A Course Material on

Information Theory and Coding

By

S.CHANDRAMOHAN
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SRI CHANDRASEKHARENDRA SARASWATHI
VISWA MAHAVIDYALAYA
(Deemed to be University established under section 3 of UGC act 1956)
ENATHUR, KANCHIPURAM – 631 561

INFORMATION THEORY AND CODING

Pre-requisite: Basic knowledge of Digital Communication

OBJECTIVE: To get exposed to information and entropy, compression technique, audio & video

UNIT I: INFORMATION THEORY

Information – Entropy, Information rate, classification of codes, Kraft McMillan inequality, Source coding theorem, Shannon-Fano coding, Huffman coding, Extended Huffman coding - Joint and conditional entropies, Mutual information - Discrete memory less channels – BSC, BEC – Channel capacity, Shannon limit

UNIT II: ERROR CONTROL CODING:

BLOCK CODES Definitions and Principles: Hamming weight, Hamming distance, Minimum distance decoding - Single parity codes, Hamming codes, Repetition codes - Linear block codes, Cyclic codes - Syndrome calculation, Encoder and decoder - CRC

UNIT III: ERROR CONTROL CODING: CONVOLUTIONAL CODES

Convolution codes – code tree, trellis, state diagram - Encoding – Decoding: Sequential search and Viterbi algorithm – Principle of Turbo coding

UNIT IV: SOURCE CODING:

TEXT, AUDIO AND SPEECH Text: Adaptive Huffman Coding, Arithmetic Coding, LZW algorithm – Audio: Perceptual coding, Masking techniques, Psychoacoustic model, MEG Audio layers I,II,III, Dolby AC3 - Speech: Channel Vocoder, Linear Predictive Coding

UNIT V: SOURCE CODING:

IMAGE AND VIDEO Image and Video Formats – GIF, TIFF, SIF, CIF, QCIF – Image compression: READ, JPEG – Video Compression: Principles-I, B, P frames, Motion estimation, Motion compensation, H.261, MPEG standard

Text Books: 1. Ranjan Bose, Information Theory, Coding and Cryptography, Publication,2005.

2. Cover, Thomas, and Joy Thomas. Elements of Information Theory. 2nd ed. New York, NY: Wiley-Interscience, 2006. ISBN: 9780471241959

E books and online learning materials:

1. http://www-public.tem-tsp.eu/~uro/cours-pdf/poly.pdf

2. http://www.cl.cam.ac.uk/teaching/0910/InfoTheory/InfoTheoryLectures.pdf

INFORMATION THEORY:– Entropy, Information rate, classification of codes, Kraft McMillan inequality, Source coding theorem, Shannon-Fano coding, Huffman coding, Extended Huffman coding - Joint and conditional entropies, Mutual information - Discrete memory less channels – BSC, BEC – Channel capacity, Shannon limit

(i) Definition

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. A practical source in a communication system is a device which produces messages, and it can be either analog or discrete.

to discrete sources through the use of sampling ...

10. As a matter of fact, a discrete information source is a source which has only a finite set of symbols as possible outputs. The set of source symbols is called the **source alphabet**, and the elements of the set are called **symbols** or **letters.**

(ii) Classification of Information Sources

Information sources can be classified as having memory or being memoryless. A source with memory is one for which a current symbol depends on the previous symbols. A memoryless source is one for which each symbol produced is independent of the previous symbols.

A discrete memoryless source (DMS) can be characterized by the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source.

DO YOU KNOW?

A discrete information source consists of a discrete set of letters or alphabet of symbols. In general, any message emitted by the source consists of a string or sequence of symbols.

****

Information is the source of a communication system, whether it is analog or digital.

Information theory is a mathematical approach to the study of coding of information along with the quantification, storage, and communication of information.

Conditions of Occurrence of Events

If we consider an event, there are three conditions of occurrence.

- If the event has not occurred, there is a condition of **uncertainty**.
- If the event has just occurred, there is a condition of **surprise**.
- If the event has occurred, a time back, there is a condition of having some **information**.

These three events occur at different times. The differences in these conditions help us gain knowledge on the probabilities of the occurrence of events.

The performance of the communication system is measured in terms of its error probability. An errorless transmission is possible when probability of error at the receiver approaches zero.

The information theory is related to the concepts of statistical properties of messages/sources, channels, noise interference etc. The information theory is used for mathematical modeling and analysis of the communication systems.

1.1.1 Uncertainty

Consider the source which emits the discrete symbols randomly from the set of fixed alphabet i.e.

$$X = \{x_0, x_1, x_2, \ldots\ldots x_{K-1}\} \quad \ldots (1.1.1)$$

The various symbols in 'X ' have probabilities of p_0, p_1, p_2, ... etc., which can be written as,

$$P(X = x_k) = p_k \qquad k = 0, 1, 2, \ldots\ldots K - 1 \quad \ldots (1.1.2)$$

This set of probabilities satisfy the following condition,

$$\sum_{k=0}^{K-1} p_k = 1 \quad \ldots (1.1.3)$$

Such information source is called discrete information source. The concept of 'Information' produced by the source is discussed in the next section. This idea of Information is related to 'Uncertainty' or 'Surprise'. Consider the emission of symbol

* BASICS OF INFORMATION SYSTEM :

* Information System is defined as the message is generated from the information source and transmitted towards receiver through transmission medium. The block diagram of an information system can be drawn as shown in fig 1.1

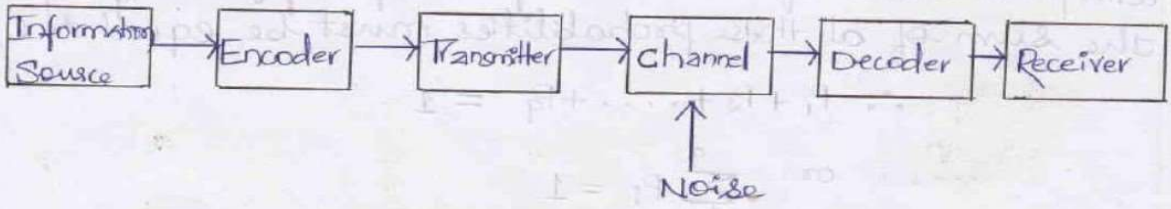


Fig 1.1 : Block diagram of an Information system.

Definition of Information:

* MEASURE OF INFORMATION :

Let us consider the communication system which transmits messages or independent sequence of symbols from source alphabet $S = \{s_1, s_2, \ldots, s_q\}$ with probabilities $P = \{P_1, P_2, \ldots, P_q\}$ respectively.

