Field GF(2^k).
- g(x) has $\alpha, \alpha^2, \ldots, \alpha^{2t}$ and their conjugates as its roots.
- We choose g(x) from $x^n + 1$ polynomial factors by taking x^{n-k} as highest term.

Contd…

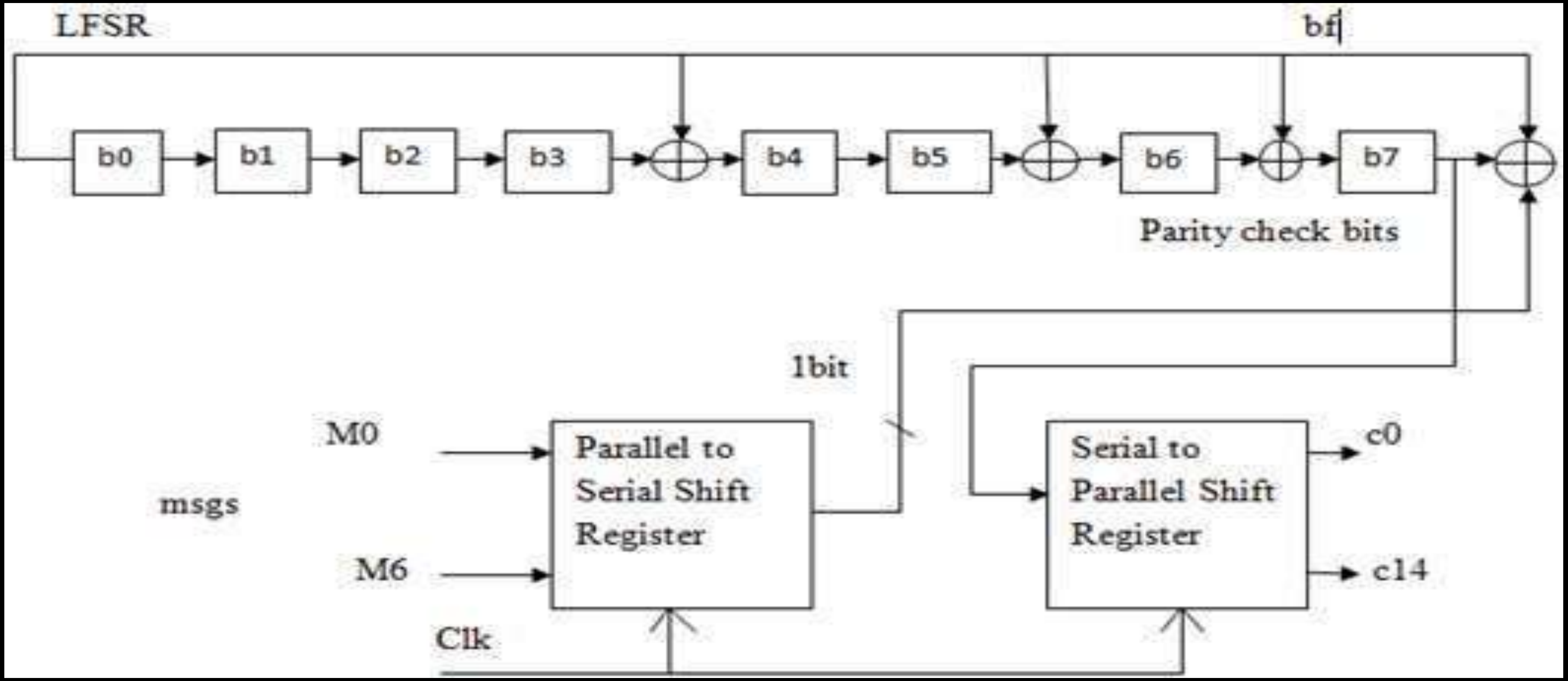
- The parameters of some useful BCH codes are:

n	k	t	Generator Polynomial
7	4	1	1 011
15	11	1	10 011
15	7	2	111 010 001
15	5	3	10 100 110 111
31	26	1	100 101
31	21	2	11 101 101 001
31	16	3	1 000 111 110 101 111
31	11	5	101 100 010 011 011 010 101
31	6	7	11 001 011 011 110 101 000 100 111

BCH Encoder

- (15, 7) BCH Encoder.
- The 7 message bits (M0, M1….M6) are applied to the parallel to serial shift register.
- The outputof parallel to serial shift register will be sent to (15, 7) BCH Encoder module.
- Using these message bits, parity bits are computed and sent to serial to parallel shift register.
- Then parity bits are appended to original message bits to obtain 15 bit encoded data.
- This entire encoding process requires 15 clock cycles.

BLOCK DIAGRAM

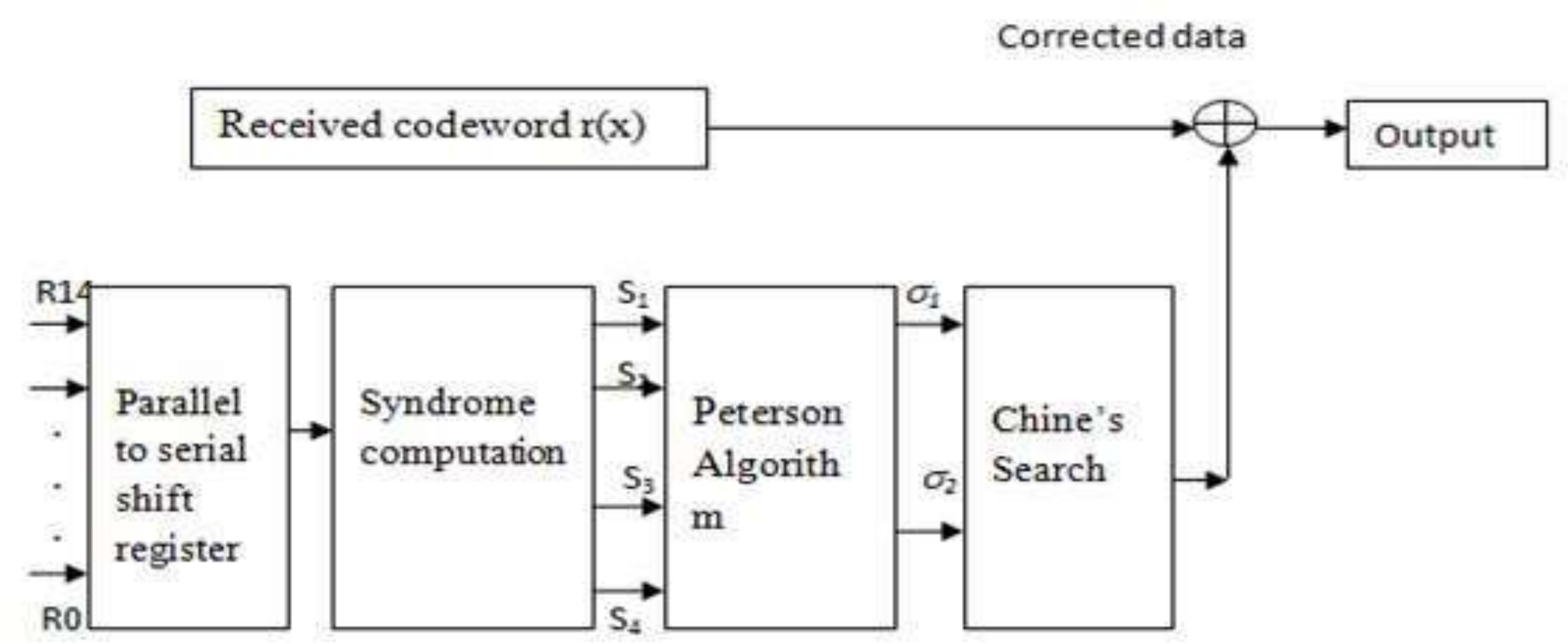

LFSR
bf
b0
b1
b2
b3
b4
b5
b6
b7
Parity check bits
1bit
M0
msgs
M6
Parallel to Serial Shift Register
Serial to Parallel Shift Register
c0
c14
Clk

Block diagram of (15,7) BCH Encoder

BCH Decoder

- (15, 7) BCH decoder.
- The decoding algorithm for BCH codes consists of three major steps.
- Calculate the syndromevalue Si, i=1,2,….,2t from the received word r(x).
- Determine the error location polynomial s*(x)*
- Find the roots of *s(x)* and then correct the errors

BLOCK DIAGRAM



Block diagram for (15, 7) BCH Decoder.

