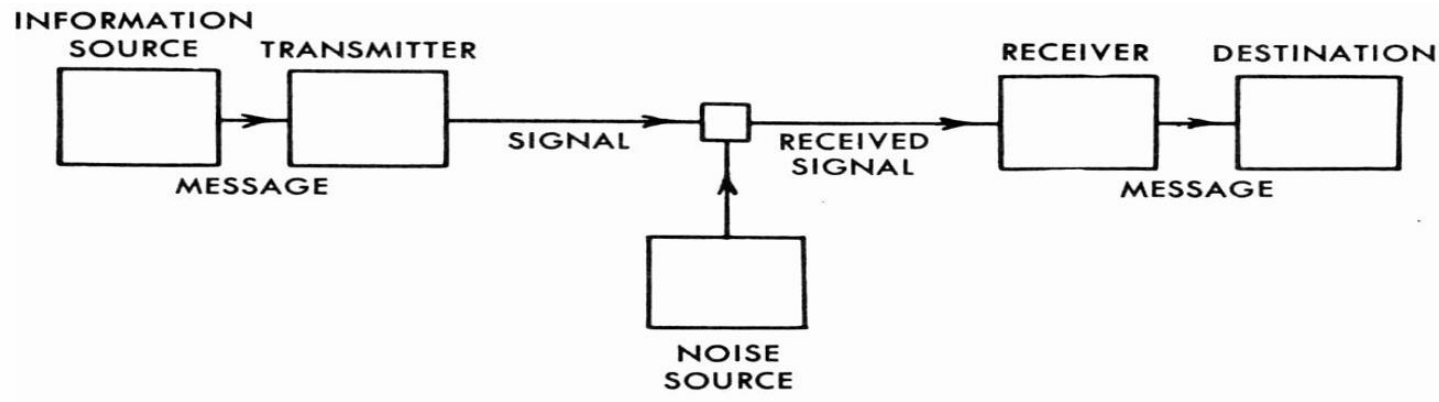
Module 3

Information theory

Information theory deals with mathematical modelling and analysis of a communication system.



INFORMATION:-

Message is the physical manifestation of information as produced by the source . The source generate messages in the form of symbols ,words ,images , sounds ets ..These messages convey information

Information is organized data which possesses some meaningful application for the receiver

The definition of information is based on the idea of lack of information ie, the less information one has ,greater is the information has to be gained

• If an event of low probability occurs it causes greater surprise and hence conveys more information

• More unexpected the event,Gives more information

• Probability of event is related to unexpectedness and it is related to information content

• The information content of a message can be expressed quantitatively as follows:

• The above concepts can now be formed in terms of probabilities as follows

• Say that, an information source emits one of q po ssible messages m1, m2 …… m q

• with p1, p2 …… pq as their probs. of occurrence. Based on the above intusion, the information content of the k th message, can be written as

..

$$I(m_k) \;\alpha\; \frac{1}{p_k}$$

Also to satisfy the intuitive concept, of information.

$I(m_k)$ must zero as p_k 1

Therefore,

$$\left.\boxed{\begin{array}{lll} I(m_k) > I(m_j); & \text{if} & p_k < p_j \\ I(m_k) \quad O(m_j); & \text{if} & p_k \quad 1 \\ I(m_k) \geq O; & \text{when} & O < p_k < 1 \end{array}}\right\} I$$

- Information content(I) can be measured using probability(P) of the occurrence of the event
- Least probable event contains more information
- I is directly proportional to (1/p)

ie, I=log(1/p)

Unit of information:-

A source puts out one of five possible messages during each message interval. The probs. of these messages are $p_1 = \frac{1}{2}$; $p_2 = \frac{1}{4}$; $p_1 = \frac{1}{4}$: $p_1 = \frac{1}{16}$, p_5 $\frac{1}{16}$

What is the information content of these messages?

$I(m_1) = -\log_2 \frac{1}{2} = 1$ bit

$I(m_2) = -\log_2 \frac{1}{4} = 2$ bits

$I(m_3) = -\log \frac{1}{8} = 3$ bits

$I(m_4) = -\log_2 \frac{1}{16} = 4$ bits

$I(m_5) = -\log_2 \frac{1}{16} = 4$ bits

Entropy(H)

The average amount of information per source symbol in a particular interval is called entropy.