Let "s_k" be a symbol choosen for transmission at any instant of time with a probability equal to P_k. Then the "amount of information" or "self-Information" of message "s_k" is given by,

$$\text{Amount of Information; } I_k = \log_2\left(\frac{1}{P_k}\right)$$

* Unit of Information :

In the above equation $\log_2\left(\frac{1}{P_k}\right) = \frac{\log_{10}(1/P_k)}{\log_{10} 2}$, if the base of the logarithm is 2, then the units are called "BITS", which is the short form of Binary units". If the base is "10", the units are "HARTLEYS" or "DECITS". If the base is "e", the units are "NATS" and if the base, in general, is "r", the units are called "r-ary units".

* The most widely used unit of information is "BITS", where the base of the logarithm is 2.

INFORMATION RATE :

Let us suppose that, the symbols are emitted by the source at a fixed time rate "r_s" symbols/sec. The average source information rate R_s" in bits/sec is defined as the product of the average information content per symbol and the message symbol rate "r_s".

$$\therefore \boxed{R_s = r_s H(S)} \text{ bits/sec or BPS}$$

1.2.1 Properties of Information

Following properties can be written for information.

i) If there is more uncertainty about the message, information carried is also more.

ii) If receiver knows the message being transmitted, the amount of information carried is zero.

iii) If I_1 is the information carried by message m_1, and I_2 is the information carried by m_2, then amount of information carried combinely due to m_1 and m_2 is $I_1 + I_2$.

iv) If there are $M = 2^N$ equally likely messages, then amount of information carried by each message will be N bits.

Prove the following statement, "If receives knows the message being transmitted, the amount of information carried is "zero".

∴ Here it is stated that receives "knows" the message. This means only one message is transmitted. Hence probability of occurrence of this message will be $P_k = 1$.

Therefore, the amount of information carried by this type of message is,

$$I_k = \log_2\left(\frac{1}{P_k}\right)$$

$$= \frac{\log_{10} 1}{\log_{10} 2}$$

$$\therefore \quad I_k = 0 \text{ bits}$$

Example 1.2.1 : *Calculate the amount of information if* $p_k = \frac{1}{4}$.

Solution : From equation 1.2.1 we know that amount of information is given as,

$$I_k = \log_2\left(\frac{1}{p_k}\right) = \frac{\log_{10}\left(\frac{1}{p_k}\right)}{\log_{10} 2}$$

$$= \frac{\log_{10} 4}{\log_{10} 2}$$

$$= 2 \text{ bits}$$

➡ **Example 1.2.3 :** *In binary PCM if '0' occur with probability* $\frac{1}{4}$ *and '1' occur with probability* $\frac{3}{4}$*, then calculate amount of information conveyed by each binit.*

Solution : Here binit '0' has $p_1 = \frac{1}{4}$

and binit '1' has $p_2 = \frac{3}{4}$

Then amount of information is given by equation 1.2.1 as,

$$I_k = \log_2\left(\frac{1}{p_k}\right)$$

$$\left.\begin{aligned} \text{with} \quad & p_1 = \frac{1}{4},\ I_1 = \log_2 4 = \frac{\log_{10} 4}{\log_{10} 2} = 2 \text{ bits} \\ \text{and with} \quad & p_2 = \frac{3}{4},\ I_2 = \log_2\left(\frac{4}{3}\right) = \frac{\log_{10}(4/3)}{\log_{10} 2} = 0.415 \text{ bits} \end{aligned}\right\} \quad \ldots (1.2.4)$$

Here observe that binit '0' has probability $\frac{1}{4}$ and it carries 2 bits of information. Whereas binit '1' has probability $\frac{3}{4}$ and it carries 0.415 bits of information. This shows that if probability of occurrence is less, information carried is more, and vice versa.

ENTROPY:

∴ Average self-information is also called "ENTROPY" of source's denoted by H(S).

$$\therefore \boxed{H(S) = \sum_{i=1}^{q} P_i \log \frac{1}{P_i}} \text{ bits/message symbol}$$

Illustration 1 :

Let us consider a binary source with source alphabet $S = \{s_1, s_2\}$ with probabilities $P = \left\{\frac{1}{256}, \frac{255}{256}\right\}$

Then, Entropy $H(S) = \sum_{i=1}^{2} P_i \log \frac{1}{P_i}$

$$= \frac{1}{256} \log_2 256 + \frac{255}{256} \log_2 \frac{256}{255}$$

$$\therefore \boxed{H(S) = 0.037 \text{ bits/msg symbol.}}$$

∴ The average uncertainty is very very small and is relatively very very easy to guess whether s_1 or s_2 will occur.

Illustration 2 :

Let $S' = \{s_3, s_4\}$ with $P' = \left\{\frac{7}{16}, \frac{9}{16}\right\}$

Then, Entropy $H(S') = \frac{7}{16} \log_2 \frac{16}{7} + \frac{9}{16} \log_2 \frac{16}{9}$

$$\therefore \boxed{H(S') = 0.989 \text{ bits/msg symbol}}$$

In this case, it is hard to guess whether s_3 or s_4 is transmitted

Illustration 3 :

Let $S'' = \{s_5, s_6\}$ with $P'' = \{1/2, 1/2\}$

Then, Entropy $H(S'') = \frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2$

$$\therefore \boxed{H(S'') = 1 \text{ bits/msg symbol}}$$

In this case, the uncertainty is maximum for a binary source and becomes impossible to guess which symbol is transmitted

Consider a source $S = \{S_1, S_2, S_3\}$ with $P = \{1/2, 1/4, 1/4\}$

Find (a) Self Information of each message

(b) Entropy of source 'S'.