EXAMPLE

- For a (31,21,2) BCH code:
- Encoder: t = 2
- $C(x)=x^{n}+1=x^{31}+1$

  $=(x+1)(x^{10}+x^{9}+x^{8}+x^{6}+x^{5}+x^{3}+1)$

  $(x^{20}+x^{17}+x^{16}+x^{13}+x^{11}+x^{7}+x^{6}+x^{5}+x^{2}+x+1)$
- Here highest order term for g(x) must be chosen as
  - $x^{n-k}=x^{31-21}=x^{10}$
- So $g(x)= (x^{10}+x^{9}+x^{8}+x^{6}+x^{5}+x^{3}+1)$

Contd…

- Message D: (0110011)
- Data: $d(x)=x^5+x^4+x+1$
- So code $C(x)=d(x).g(x)$

$$= (x^5+x^4+x+1)(x^{10}+x^9+x^8+x^6+x^5+x^3+1)$$

$$= x^{15}+2x^{14}+2x^{13}+x^{12}+2x^{11}+4x^{10}+3x^9+2x^8$$

$$+2x^7+2x^6+2x^5+2x^4+x^3+x+1$$

$$= x^{15}+x^{12}+ x^9+ x^3+x+1$$

- Codeword, C: (1001001000001011)

MERITS

- The principal advantage is the ease with which they can be decoded using ‘syndrome decoding’ method.
- Allows very simple electronic hardware to perform the task, obviating the need for a computer, and meaning that a decoding device may be made small and low-powered.
- Low amount of redundancy
- Easy to implement in hardware
- Widely used

DEMERITS

- Complexity
- Iterative and complex decoding algorithm
- Decoder cannot decide whether a decoded package is false or not.

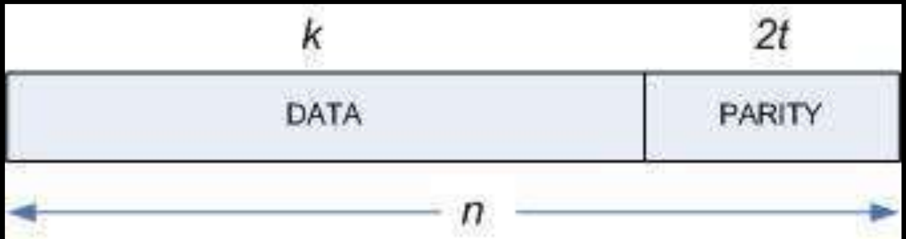
Reed Solomon Codes

- It is a subclass of non-binary BCH codes.
- The encoder for an RS code differs from binary encoder in that it operates on multiple bits rather than individual bits

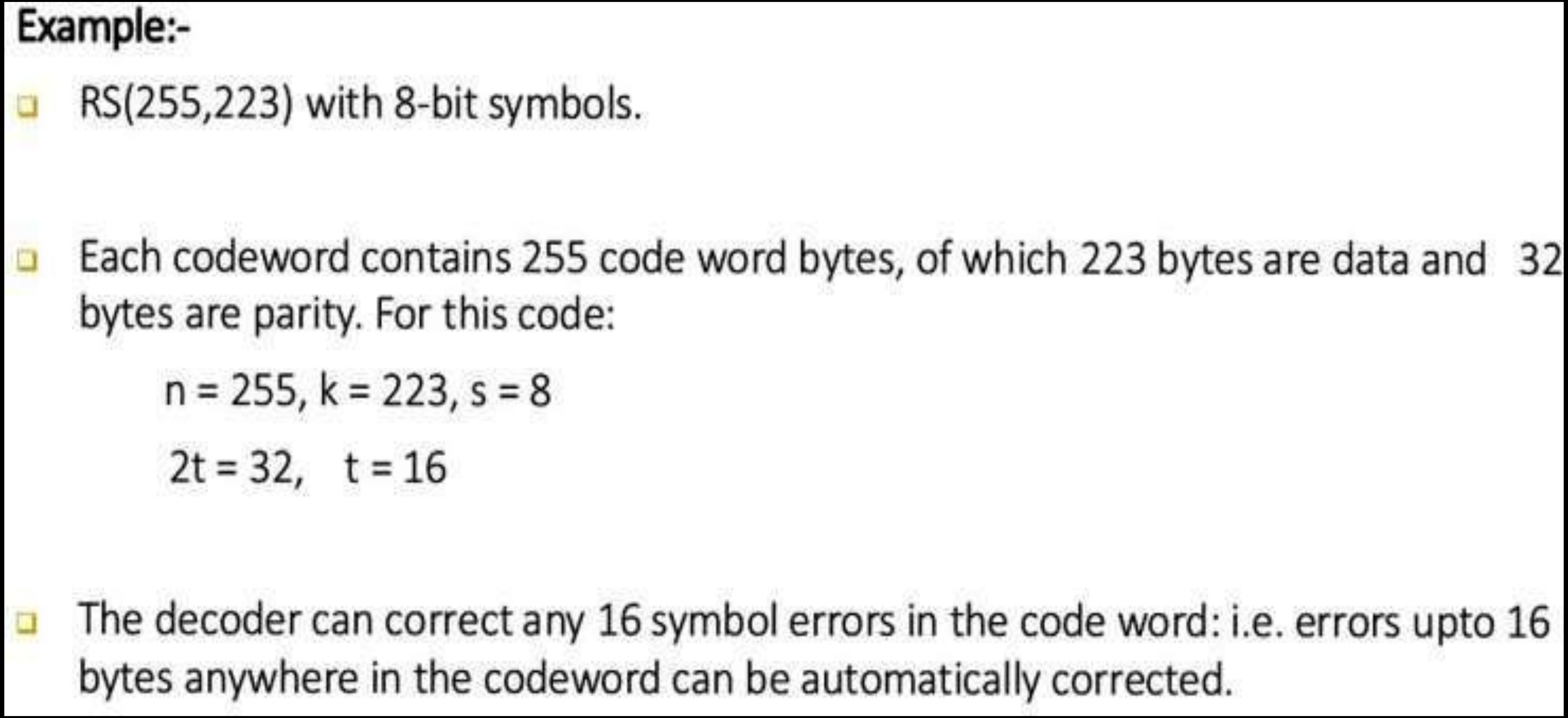
- Properties:
- Block length, n=2^m-1
- Message size k symbols
- Parity check size (n-k)=2t symbols
- Minimum distance, d_min=(2t+1)
- (n, k) code is used to encode m-bit symbol, Redundant bits=(n-k)

Contd…

- Before data transmission, the encoder attaches parity symbols to the data using a predetermined algorithm before transmission.
- At the receiving side, the decoder detects and corrects a limited predetermined number of errors occurred during transmission.
- Transmitting additional symbols introduced by FEC is better than retransmitting the whole package when at least an error has been detected by the receiver.
- A Reed-Solomon code is a block code and can be specified as RS(n,k)



Contd…



Example:-

- RS(255,223) with 8-bit symbols.