- Let us assume an information source which emit K distinct symbols $m_1, m_2, ..., m_k, ..., m_K$ with probability $p_1, p_2, ..., p_k, ..., p_K$ respectively
- Assume source have delivered a statistically independent sequence of N symbols ($N \Rightarrow \infty$) then symbol m_k occurs Np_k times and occurrence of each m_k conveys information $I_k = -\log p_k$
- Therefore the total information due to m_k symbol in a sequence $= -Np_k \log p_k$
- The total information due to all symbols in a sequence

$$I_{total} = -N\sum p_k \log p_k$$

- Entropy H of the source is define as the average information per symbol

$$\boxed{H = I_{total}/N = -\sum_{k=1}^{K} p_k \log p_k}$$

• Entropy will attain its maximum value, when the symbol probabilities are equal.

1. Consider a discrete memoryless source with a source alphabet A = { s_0, s_1, s_2} with respective probs. $p_0 = ¼$, $p_1 = ¼$, $p_2 = ½$. Find the entropy of the source.

Solution: By definition, the entropy of a source is given by

$$H = \sum_{i=1}^{M} p_i \log \frac{1}{p_i} \text{ bits/ symbol}$$

H for this example is

$$H(A) = \sum_{i=0}^{2} p_i \log \frac{1}{p_i}$$

Substituting the values given, we get

$$H(A) = p_o \log \frac{1}{P_o} + P_1 \log \frac{1}{p_1} + p_2 \log \frac{1}{p_2}$$

$$= ¼ \log_2 4 + ¼ \log_2 4 + ½ \log_2 2$$

$$= \frac{3}{2} = 1.5 \text{ bits}$$

if $r_s = 1$ per sec, then

$$\mathbf{H'(A) = r_s\ H(A) = 1.5\ bits/sec}$$

CHANNEL CAPACITY(C):-

• Maximum amount of information that can be transmitted through a channel . Usually measured in bits per second.

SHANNON HARTLEY THEOREM:-

• Shannon–Hartley theorem tells the maximum rate at which information can be transmitted over a communications channel of a specified bandwidth in the presence of noise

• C=Blog(1+S/N)

• C=Channel capacity

• B=Bandwidth

• S/N=Signal to noise ratio

• Unit=bits/second

PROBLEM:-

A system has a bandwidth of 4 kHz and a SNR of 28dB at the input to the receiver. Calculate the information carrying capacity of the system.

$C=B\log_2(1+S/N)$

B=4Khz

$20\log_{10}(S/N)=28$db(Decibel)

$\log_{10}(S/N)=28/20=1.4$

$S/N=10^{1.4}=25.11$

$C=4\log_2(26.11)=4\log_{10}(26.11)\times\log_2(10)$

$=4\times1.4168\times3.32$

$=18.824$

CODING:-

• Refers to the process of converting digital data (such as text, images, audio, or video) into a specific format that can be easily transmitted, received, and decoded accurately

NEED FOR CODING

1. Error Detection and Correction
2. Efficiency in Bandwidth Usage
3. Robustness to Channel Distortions:
4. Security and Privacy
5. Compatibility and Interoperability
6. Capacity Enhancement
7. Adaptability to Channel Conditions

SHANNON FANO ALGORITHM:-

- A data compression technique
- Assigning variable-length codes to symbols in the input data based on their probabilities of occurrence
- Symbols that are more probable are assigned shorter codes, while less probable symbols are assigned longer codes

WORKING:-

- Calculate the probabilities of occurrence for each symbol in the input data.
- Sort the symbols in descending order of their probabilities.
- Divide the symbols into two subsets such that the sum of probabilities in each subset is approximately equal.
- Assign the code '0' to symbols in the first subset and '1' to symbols in the second subset.
- Repeat steps 3 and 4 for each subset until all symbols have been assigned codes.
- Encode the input data by replacing each symbol with its corresponding code.