(a) Self Information of $S_1 = I_1 = \log_2 \frac{1}{P_1} = \log_2 2 = 1$ bits

self Information of $S_2 = I_2 = \log_2 \frac{1}{P_2} = \log_2 4 = 2$ bits

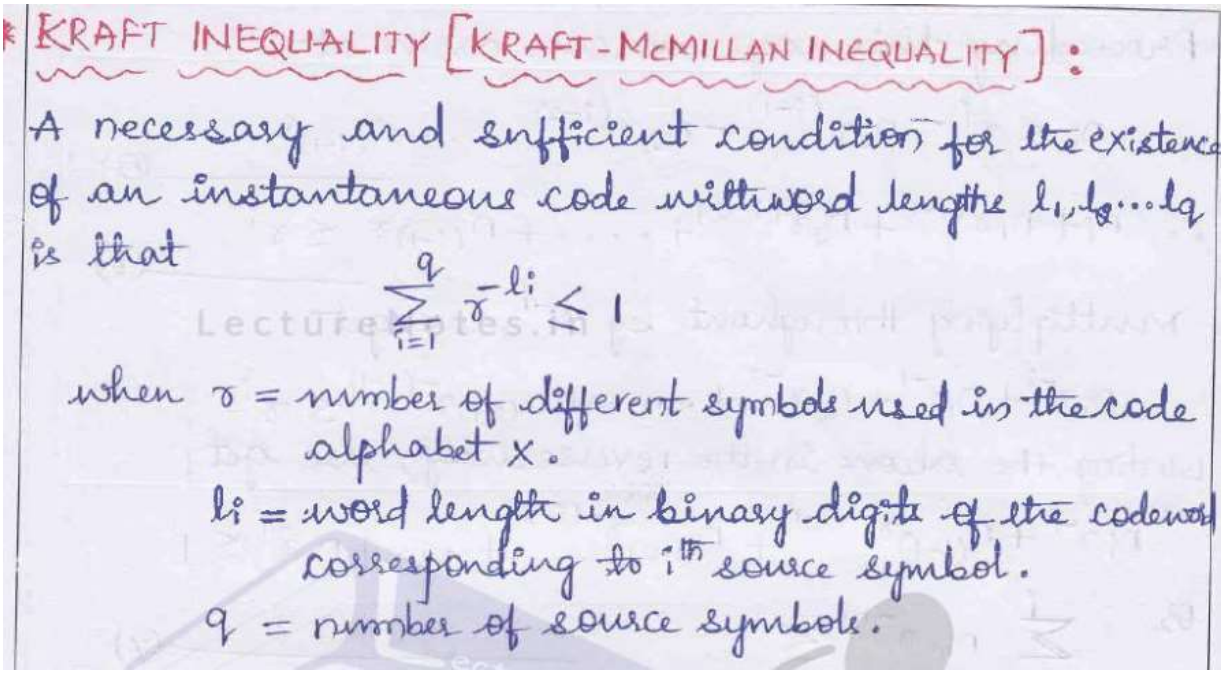
self Information of $S_3 = I_3 = \log_2 \frac{1}{P_3} = \log_2 4 = 2$ bits

b). Average Information content or Entropy is given by,

$$H(S) = \sum_{i=1}^{3} P_i I_i = P_1 I_1 + P_2 I_2 + P_3 I_3$$

$$= \left(\frac{1}{2}\right)(1) + \frac{1}{4}(2) + \frac{1}{4}(2)$$

$\therefore$ $\boxed{H(S) = 1.5 \text{ bits/msg symbol}}$



Entropy: When we observe the possibilities of the occurrence of an event, how surprising or uncertain it would be, it means that we are trying to have an idea on the average content of the information from the source of the event.

Entropy can be defined as a measure of the average information content per source symbol.

$$H = -\sum_i p_i \log_b p_i$$

Where $\mathbf{p_i}$ is the probability of the occurrence of character number **i** from a given stream of characters and **b** is the base of the algorithm used. Hence, this is also called as **Shannon's Entropy**.

Conditional Entropy: The amount of uncertainty remaining about the channel input after observing the channel output, is called as Conditional Entropy.

It is denoted by $H(x \mid y)$

15.10 THE CONDITIONAL AND JOINT ENTROPIES

Using the input probabilites $P(x_i)$, output probabilities $P(y_i)$, transition probabilities $P(y_j | x_i)$, and joint probabilities $P(x_i, y_j)$, let us define the following various entropy functions for a channel with m inputs and n outputs:

$$H(X) = -\sum_{i=1}^{m} P(x_i) \log_2 P(x_i) \quad ...(15.23)$$

$$H(Y) = -\sum_{j=1}^{n} P(y_j) \log_2 P(y_j) \quad ...(15.24)$$

$$H(X|Y) = -\sum_{j=1}^{n} \sum_{i=1}^{m} P(x_i, y_j) \log_2 P(x_i | y_j) \quad ...(15.25)$$

$$H(Y|X) = -\sum_{j=1}^{n} \sum_{i=1}^{m} P(x_i, y_j) \log_2 P(y_j | x_i) \quad ...(15.26)$$

$$H(X, Y) = -\sum_{j=1}^{n} \sum_{i=1}^{m} P(x_i, y_j) \log_2 P(x_i, y_j) \quad ...(15.27)$$

These entropies can be interpreted as under:

$H(X)$ 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. Also, $H(X|Y)$ is sometimes called the *equivocation* of X with respect to Y. The conditional entropy $H(Y|X)$ is the average uncertainty of the channel output given that X was transmitted. The joint entropy $H(Y|Y)$ is the average uncertainty of the communication channel as a whole. Two useful relationships among the above various entropies are as under:

DO YOU KNOW?

The maximum rate of transmission occurs when the source is matched to the channel.

$$H(X, Y) = H(X|Y) + H(Y) \quad ...(15.28)$$

$$H(X,Y) = H(Y|X) + H(X) \quad ...(15.29)$$

15.11 THE MUTUAL INFORMATION

The *mutual information* denoted by $I(X; Y)$ of a channel is defined by

$$I(X; Y) = H(X) - H(X|Y) \text{ b/symbol} \quad ...(15.30)$$

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.

15.12.1. Channel Capacity Per Symbol C_s

The *channel capacity per symbol* of a discrete memoryless channel (DMC) is defined as

$$C_s = \max_{\{P(x_i)\}} I(X;Y) \text{ b/symbol} \qquad ...(15.35)$$

where the maximization is over all possible input probability distributions $\{P(x_i)\}$ on X. Note that the channel capacity C_s is a function of only the channel transition probabilities which define the channel.

15.12.2. Channel Capacity Per Second C

If r symbols are being transmitted per second, then the maximum rate of transmission of information per second is rC_s. This is the *channel capacity per second* and is denoted by $C(b/s)$, *i.e.*,

$$C = rC_s \text{ b/s} \qquad ...(15.36)$$

Mutual Information :

Let us consider a channel whose output is **Y** and input is **X**

Let the entropy for prior uncertainty be **X** = **H**x

Thisisassumedbeforetheinputisapplied

To know about the uncertainty of the output, after the input is applied, let us consider Conditional Entropy, given that **Y** = $\mathbf{y_k}$

$$H(x \mid y_k) = \sum_{j=0}^{j-1} p(x_j \mid y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]$$

This is a random variable for

$H(X \mid y = y_0) \ldots \ldots \ldots \ldots \ldots H(X \mid y = y_k)$ With probabilities p (yo)….p(y_{k-1}) respectively

The mean value of $H(X \mid y = y_k)$ for output alphabet **y** is –

$$H(X \mid Y) = \sum_{k=0}^{k-1} H(X \mid y = y_k)\, p(y_k)$$

$$= \sum_{k=0}^{k-1} \sum_{j=0}^{j-1} p(x_j \mid y_k)\, p(y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]$$

$$= \sum_{k=0}^{k-1} \sum_{j=0}^{j-1} p(x_j, y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]$$

Now, considering both the uncertainty conditions beforeandafterapplyingtheinputsbeforeandafterapplyingtheinputs,

we come to know that the difference, i.e. $H(x) - H(x|y)$must represent the uncertainty about the channel input that is resolved by observing the channel output.