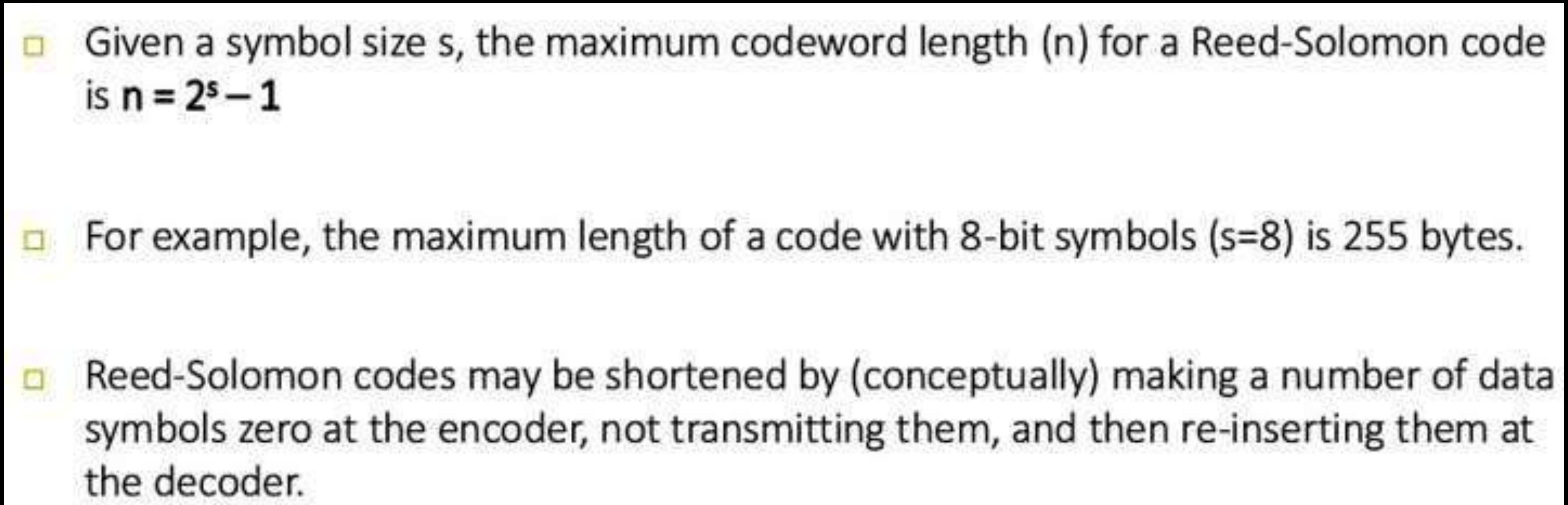
- Each codeword contains 255 code word bytes, of which 223 bytes are data and 32 bytes are parity. For this code:

  $n = 255, k = 223, s = 8$

  $2t = 32, \quad t = 16$

- The decoder can correct any 16 symbol errors in the code word: i.e. errors upto 16 bytes anywhere in the codeword can be automatically corrected.

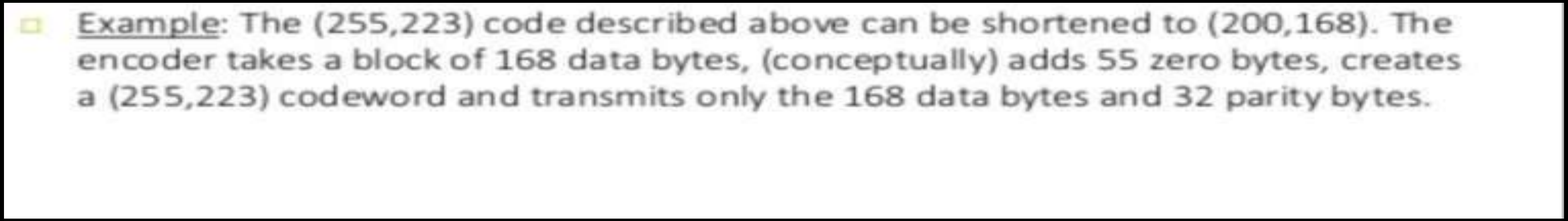
Contd…



- □ Given a symbol size s, the maximum codeword length (n) for a Reed-Solomon code is $\mathbf{n = 2^s - 1}$

- □ For example, the maximum length of a code with 8-bit symbols (s=8) is 255 bytes.

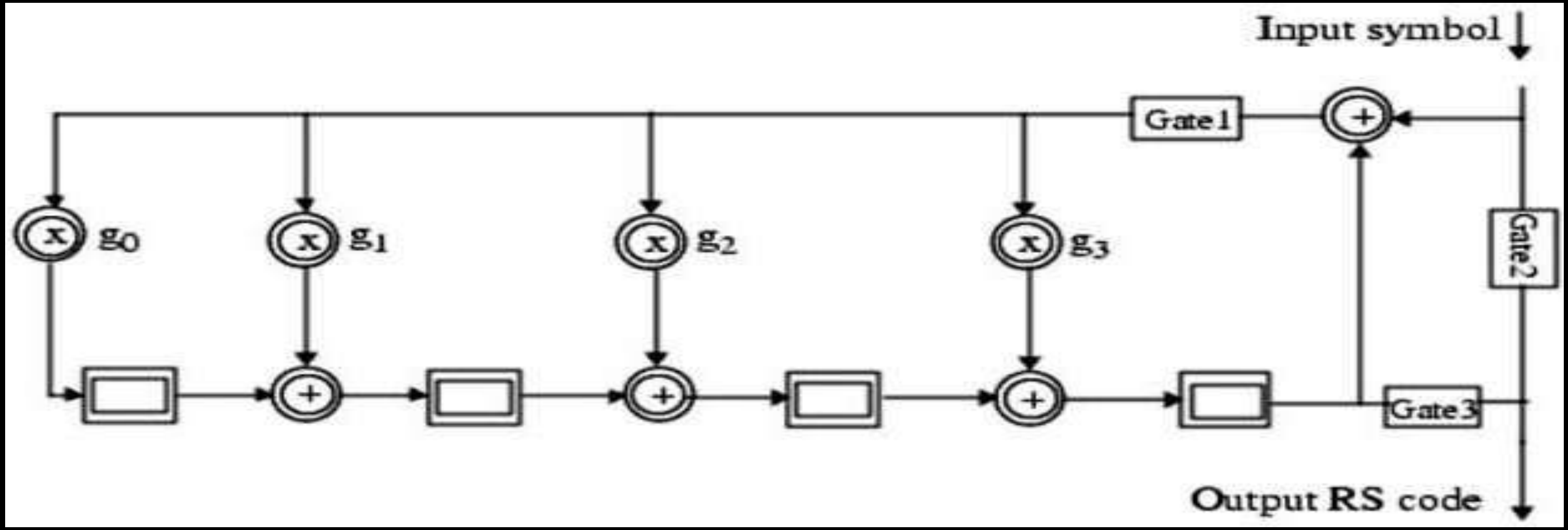
- □ Reed-Solomon codes may be shortened by (conceptually) making a number of data symbols zero at the encoder, not transmitting them, and then re-inserting them at the decoder.

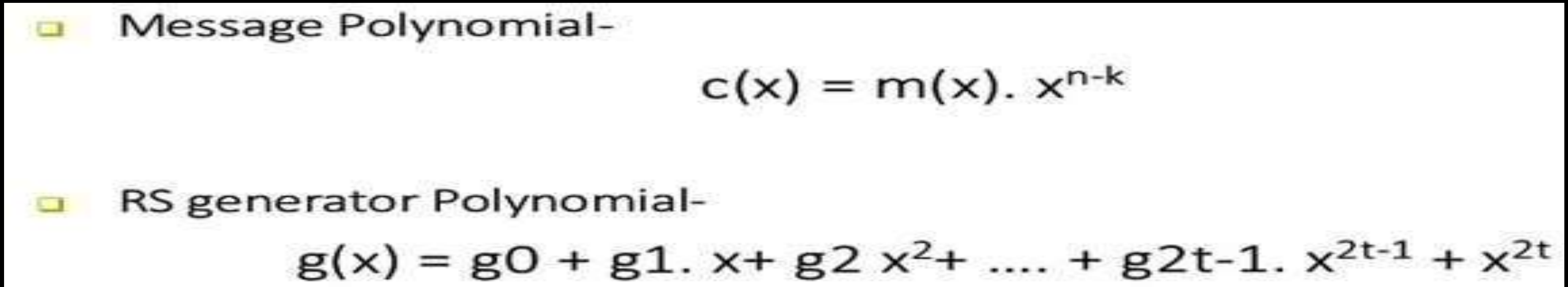


- □ <u>Example</u>: The (255,223) code described above can be shortened to (200,168). The encoder takes a block of 168 data bytes, (conceptually) adds 55 zero bytes, creates a (255,223) codeword and transmits only the 168 data bytes and 32 parity bytes.