EXAMPLE:-

X	P(X)	STEPS	CODE
E	0.4	0	0
A	0.3	1 0	10
D	0.15	1 1 0	110
B	0.1	1 1 1 0	1110
F	0.03	1 1 1 1 0	11110
C	0.02	1 1 1 1 1	11111

- Apply Shannon fano coding for the following message a1,a2,a3,a4,a5,a6 with there probabilities 0.36,.18,.18,0.12,0.09,0.07 respectively

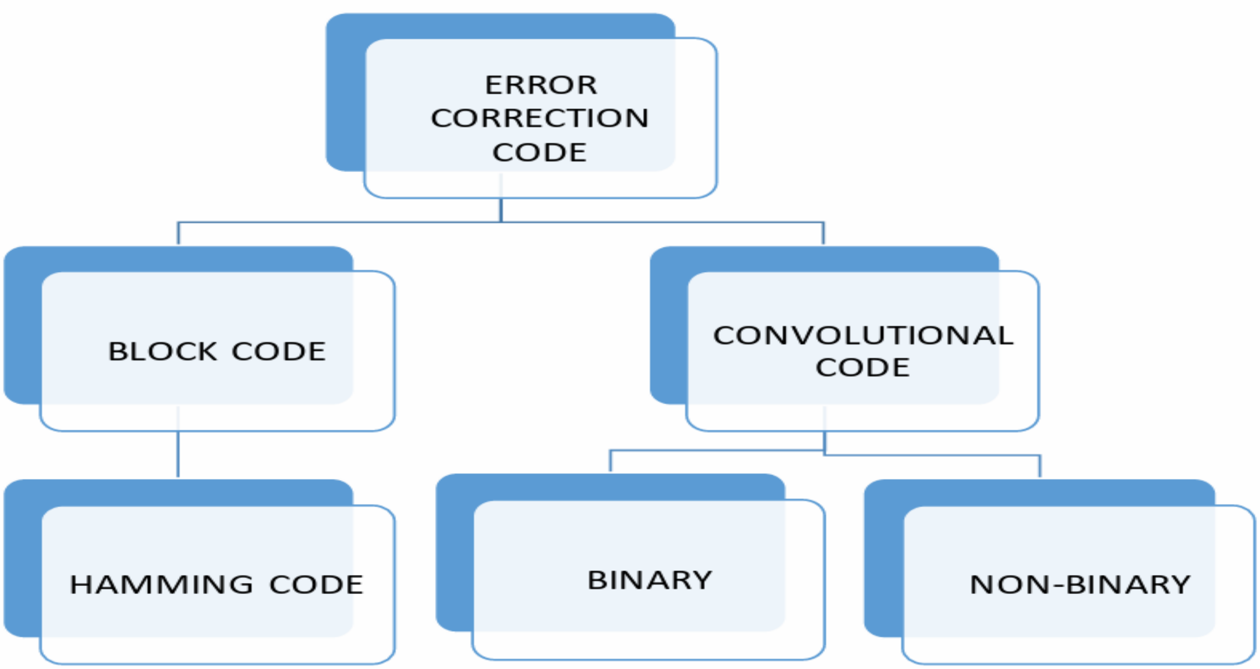
SOLUTION:-

a_i	$p(a_i)$	1	2	3	4	Code
a_1	0.36	0	00			00
a_2	0.18		01			01
a_3	0.18	1	10			10
a_4	0.12		11	110		110
a_5	0.09			111	1110	1110
a_6	0.07				1111	1111

ERROR DETECTION AND CORRECTION CODES:-

- In digital communication a bit ‘0’ may change to ‘1’ or vice versa due to noise
- Original data can be retrieved by first detecting the error and then correcting it.
- For this purpose we can use following codes

1. Error Detection Code-Detect errors(Parity Code , Hamming Code)
2. Error Correction Code-Find the correct number of corrupted bits and their positions in the message



PARITY BIT METHOD:-

• To detect error in the received message, we add some extra bits called redundant bits to the actual data

• A parity bit is an extra bit included in binary message to make total number of 1's either odd or even

1. Even Parity-Total number of 1's should be even

2. Odd Parity-Total number of 1's should be odd

Message	odd	Even
000	1	0
001	0	1
010	0	1
011	1	0
100	0	1
101	1	0
110	1	0
111	0	1

HAMMING CODE:-

• It is a Block Code-Message is divided in to fixed sized blocks , to which redundant bits are added

• It is used to detection and correction of error

• Here we sent data along with parity bits or redundant bits

• It is represented by (n,k) code ,where n-total bit in a block of code ,k message bit or data bit

ie, parity bit, r =n-k

• To identify parity bit r it should satisfy the following condition $2^r \geq k + r + 1$

• let us consider ASCII code having 7 bit data ie, k=7,from above condition we get r=4 that is 4 parity bits

• In the total bits transmitted in a block code power of 2 marked as PARITY BITS(1,2,4,8….)