This is called as the **Mutual Information** of the channel.

Denoting the Mutual Information as $I(x;y)$,we can write the whole thing in an equation, as follows

$$I(x;y) = H(x) - H(x \mid y)$$

Hence, this is the equational representation of Mutual Information.

Discrete Memoryless Channel (DMC):

15.8.1. Channel Representation

A communication channel may be defined as the path or medium through which the symbols flow to the receiver end. A *discrete memoryless channel* (DMC) is a statistical model with an input X and an output Y as shown in figure 15.1. During each unit of the time (signaling interval), the channel accepts an input symbol from X, and in response it generates an output symbol from Y. The channel is said to be "discrete" when the alphabets of X and Y are both finite. Also, it is said to be "memoryless" when the current output depends on only the current input and not on any of the previous inputs.

A diagram of a DMC with m inputs and n outputs has been illustrated in figure 15.1. The input X consists of input symbols $x_1, x_2, \ldots x_m$. The a priori probabilities of these source symbols $P(x_i)$ are assumed to be known. The outputs Y consists of output symbols $y_1, y_2, \ldots y_n$. 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_i, and is called a **channel transition probability.**

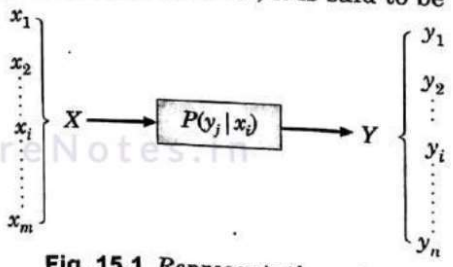


Fig. 15.1 *Representation of a discrete memoryless channel (DMC).*

15.8.2. The Channel Matrix

A channel is completely specified by the complete set of transition probabilities. Accordingly, the channel in figure 15.1 is often specified by the matrix of transition probabilities $[P(Y|X)]$. This matrix is given by

$$[P(Y|X)] = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) & \cdots & P(y_n|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & \cdots & P(y_n|x_2) \\ \cdots & \cdots & \cdots & \cdots \\ P(y_1|x_m) & P(y_2|x_m) & \ldots & P(y_n|x_m) \end{bmatrix} \quad \ldots(15.13)$$

This matrix $[P(Y|X)]$ is called the **channel matrix.**

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

$$\sum_{j=1}^{n} P(y_j | x_i) = 1 \text{ for all } i \quad \ldots(15.14)$$

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

$$[P(X)] = [P(x_1)\ P(x_2) \ldots P(x_m)] \quad \ldots(15.15)$$

Also, the output probabilities $P(Y)$ are represented by the row matrix as under:

$$[P(Y)] = [P(y_1)\ P(y_2) \ldots P(y_n)] \quad \ldots(15.16)$$

then

$$[P(Y)] = [P(X)][P(Y|X)] \quad \ldots(15.17)$$

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

$$[P(X)]_d = \begin{bmatrix} P(x_1) & 0 & \cdots & 0 \\ 0 & P(x_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & P(x_m) \end{bmatrix} \quad \ldots(15.18)$$

then $$[P(X,Y)] = [P(X)]_d\,[P(Y|X)] \quad \ldots(15.19)$$

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.

15.9.1. Lossless Channel

A channel described by a channel matrix with only one non-zero element in each column is called a *lossless channel*. An example of a lossless channel has been shown in figure 15.2, and the corresponding channel matrix is given in equation (15.20) as under:

$$[P(Y|X)] = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \ldots(15.20)$$

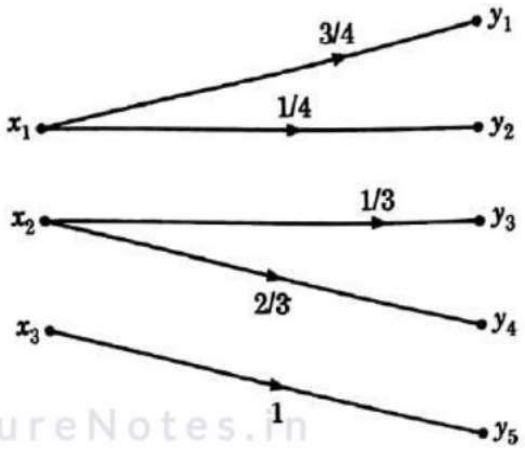


Fig1: Noiseless Channel

It can be shown that in the lossless channel, no source information is lost in transmission.

15.9.2. Deterministic Channel

A channel described by a channel matrix with only one non-zero element in each row is called a *deterministic channel*. An example of a deterministic channel has been shown in figure 15.3, and the corresponding channel matrix is given by equation (15.21) as under:

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \ldots(15.21)$$

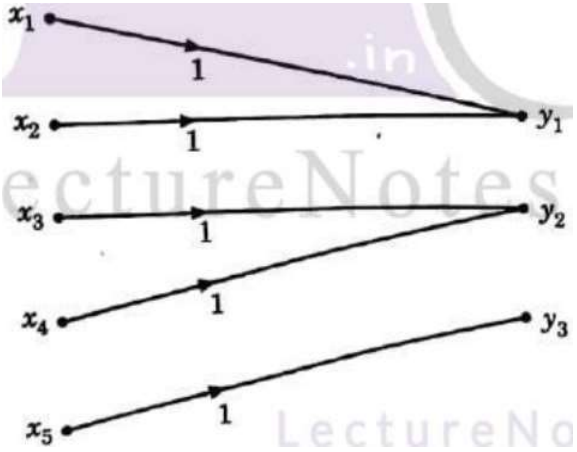


Fig2: Deterministic Channel

Important Point: *It may be noted that since each row has only one non-zero element, therefore, this element must be unity by equation (15.14). Thus , when a given source symbol is sent in the deterministic channel, it is clear which output symbol will be received.*

15.9.3. Noiseless Channel

A channel is called *noiseless* if it is both lossless and deterministic. A noiseless channel has been shown in figure 15.4. The channel matrix has only one element in each row and in each column, and this element is unity. Note that the input and output alphabets are of the same size, that is, $m = n$ for the noiseless channel.