Contd…

Encoder:





- Message Polynomial-

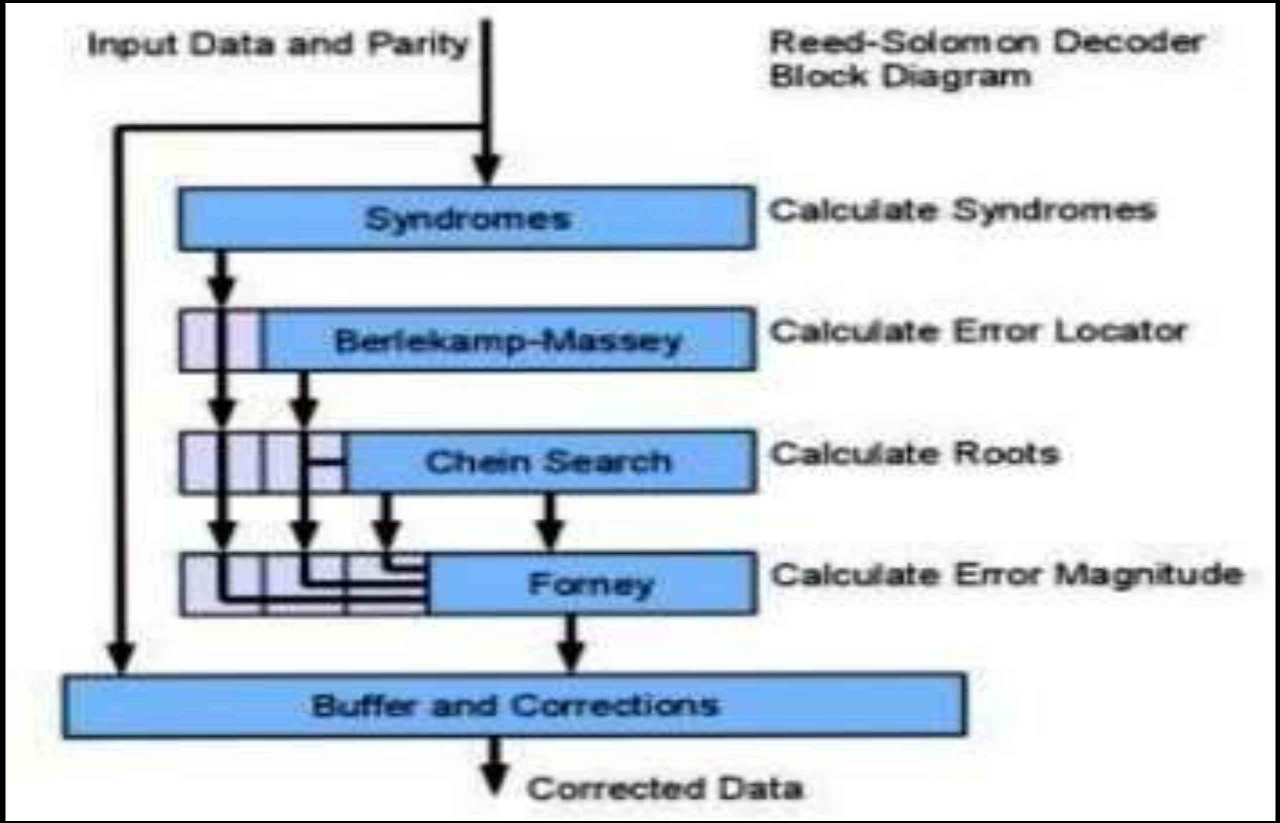
$$c(x) = m(x).\ x^{n-k}$$

- RS generator Polynomial-

$$g(x) = g0 + g1.\ x + g2\ x^2 + \ \ldots. + g2t\text{-}1.\ x^{2t-1} + x^{2t}$$

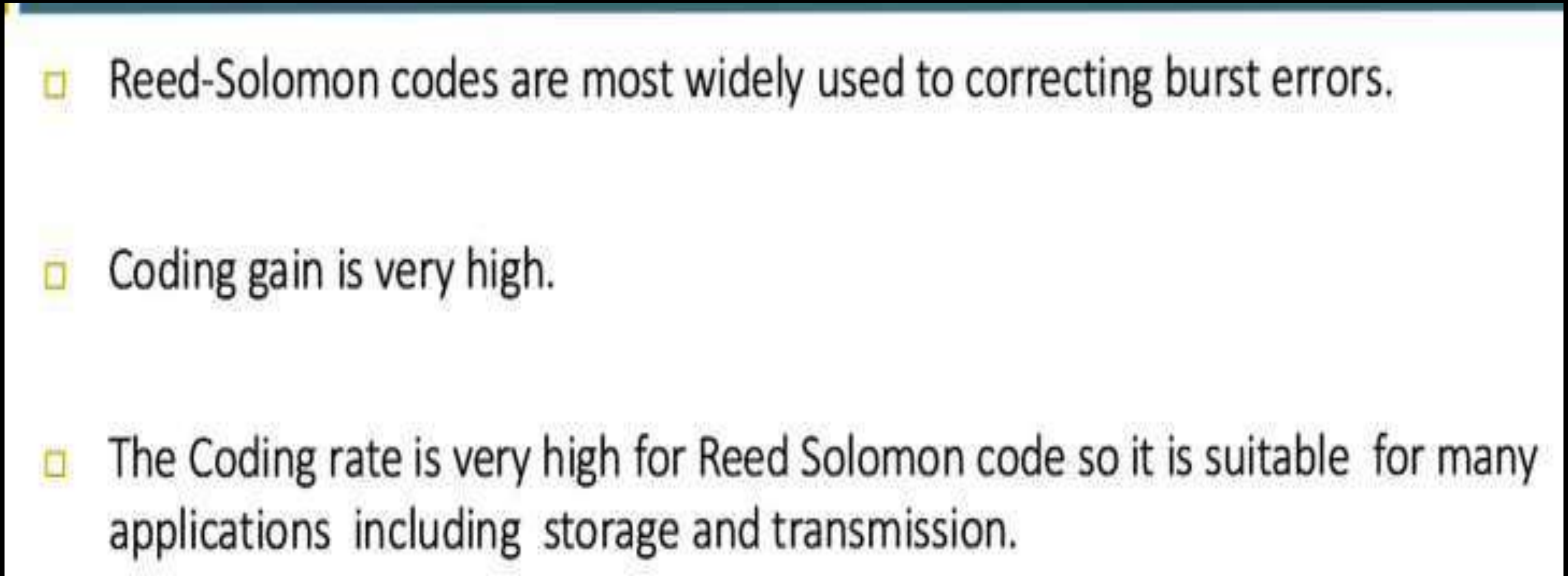
Contd…

Decoder:



Contd…

Advantages:



- Reed-Solomon codes are most widely used to correcting burst errors.
- Coding gain is very high.
- The Coding rate is very high for Reed Solomon code so it is suitable for many applications including storage and transmission.

Convolutional Codes

- Convolutional codes are introduced by Elias in 1955.
- Convolution coding is a popular error-correcting coding method used to improve the reliability of communication system.
- A message is convoluted, and then transmitted into a noisy channel.
- This convolution operation encodes some redundant information into the transmitted signal, thereby improving the data capacity of the channel.
- Convolution codes are error detecting codes used to reliably transmit digital data over unreliable communication channel system to channel noise.

Contd…

- The convolutional codes map information to code bits, but sequentially convolve the sequence of information bit according to some rule.
- The convolutional coding can be applied to a continuous data stream as well as to blocks of data whereas the block codes can be applied only for the block of data.

Convolutional encoder

- Convolutional encoder is a finite state machine (FSM), processing information bits in a serial manner.
- Convolutional encoding of data is accomplished using a shift register and associated combinatorial logic that performs modulo-two addition.
- A shift register is merely a chain of flip-flops wherein the output of the nth flip-flop is tied to the input of the (n+1)th flip flop.
- Every time the active edge of the clock occurs, the input to the flip-flop is clocked through to the output, and thus the data are shifted over one stage.