• We can choose odd and even parity is there

Algorithm:-

1. Write the bit positions starting from 1 in binary form (1, 10, 11, 100, etc).

2. All the bit positions that are a power of 2 are marked as parity bits (1, 2, 4, 8, etc).

3. All the other bit positions are marked as data bits.

4. Each data bit is included in a unique set of parity bits, as determined its bit position in binary form

Parity bit 1---Covers all bit position whose binary representation includes a ‘1’ in LSB(1,3,5,7,9,11..)

Parity bit 2---Covers all bit position whose binary representation includes a ‘1’ in second position from LSB(2,3,6,7,10,11,15..)

Parity bit 4---Covers all bit position whose binary representation includes a ‘1’ in third position from LSB(4,5,6,7,12,13,14,15..)

• Parity bit 8---Covers all bit position whose binary representation includes a ‘1’ in fourth position from LSB(8,9,10,11,12,13,14,15..)

• Each parity bit covers all bits where the bitwise AND of the parity position and the bit position is non-zero

Position	R8	R4	R2	R1
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1

R1 -> 1,3,5,7,9,11
R2 -> 2,3,6,7,10,11
R3 -> 4,5,6,7
R4 -> 8,9,10,11

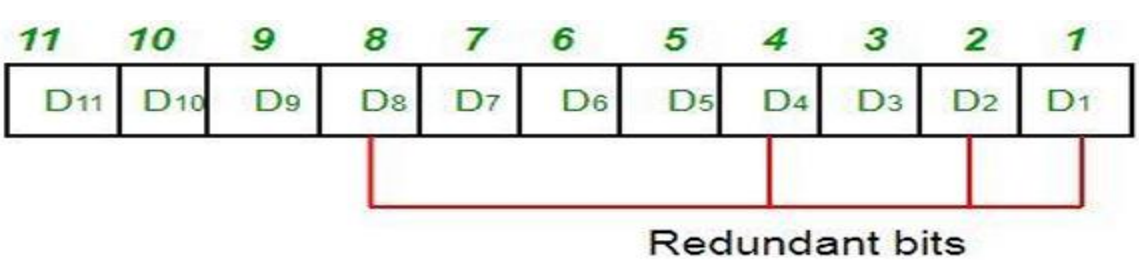
Determining the position of redundant bits:-

• The number of data bits = 7

• The number of redundant bits = 4

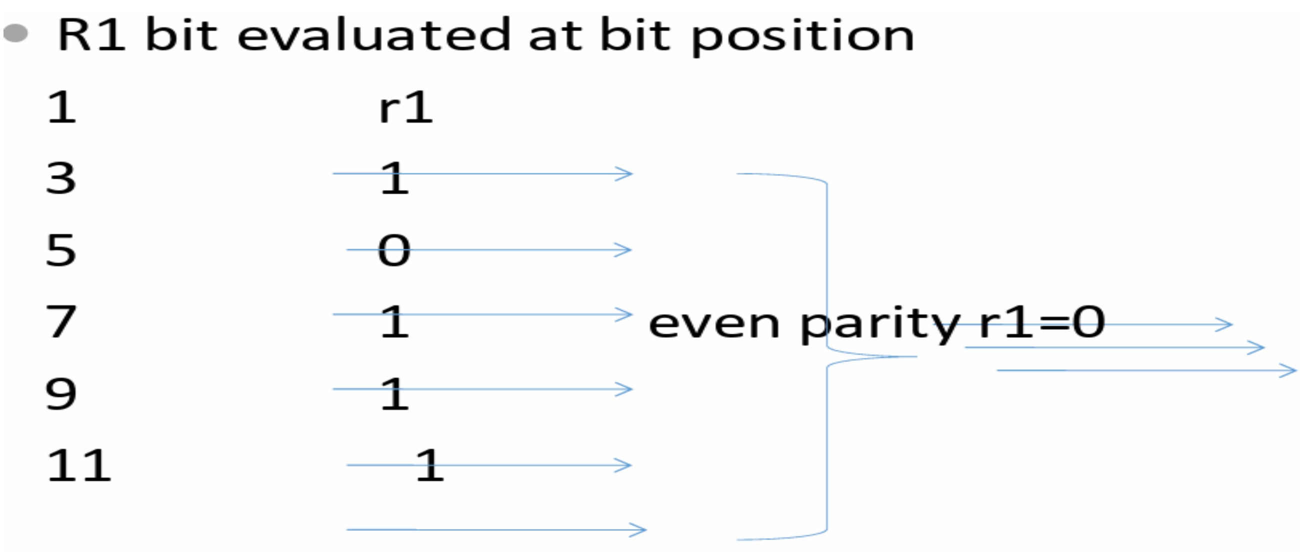
• The total number of bits = 11

• The redundant bits are placed at positions corresponding to power of 2- 1, 2, 4, and 8