The matrix for a noiseless channel is given by

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

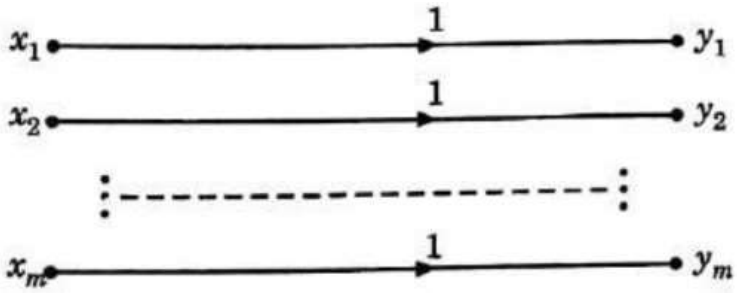


Fig2: Noiseless Channel

15.9.4. Binary Symmetric Channel (BSC)

The binary symmetric channel (BSC) is defined by the channel diagram shown in figure 15.5, and its channel matrix is given by

$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \quad \ldots(15.22)$$

A BSC channel has two inputs ($x_1 = 0, x_2 = 1$) and two outputs ($y_1 = 0, y_2 = 1$). 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. This common transition probability is denoted by p as shown in figure 15.5.

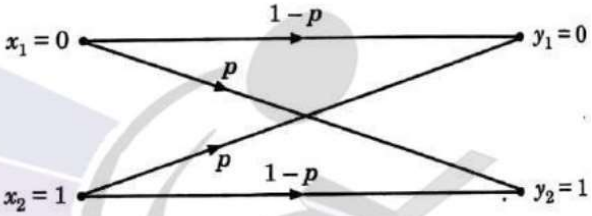


Fig. 15.5 *Binary symmetrical channel.*

The binary symbols '0' and '1' are transmitted with probabilities 1/4 and 3/4 respectively. Find the corresponding self Informations.

∴ Self Information in a '0' $= I_0 = \log \frac{1}{P_0} = \log 4$

$$\Rightarrow I_0 = \frac{\log 4}{\log 2} = 2 \text{ bits}$$

$$\therefore \boxed{I_0 = 2 \text{ bits}}$$

Self Information in a '1' $= I_1 = \log \frac{1}{P_1} = \log_2 \left(\frac{4}{3}\right)$

$$\Rightarrow I_1 = \frac{\log_{10} \frac{4}{3}}{\log_{10} 2}$$

$$\therefore \boxed{I_1 = 0.415 \text{ bits}}$$

Thus it can be observed that more information is carried by a less likely message.

3) If there are M equally likely and independent messages, then Prove that amount of information carried by each message will be $I = N$ bits, where $M = 2^N$ and N is an integer.

Soln: Since all the messages are equally likely and independent, probability of occurrence of each message will be $\frac{1}{M}$.

We know that, the amount of information is given by,

$$I_K = \log_2 \frac{1}{P_K}$$

Since $P_K = \frac{1}{M}$ then $I_K = \log_2 M$

As wkt, $M = 2^N$, Hence above equation becomes,

$$I_K = \log_2 2^N$$

$$= N \log_2 2 = N \frac{\log_{10} 2}{\log_{10} 2}$$

$$\Rightarrow \boxed{I_K = N \text{ bits}}$$ Hence the proof.

Prove the following statement, "If receiver knows the message being transmitted, the amount of information carried is "Zero".

Soln: Here it is stated that receiver "knows" the message. This means only one message is transmitted. Hence probability of occurrence of this message will be $P_K = 1$.

Therefore, the amount of information carried by this type of message is,

$$I_K = \log_2\left(\frac{1}{P_K}\right)$$

$$= \frac{\log_{10} 1}{\log_{10} 2}$$

$$\therefore \boxed{I_K = 0 \text{ bits}}$$

Justification for logarithmic measure of Information

Logarithmic expression is chosen for measuring information because of the following reasons :

1) The information content or self-information of any message cannot be negative. Each message must contain certain amount of information.
2) The lowest possible self-information is "zero" which occurs for a sure event, since P(sure Event) = 1 .
   Eg : "zero-Level Talk"
3) More information is carried by a less likely message.
4) When independent symbols are transmitted, the total self-information must be equal to the sum of individual self-information.

Comment on the information content of the following messages

i) Tomorrow the sun will rise from the east

ii) It will snow in Bangalore this winter

iii) The phone will ring in the next one hour.

Ans: Information content of the messages:

i) The first statement does not carry any information since it is sure that sun always rises from east. The probability of occurrance of first event is high or sure. Hence it carries less or negligible information.

$$\text{i.e } I_k = \log_2 \frac{1}{P_k} = \log_2 \frac{1}{1} = 0$$

ii) In the winter season snow fall in Bangalore is very rare. Hence probability of occurrance of this event is very rare. So it carries large amount of information.

iii) The third statement predicts about phone ring in the time span of one hour. It doesnot mention exact time but span of one hour is mentioned. Hence it carries moderate information.

AVERAGE INFORMATION CONTENT (ENTROPY) OF SYMBOLS IN LONG INDEPENDENT SEQUENCES :

Let us consider a zero memory source producing independent sequences of symbols with source alphabet $S = \{s_1, s_2, \ldots, s_q\}$ with probabilities $P = \{P_1, P_2, \ldots P_q\}$ respectively.

Let us consider a long independent sequence of length 'L' symbols. This long sequence then contains

$P_1 L$ number of messages of type s_1,
$P_2 L$ number of messages of type s_2,
⋮ ⋮ ⋮ ⋮
and $P_q L$ number of messages of type s_q.

Wkt, the self Information of $S_1 = \log \frac{1}{P_1}$ bits

$\therefore$ $P_1 L$ number of messages of type S_1 contain $P_1 L \log \frac{1}{P_1 L}$ bits of information

$P_2 L$ number of messages of type S_2 contain $P_2 L \log \frac{1}{P_2 L}$ bits of information

$P_q L$ number of messages of type S_q contain $P_q L \log \frac{1}{P_q L}$ bits of information

$\therefore$ The total self-information content of all these message symbols is given by,

$$I_{total} = P_1 L \log \frac{1}{P_1} + P_2 L \log \frac{1}{P_2} + \cdots + P_q L \log \frac{1}{P_q} \text{ bits}$$

$$\therefore I_{total} = L \sum_{i=1}^{q} P_i \log \frac{1}{P_i} \text{ bits}$$

and Average self-information $= \frac{I_{total}}{L}$

$$= \sum_{i=1}^{q} P_i \log \frac{1}{P_i} \text{ bits/msg symbol}$$

$\therefore$ Average self-information is also called "ENTROPY" of source 'S' denoted by H(S).

$$\therefore \boxed{H(S) = \sum_{i=1}^{q} P_i \log \frac{1}{P_i}} \text{ bits/message symbol}$$

Thus, H(S) represents the "average uncertainty" or the "average amount of surprise" of the source.