Contd…

- In convolutional code the block of n code bits generated by the encoder in a particular time instant depends not only on the block of k message bits within that time instant but also on the block of data bits within a previous span of $N-1$ time instants ($N>1$).
- A convolutional code with constraint length N consists of an N-stage shift register (SR) and ν modulo-2 adders.

Contd…

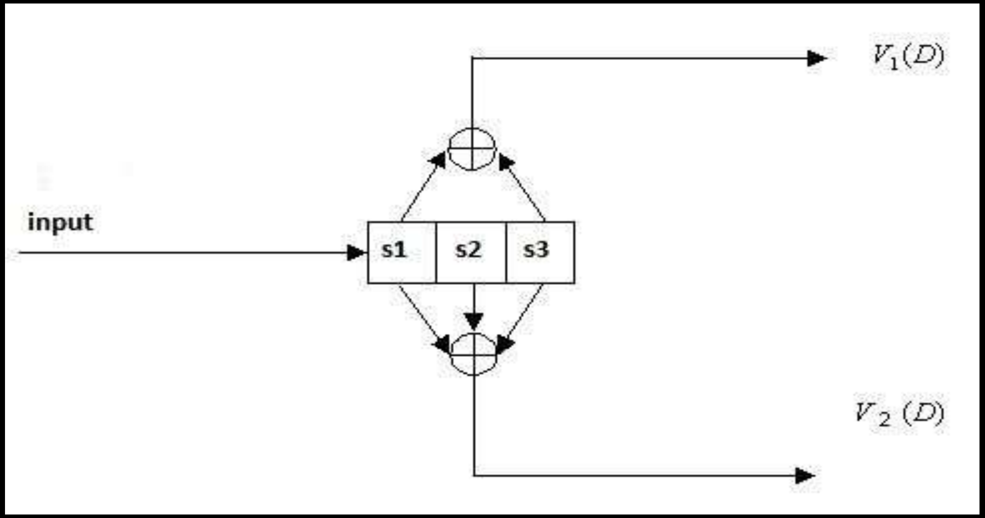


Fig (a) Convolutional Encoder with N=3 and υ=2

- The message bits are applied at the input of the shift register (SR). The coded digit stream is obtained at the commutator output. The commutator samples the ν modulo-2 adders in a sequence, once during each input-bit interval.

Contd…

- ❖ Example: Assume that the input digits are 1010. Find the coded sequence output for previous Fig (a).
- ➢ Initially, the Shift Registers s1=s2=s3=0.
- ➢ When the first message bit 1 enters the SR, s1= 1, s2 = s3=0.Then ν1=1, ν2=1 and the coder output is 11.
- ➢ When the second message bit 0 enters the SR, s1=0, s2=1, s3=0.Then ν1=1 and ν2=0 and the coder output is 10.
- ➢ When the third message bit 1 enters the SR, s1=1, s2=0 and s3=1Then ν1=0 and ν2=0 and the coder output is 00.
- ➢ When the fourth message bit 0 enters the SR, s1=0, s2=1 and s3=0.Then ν1=1 and ν2=0 and the coder output is 10.
- ➢ The coded Output Sequence is : 11100010

Contd…

❖ PARAMETERS OF A CONVOLUTIONAL ENCODER

- Convolutional codes are commonly specified by three parameters: (n,k,m):

n = number of output bits

k = number of input bits

m = number of memory registers

- **Code Rate:** The quantity k/n is called as code rate. It is a measure of the efficiency of the code.
- Constraint Length: The quantity L(or K) is called the constraint length of the code. It represents the number of bits in the encoder memory that affect the generation of the n output bits. It is defined by

Constraint Length, L = k (m-1)

Encoder Representation

The encoder can be represented in several different but equivalent ways. They are:

a) Generator Representation

b) State Diagram Representation

c) Tree Diagram Representation

d) Trellis Diagram Representation

Contd…

a) Generator Representation

- Generator representation shows the hardware connection of the shift register taps to the modulo-2 adders. A generator vector represents the position of the taps for an output. A "1" represents a connection and a "0" represents no connection.
- (*n*, *k*, *L*) Convolutional code can be described by the generator sequences that are the impulse response for each coder n output branches.
- Generator sequences specify convolutional code completely by the associated generator matrix.
- Encoded convolution code is produced by matrix multiplication of the input and the generator matrix.

Contd…

- For example, the two generator vectors for the encoder in Fig (a) are **g1** = [111] and **g2** = [101], where the subscripts 1 and 2 denote the corresponding output terminals.

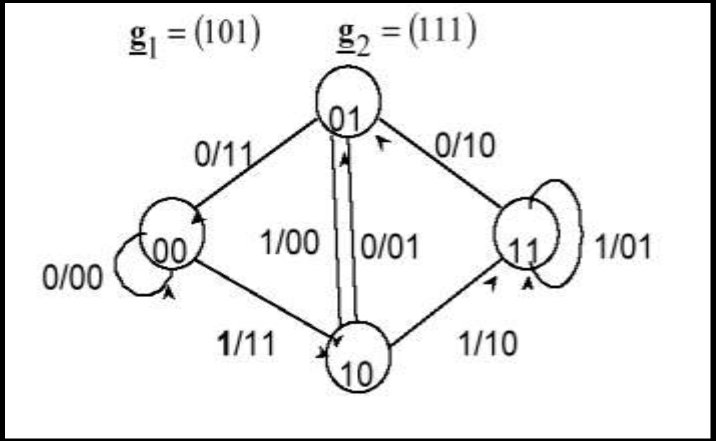
b) State Diagram Representation

- In the state diagram, the state information of the encoder is shown in the circles. Each new input information bit causes a transition from one state to another.
- Contents of the rightmost (K-1) shift register stages define the states of the encoder. The transition of an encoder from one state to another, as caused by input bits, is depicted in the state diagram.

Contd…

- The path information between the states, denoted as x/***c***, represents input information bit x and output encoded bits ***c***.
- It is customary to begin convolutional encoding from the all zero state.

Example: State diagram representation of convolutional codes.



Fig(b):state diagram Here k=1,n=2,K=3.

Contd…

- From the state diagram

Let 00 State a ; 01 State b; 10 State c; 11 State d;