Suppose the data to be transmitted is 1011001, the bits will be placed as follows:-

11	10	9	8	7	6	5	4	3	2	1
1	0	1	R8	1	0	0	R4	1	R2	R1



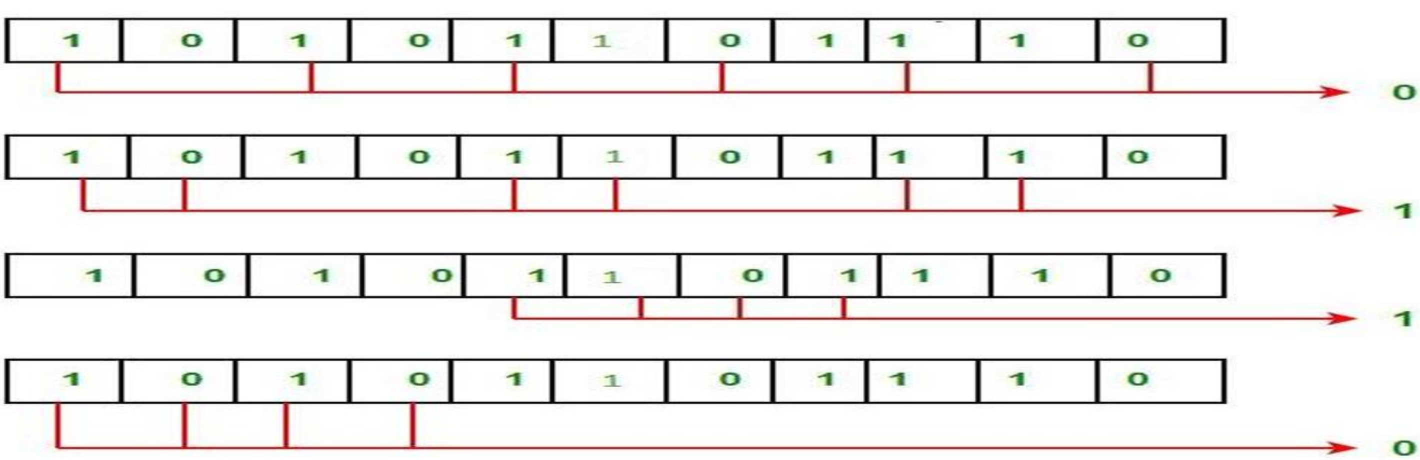
Similarly r2=1,r4=1,r8=0

Data transferred as

11	10	9	8	7	6	5	4	3	2	1
1	0	1	0	1	0	0	1	1	1	0

Error detection and correction:-

• Suppose in the above example the 6th bit is changed from 0 to 1 during data transmission, then it gives new parity values in the binary number:



• The bits give the binary number 0110 whose decimal representation is 6. Thus, bit 6 contains an error. To correct the error the 6th bit is changed from 1 to 0.

• Hamming code is designed to detect and correct single-bit errors that may occur during the transmission of data

An even parity hamming code for 5 data bits is transmitted and received. The received codeword is 110001101.Detect the error and correct it.

Example

• An even parity hamming code for 5 data bits is transmitted and received. The received code word is 110001101.Detect the error and correct it.

• If received hamming code is 1110101 with even parity the detect and correct error

LINEAR BLOCK CODES

• Definition :a block code is said to be linear code if its code words satisfy the condition that the sum of any two code words gives another code word

ie, $C_p = C_i + C_k$

• Type of error-correcting code used in digital communication

• Generator and parity matrix are there

• Properties

1. Linearity

2. Error Detection and Correction

3. Generator and Parity-Check Matrices

4. Error Correction Algorithms

5. Code rate: Ratio of the number of information bits (k) to the total number of bits (n) in a code word.

• Higher code rate higher efficiency

• Efficient Implementation