INFORMATION RATE :

Let us suppose that the symbols are emitted by the source at a fixed time rate "r_s" symbols/sec. The average source information rate R_s" in bits/sec is defined as the product of the average information content per symbol and the message symbol rate "r_s".

$$\therefore \boxed{R_s = r_s H(S)} \text{ bits/sec or BPS}$$

Consider a source $S = \{S_1, S_2, S_3\}$ with $P = \{1/2, 1/4, 1/4\}$

Find (a) Self Information of each message

(b) Entropy of source 'S'.

(a) Self Information of $S_1 = I_1 = \log_2 \frac{1}{P_1} = \log_2 2 = 1$ bits

Self Information of $S_2 = I_2 = \log_2 \frac{1}{P_2} = \log_2 4 = 2$ bits

Self Information of $S_3 = I_3 = \log_2 \frac{1}{P_3} = \log_2 4 = 2$ bits

b). Average Information content or Entropy is given by,

$$H(S) = \sum_{i=1}^{3} P_i I_i = P_1 I_1 + P_2 I_2 + P_3 I_3$$

$$= \left(\frac{1}{2}\right)(1) + \frac{1}{4}(2) + \frac{1}{4}(2)$$

$$\therefore \boxed{H(S) = 1.5 \text{ bits/msg symbol}}$$

* A source emits one of four symbols s_0, s_1, s_2 and s_3 with probabilities $\frac{1}{3}, \frac{1}{6}, \frac{1}{4}$ and $\frac{1}{4}$ respectively. The successive symbols emitted by the source are statistically independent. Calculate the entropy of the source.

Sol: The entropy of the source is given by,

$$H(S) = \sum_{i=0}^{3} P_i \log \frac{1}{P_i}$$

$$= P_0 \log_2 \frac{1}{P_0} + P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + P_3 \log_2 \frac{1}{P_3}$$

$$= \frac{1}{3} \log_2 3 + \frac{1}{6} \log_2 6 + \frac{1}{4} \log_2 4 + \frac{1}{4} \log_2 4$$

$$= 0.5282 + 0.43082 + 0.5 + 0.5$$

$\therefore$ $\boxed{H(S) = 1.95914 \text{ bits/msg symbol}}$

A source emits one of 4 possible symbols "x_0 to x_3" during each signalling interval. The symbols occur with probabilities as given in table below:

Symbol	Probability
x_0	$P_0 = 0.4$
x_1	$P_1 = 0.3$
x_2	$P_2 = 0.2$
x_3	$P_3 = 0.1$

Find the amount of information gained by observing the source emitting each of these symbols and also the entropy of source.

∴ The self-information "I_k" is given by

$$I_k = \log_2 \frac{1}{P_k} \text{ bits} \quad \text{---------- (1)}$$

Since we have four symbols, so $K = 0, 1, 2, 3$

∴ when $K=0$, $I_0 = \log_2 \frac{1}{P_0} = \log_2 \frac{1}{0.4} = 1.322$ bits

$K=1$, $I_1 = \log_2 1/P_1 = \log_2 1/0.3 = 1.737$ bits

$K=2$, $I_2 = \log_2 1/P_2 = \log_2 1/0.2 = 2.322$ bits

$K=3$, $I_3 = \log_2 1/P_3 = \log_2 1/0.1 = 3.322$ bits

The entropy of the source is given by,

$$H(X) = \sum_{K=0}^{3} P_K \log \frac{1}{P_K} \text{ bits/msg symbol}$$

$$= \sum_{K=0}^{3} P_K I_K \qquad \because I_K = \log \frac{1}{P_K}$$

$$= P_0 I_0 + P_1 I_1 + P_2 I_2 + P_3 I_3$$

$$= 0.4(1.322) + 0.3(1.737) + 0.2(2.322) + 0.1(3.322)$$

$$\therefore \boxed{H(X) = 1.8465 \text{ bits/msg symbol}}$$

Next Page 72 ?????

SHANNON- FANO CODING:

* SHANNON – FANO ENCODING ALGORITHM :

Shannon – fano Encoding procedure for getting a compact code with minimum redundancy is given below :

i) The symbols are arranged according to non-increasing probabilities.

ii) The symbols are divided into two groups so that the sum of probabilities in each group is approximately equal.

iii) All the symbols in the 1st group are designated by "1" and the 2nd group by "0".

iv) The 1st group is again subdivided into two subgroups such that each subgroup probabilities are approximately same.

v) All the symbols of the 1st subgroup are designated by "1" and 2nd subgroup by "0".

vi) The second subgroup is subdivided into two more subgroups and step (v) is repeated.

vii) This process is continued till further sub-division is impossible

Given the messages X_1, X_2, X_3, X_4, X_5 and X_6 with respective probabilities 0.4, 0.2, 0.2, 0.1, 0.07 and 0.03, construct a binary code by applying shannon-fano Encoding procedure. Determine code efficiency and redundancy of the code.

Sol: Applying Shannon-Fano Encoding Procedure

X_1	0.4	1									
X_2	0.2	0	0.2	1	0.4	1					
X_3	0.2	0	0.2	1	0.2	0					
X_4	0.1	0	0.1	0	0.1	1					
X_5	0.07	0	0.07	0	0.07	0	0.07	1			
X_6	0.03	0	0.03	0	0.03	0	0.03	0			

code	l_i in bits
1	1
011	3
010	3
001	3
0001	4
0000	4

You are given 4 messages X_1, X_2, X_3 and X_4 with respective probabilities 0.1, 0.2, 0.3, 0.4

i) Device a code with prefix property (shannon-fano code) for these messages and draw the code tree.

ii) Calculate the efficiency and redundancy of the code.

iii) calculate the probabilities of 0's and 1's in the code.

Sol:

X_4	0.4	1					
X_3	0.3	0	0.3	1			
X_2	0.2	0	0.2	0	0.2	1	
X_1	0.1	0	0.1	0	0.1	0	

code	l_i in bits
1	1
01	2
001	3
000	3

Now, Average length "L" is given by

$$L = \sum_{i=1}^{6} P_i l_i$$

$$= 0.4\times1 + 0.2\times3 + 0.2\times3 + 0.1\times3 + 0.07\times4 + 0.03\times4$$

$$\therefore \boxed{L = 2.3 \text{ bits/msg symbol}}$$

Entropy, $H(S) = \sum_{i=1}^{6} P_i \log 1/P_i$

$$= 0.4\log 1/0.4 + 2\times0.2\log\frac{1}{0.2} + 0.1\log\frac{1}{0.1} + 0.07\log 1/0.07 + 0.03\log\frac{1}{0.03}$$

$$\therefore \boxed{H(S) = 2.209 \text{ bits/msg symbol}}$$

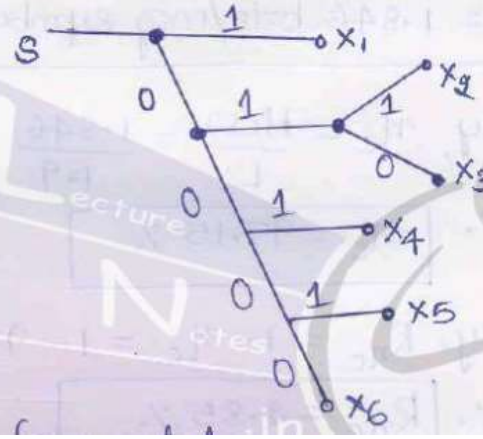
$\therefore$ Code Efficiency $\eta_c = \dfrac{H(S)}{L} = \dfrac{2.209}{2.3}$

$$\therefore \boxed{\eta_c = 96.04\%}$$

$\therefore$ Code Redundancy, $R_{\eta_c} = 1 - \eta_c$

$$\therefore \boxed{R_{\eta_c} = 3.96\%}$$

The code tree is constructed as shown below.