(1) State a goes to State a when the input is 0 and the output is 00

(2) State a goes to State b when the input is 1 and the output is 11

(3) State b goes to State c when the input is 0 and the output is 10

(4) State b goes to State d when the input is 1 and the output is 01

(5) State c goes to State a when the input is 0 and the output is 11

(6) State c goes to State b when the input is 1 and the output is 00

(7) State d goes to State c when the input is 0 and the output is 01

(8) State d goes to State d when the input is 1 and the output is 10

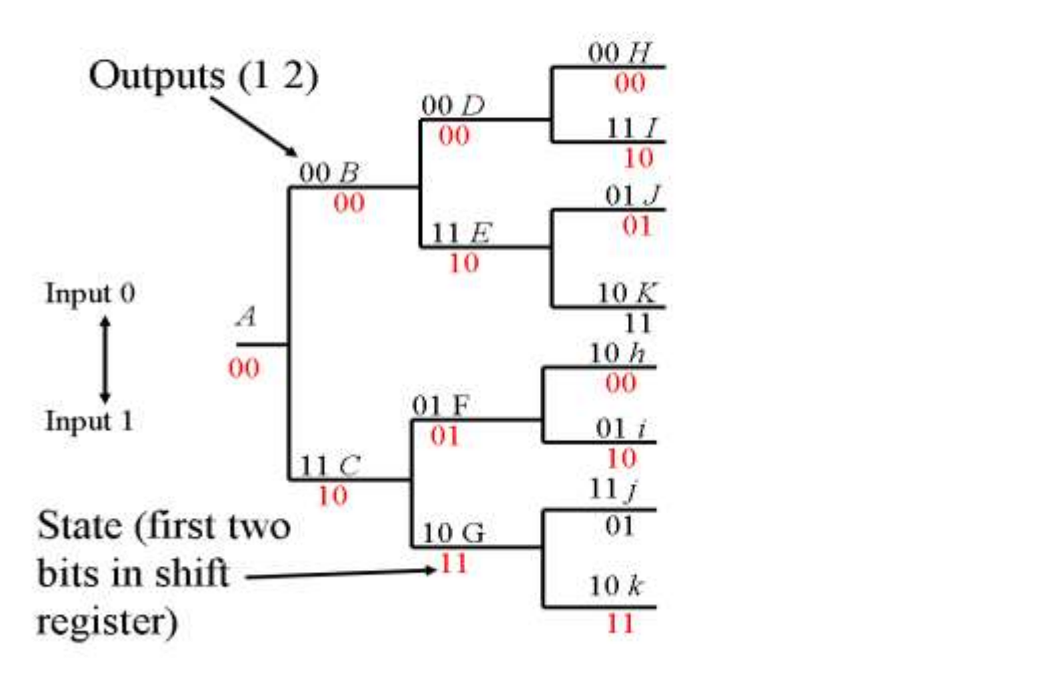
Contd…

c) Tree Diagram Representation

- The tree diagram representation shows all possible information and encoded sequences for the convolutional encoder.
- In the tree diagram, a solid line represents input information bit 0 and a dashed line represents input information bit 1.
- The corresponding output encoded bits are shown on the branches of the tree.
- An input information sequence defines a specific path through the tree diagram from left to right.

Contd…

Example: Tree Diagram representation of convolutional codes



Fig(c): Tree diagram

Contd…

- The tree diagram in **Fig(b)** tends to suggest that there are eight states in the last layer of the tree and that this will continue to grow. However some states in the last layer (i.e. the stored data in the encoder) are equivalent as indicated by the same letter on the tree (for example H and h).
- These pairs of states may be assumed to be equivalent because they have the same internal state for the first two stages of the shift register and therefore will behave exactly the same way to the receipt of a new (0 or 1) input data bit.

Contd…

d) Trellis Diagram Representation

- The trellis diagram is basically a redrawing of the state diagram. It shows all possible state transitions at each time step.
- The trellis diagram is drawn by lining up all the possible states (2L) in the vertical axis. Then we connect each state to the next state by the allowable codeword's for that state.
- There are only two choices possible at each state. These are determined by the arrival of either a 0 or a 1 bit.
- The arrows show the input bit and the output bits are shown in parentheses.
- The arrows going upwards represent a 0 bit and going downwards represent a 1 bit.

Contd…

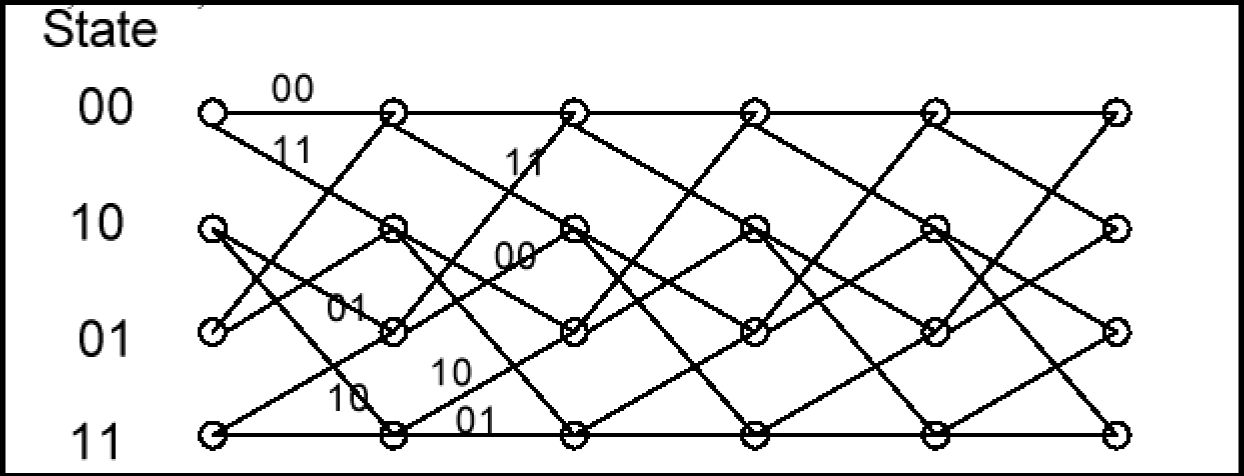
Steps to construct trellis diagram

- It starts from scratch (all 0’s in the SR, i.e., state a) and makes transitions corresponding to each input data digit.
- These transitions are denoted by a solid line for the next data digit 0 and by a dashed line for the next data digit 1.
- Thus when the first input digit is 0, the encoder output is 00 (solid line)
- When the input digit is 1, the encoder output is 11 (dashed line).
- We continue this way for the second input digit and so on as depicted in Fig (e) that follows.

Contd…

Example: Encoding of convolutional codes using Trellis Representation

k=1, n=2, K=3 convolutional code



We begin in state 00:

Fig (d)

Input Data: 0 1 0 1 1 0 0

Output: 0 0 1 1 0 1 0 0 10 10 1 1

Decoding of Convolutional Codes

- There are several different approaches to decoding of convolutional codes.
- These are grouped in two basic categories:

  A. Sequential Decoding

  -Fano Algorithm.

  B. Maximum Likelihood Decoding

  -Viterbi Algorithm.

- Both of these two methods represent two different approaches .

Contd…

- Each node examined represents a path through part of the tree.
- TheFano algorithm can only operate over a code tree because it cannot examine path merging.
- At each decoding stage, the Fano-algorithm retains the information regarding three paths:

  -the current path,

  -its immediate predecessor path,

  -one of its successor paths.
- Based on this information, the Fano algorithm can move from the current path to either its immediate predecessor path or the selected successor path.

Contd…

- It allows both the forward and backward movement through the Trellis diagram flow.
- **Example: Decoding using Sequential decoding-Fano algorithm**

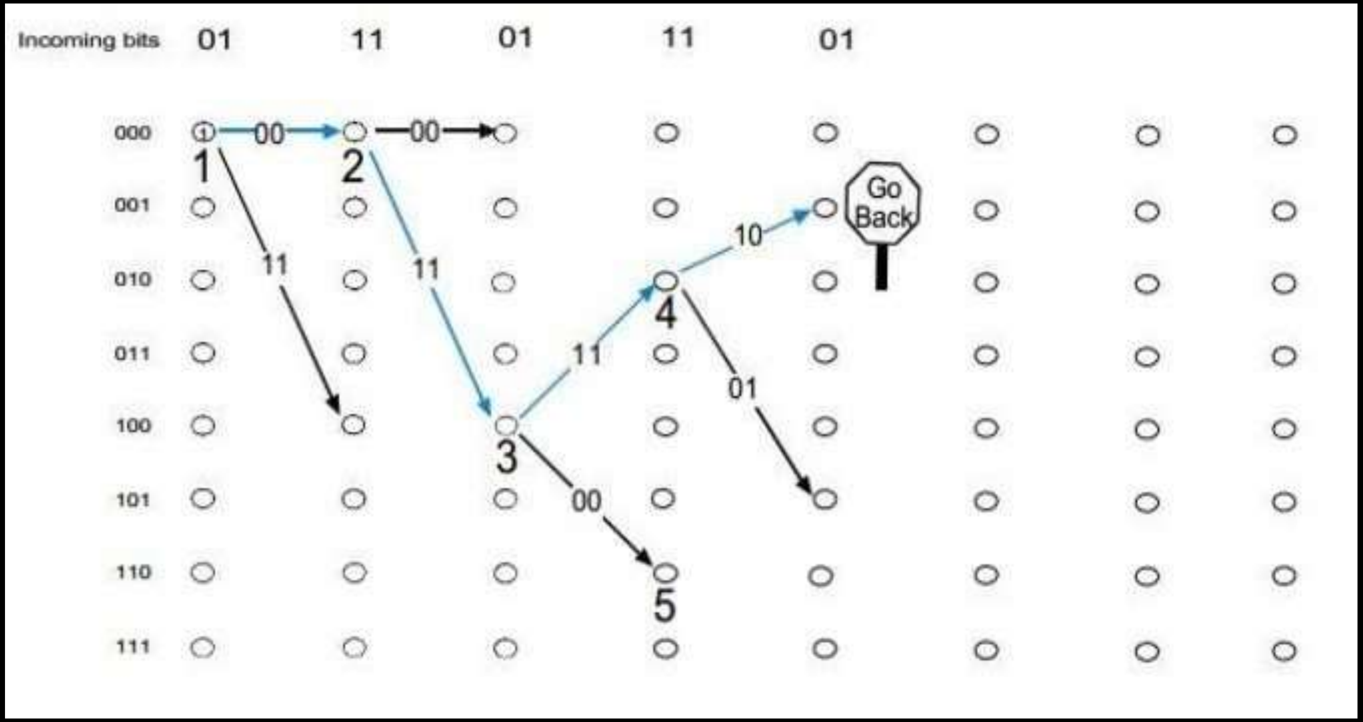
- Consider the code 01 11 01 11 01 01 11 is received, the algorithm will take a start and tally with the outputs it finds on the way. If an output does not tally, it retraces the position back to the previous ambiguous decision.