The code tree can be drawn as shown below.

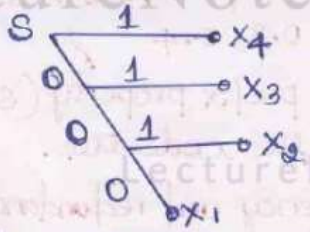


fig : code tree

> The probability of 0's and 1's in the code are found using

$$P(0) = \frac{1}{L}\sum_{i=1}^{4}(\text{no. of 0's in code } x_i)(P_i) = \frac{1}{1.9}\left[3\times0.1 + 2(0.2) + 1\times0.3 + 0(0.4)\right]$$

$$\therefore \boxed{P(0) = 0.5263}$$

and $$P(1) = \frac{1}{L}\sum_{i=1}^{4}(\text{no. of 1's in code } x_i)(P_i) = \frac{1}{1.9}\left[0\times0.1 + 1\times0.2 + 1\times0.3 + 1\times0.4\right]$$

$$\therefore \boxed{P(1) = 0.4737}$$

> Average Length, $L = \sum_{i=1}^{4} P_i L_i$

$$= (0.4)(1) + (0.3)(2) + (0.2)(3) + (0.1)(3)$$

$$\therefore \boxed{L = 1.9 \text{ bits/msg symbol}}$$

Entropy, $H(S) = \sum_{i=1}^{4} P_i \log \frac{1}{P_i}$

$$= 0.4 \log \frac{1}{0.4} + 0.3 \log \frac{1}{0.3} + 0.2 \log \frac{1}{0.2} + 0.1 \log \frac{1}{0.1}$$

$$\therefore \boxed{H(S) = 1.846 \text{ bits/msg symbol}}$$

$\therefore$ Code efficiency, $\eta_c = \frac{H(S)}{L} = \frac{1.846}{1.9}$

$$\therefore \boxed{\eta_c = 97.15\ \%}$$

$\therefore$ Code Redundancy, $R_{nc} = 1 - \eta_c$

$$\therefore \boxed{R_{nc} = 2.85\ \%}$$

length shift register and they are combined with the help of mod-2 adders. This operation is equivalent to binary convolution and hence it is called *convolutional coding*. This concept is illustrated with the help of simple example given below.

Fig. 5.1.1 shows a convolutional encoder.

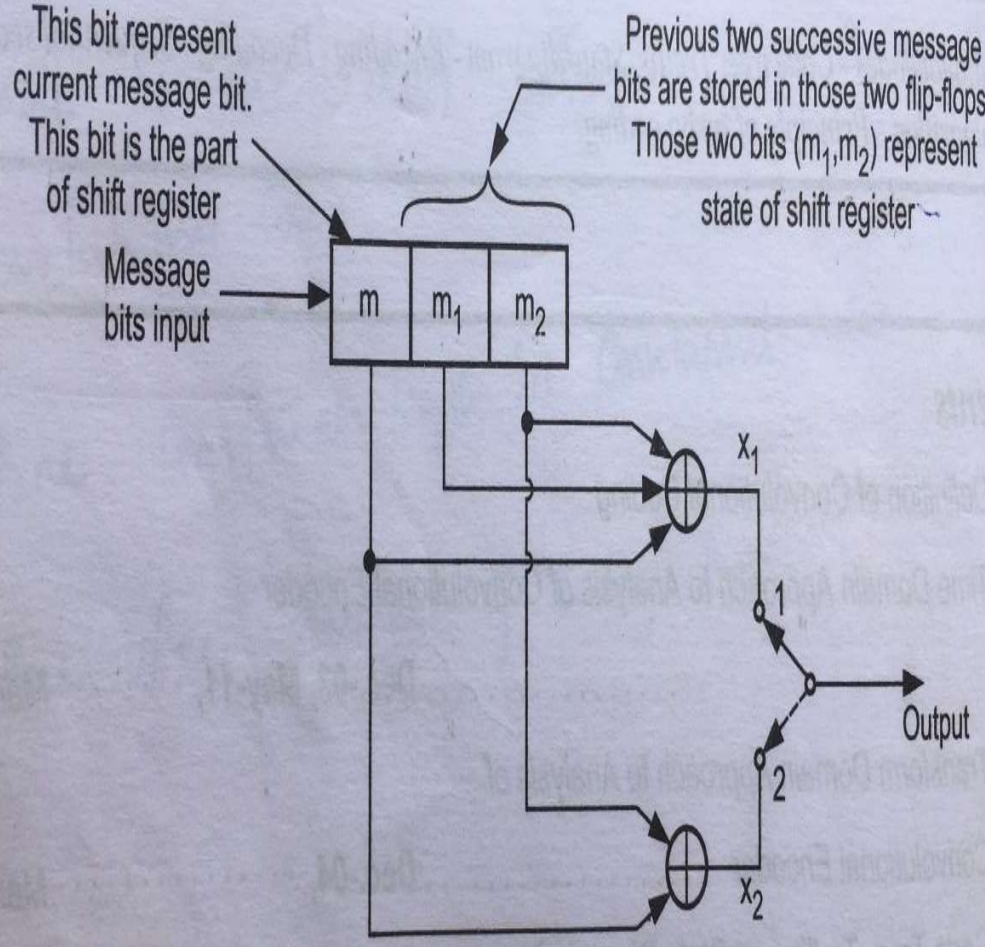


Fig. 5.1.1 Convolutional encoder with K = 3, k = 1 and n = 2

The above convolutional encoder operates as follows.

Operation :

... the new values of x_1 and x_2 are

The code rate of this encoder is,

$$r = \frac{k}{n} = \frac{1}{2} \qquad \ldots (5.1.4)$$

In the encoder of Fig. 5.1.1, observe that whenever a particular message bit enters a shift register, it remains in the shift register for three shifts i.e.,

First shift → Message bit is entered in position 'm'.