Contd…

A. Sequential Decoding – Fano Algorithm.

- It was one of the first methods proposed for decoding of a convolution ally coded bit stream.
- It was first proposed by Wozencraft and later a better version was proposed by Fano.
- Sequential decoding concentrates only on a certain number of likely codeword's.
- The purpose of sequential decoding is to search through the nodes of the code tree in an efficient way to find the maximum likelihood path.

Contd…



Fig(e):decoding using sequential decoding-Fano Algorihm

Contd…

B. Maximum Likelihood Decoding-Virtebi Algorithm

- The Viterbi decoder examines the entire received sequence of a given length.
- It works on maximum likelihood decoding rule which tried to reduce the error between the detected sequence and the original transmitted sequence.
- Trellis diagram is constructed for a system based on the received sequence the path is straight and the trellis level by level.
- If a condition raises in such a way that there is no path for the corresponding sequence then the viterbi decoding helps to detect the best path based on the subsequent sequence.

Contd…

- The best path is termed as survivor

Example: Maximum Likelihood decoding –Viterbi algorithm.

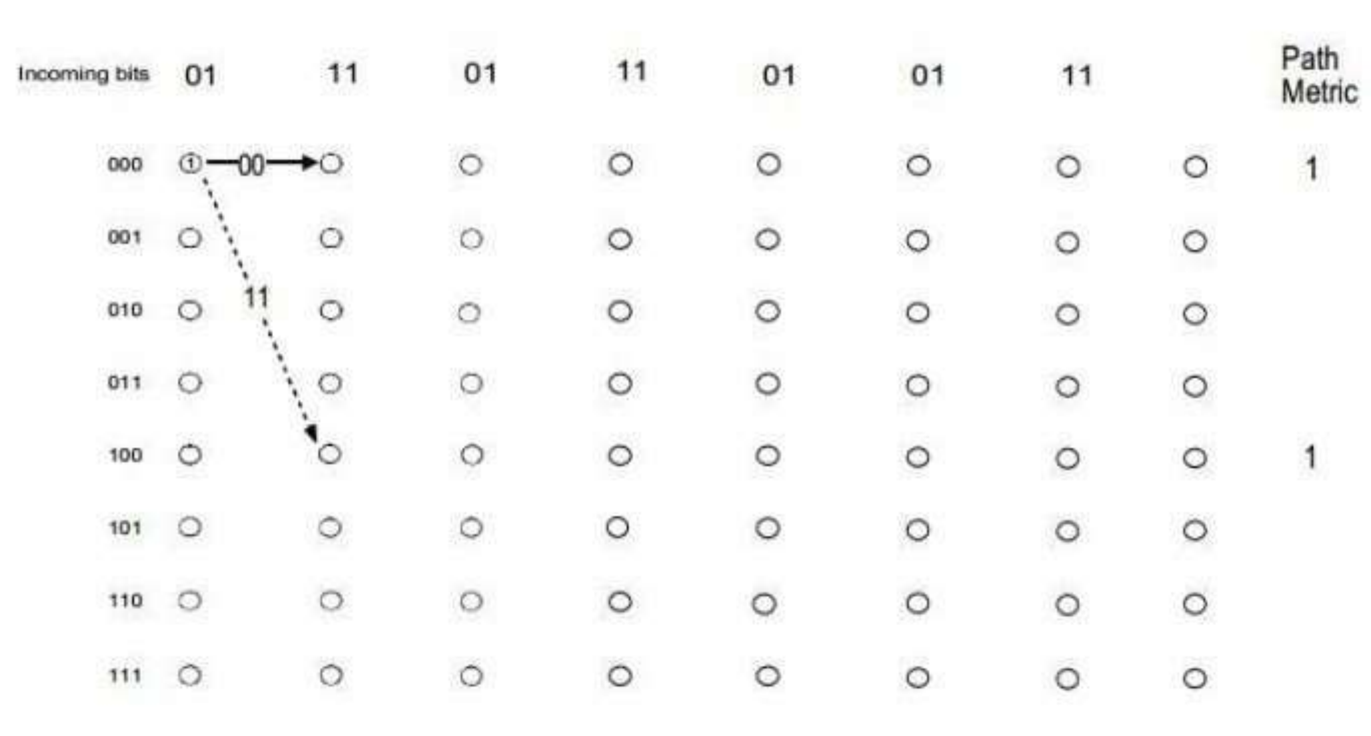


Fig: Maximum Likelihood decoding –Viterbi algorithm (Step1)

Contd…

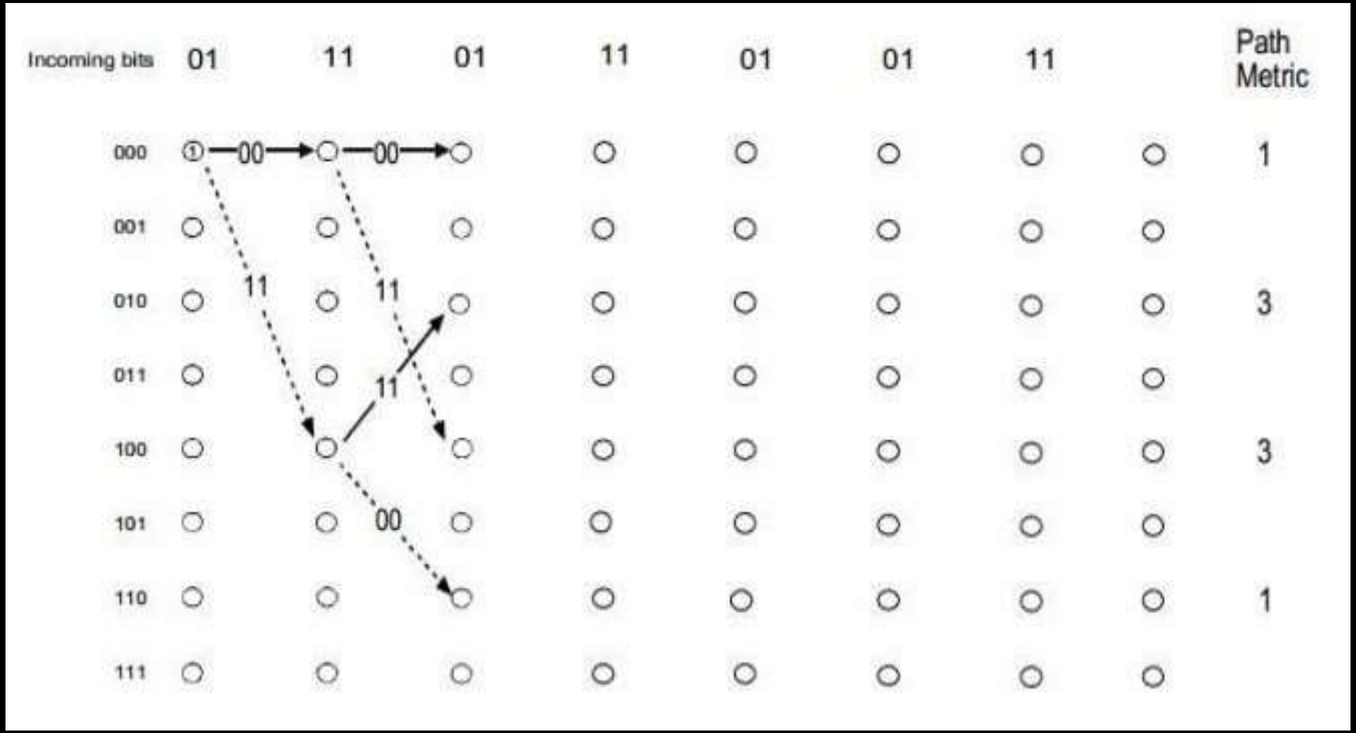


Fig: Maximum Likelihood decoding –Viterbi algorithm (Step2)

Contd…

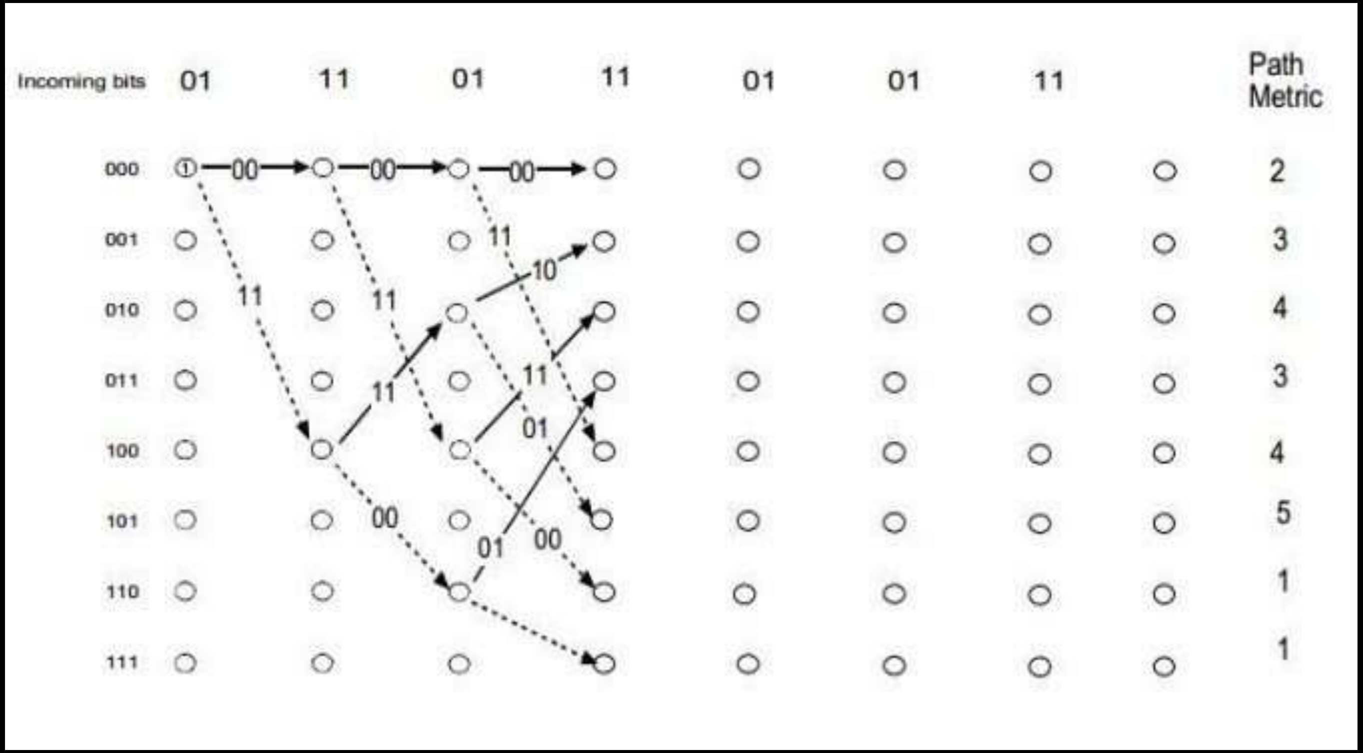


Fig: Maximum Likelihood decoding –Viterbi algorithm (Step3)

Contd…

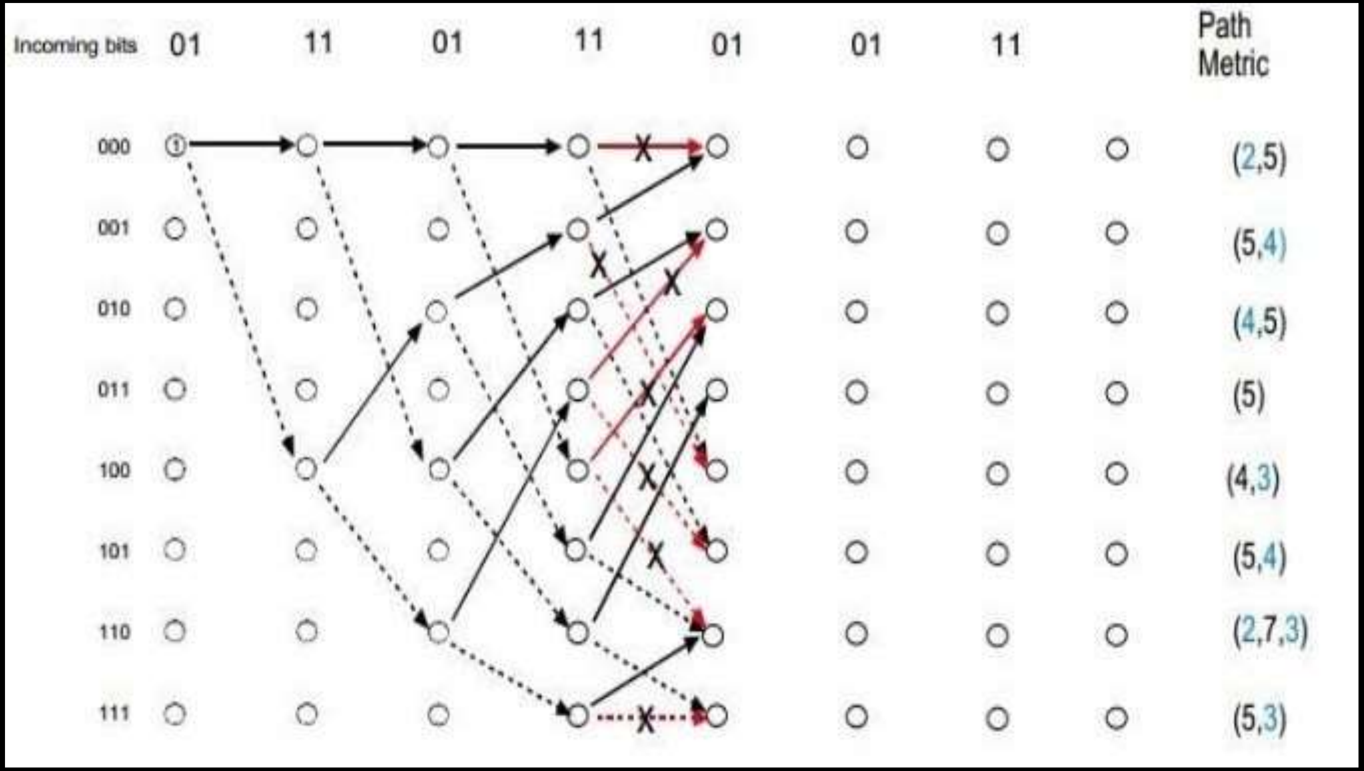


Fig: Maximum Likelihood decoding –Viterbi algorithm (Step4)

Advantages of Convolutional Codes

- Convolution coding is a popular error-correcting coding method used in digital communications.
- The convolution operation encodes some redundant information into the transmitted signal, thereby improving the data capacity of the channel.
- Convolution Encoding with Viterbi decoding is a powerful FEC technique that is particularly suited to a channel in which the transmitted signal is corrupted mainly by AWGN.
- It is simple and has good performance with low implementation cost.

Thank you