Second shift → Message bit is shifted in position m_1.

Third shift → Message bit is shifted in position m_2.

And at the fourth shift the message bit is discarded or simply lost by overwriting. We know that x_1 and x_2 are combinations of m, m_1, m_2. Since a single message bit remains in m during first shift, in m_1 during second shift and in m_2 during third shift; it influences output x_1 and x_2 for 'three' successive shifts.

5.1.2 Constraint Length (K)

The constraint length of a convolution code is defined as the number of shifts over which a single message bit can influence the encoder output. It is expressed in terms of message bits.

For the encoder of Fig. 5.1.1 constraint length K = 3 bits. This is because in the encoder, a single message bit influences encoder output for three successive shifts. At the fourth shift, the message bit is lost and it has no effect on the output.

5.1.3 Dimension of the Code

The dimension of the code is given by n and k. We know that 'k' is the number of message bits taken at a time by the encoder. And 'n' is the encoded output bits for one message bits. Hence the dimension of the code is (n, k). And such encoder is ... Fig. 5.1.1 has the dimensi...

sequence x_1 is given as, ... given. The

$$x_1 = x_i^{(1)} = \sum_{l=0}^{M} g_l^{(1)} m_{i-l} \qquad i = 0, 1, 2 \ldots\ldots \qquad \ldots (5.2.1)$$

Here $m_{i-l} = 0$ for all $l > i$. Similarly the sequence x_2 is given as,

$$x_2 = x_i^{(2)} = \sum_{l=0}^{M} g_l^{(2)} m_{i-l} \quad i = 0, 1, 2 \ldots\ldots \qquad \ldots (5.2.2)$$

Note : All additions in above equations are as per mod-2 addition rules.

As shown in the Fig. 5.1.1, the two sequences x_1 and x_2 are multiplexed by the switch. Hence the output sequence is given as,

$$\{x_i\} = \left\{ x_0^{(1)} \; x_0^{(2)} \; x_1^{(1)} \; x_1^{(2)} \; x_2^{(1)} \; x_2^{(2)} \; x_3^{(1)} \; x_3^{(2)} \ldots\ldots \right\} \qquad \ldots (5.2.3)$$

Here $v_1 = x_i^{(1)} = \left\{ x_0^{(1)} \; x_1^{(1)} \; x_2^{(1)} \; x_3^{(1)} \ldots\ldots \right\}$

and $v_2 = x_i^{(2)} = \left\{ x_0^{(2)} \; x_1^{(2)} \; x_2^{(2)} \; x_3^{(2)} \ldots\ldots \right\}$

Observe that bits from above two sequences are multiplexed in equation (5.2.3) The sequence $\{x_i\}$ is the output of the convolutional encoder.

* Some authors define constraint length as number of output bits influenced by a single message bit i.e.

$$\text{Constraint length } (k) = (n \times M) \text{ bits} \qquad \ldots (5.2.4)$$

MCQ Test

1. Self information should be

a) Negative
b) Positive
c) Positive & Negative
d) None of the mentioned Ans: b

2. The unit of average mutual information is
   a) Bytes per symbol
   b) Bytes
   c) Bits per symbol
   d) Bits Ans: d

3. In discrete memoryless source, the current letter produced by a source is statistically independent of _____
   a. Past output
   b. Future output
   c. Both a and b
   d. None of the aboveAns: c
4. When the base of the logarithm is 2, then the unit of measure of information is
   a) Bits
   b) Bytes
   c) Nats
   d) None of the mentioned Ans: a
5. The self information of random variable is
   a) 0
   b) 1
   c) Infinite
   d) Cannot be determined Ans: c
6. Entropy of a random variable is
   a) 0
   b) 1
   c) Infinite
   d) Cannot be determined Ans: c

7. Which is more efficient method?
   a) Encoding each symbol of a block
   b) Encoding block of symbols

c) Encoding each symbol of a block & Encoding block of symbols
d) None of the mentioned Ans: b

8. The mutual information between a pair of events is
   a) Positive
   b) Negative
   c) Zero
   d) All of the mentionedAns: d
9. When the base of the logarithm is e, the unit of measure of information is
   a) Bits
   b) Bytes
   c) Nats
   d) None of the mentionedAns: c

10. When probability of error during transmission is 0.5, it indicates that
    a) Channel is very noisy
    b) No information is received
    c) Channel is very noisy & No information is received
    d) None of the mentioned Ans: c

11. Types of compression

    a. Lossless
    b. Lossy.
    c. both a and b
    d. None of the aboveAns: C

12. What is significance of D- frames in video coding

    a. They generate low resolution picture.
    b. highly compressed technique.
    c. They generate high resolution picture.
    d. None of the aboveAns: A

13. MPEG coders are used for

    a. compression of audio
    b. compression of text
    c. compression of audio and video
    d. None of the aboveAns: A

14. B-frame is also known as

a. unidirectional.
b. B- De-compression technique
c. B- compression technique
d. bidirectional frame. Ans: D

15. I-frame is

a. It basically searches the frames.
b. It basically predicts the movement of objects.
c. It basically compress the movement of objects
d. None of the above.Ans: B

74. Video Coding consists of two process

a. Processing for reducing Temporal Redundancy.
b. Processing for reducing Spatial Redundancy
c. Both a and b
d. None of the above.Ans: C

16. H.261 is

a. compression of audio
b. De-compression of audio
c. Video Compression standard
d. None of the aboveAns: A

Assignment

(1)Explain the terms (i) Self information (ii) Average information (iii) Mutual Information.
(2)Discuss the reason for using logarithmic measure for measuring the amount of information.
(3)Explain the concept of amount of information associated with message.

(4) A binary source emitting an independent sequence of 0's and 1's with probabilities p and (1p) respectively. Plottheentropyofthesource.

(5) Explain the concept of information, average information, information rate and redundancy as referred to information transmission.

(6)Let X represents the outcome of a single roll of a fair dice. What is the entropy of X?

(7)A code is composed of dots and dashes. Assume that the dash is 3 times as long as the dot and has one-third the probability of occurrence. (i) Calculate the information in dot and that in a dash; (ii) Calculate the average information in dot-dash code; and (iii) Assume that a dot lasts for 10 ms and this same time interval is allowed between symbols. Calculate the average rate of information transmission.
(8)What do you understand by the term extension of a discrete memory less source? Show that the entropy of the nth extension of a DMS is n times the entropy of the original source.

(9)A card is drawn from a deck of playing cards. A) You are informed that the card you draw is spade. How much information did you receive in bits? B) How much information did you receive if you are told that the card you drew is an ace? C) How much information did you receive if you are told that the card you drew is an ace of spades? Is the information content of the message "ace of spades" the sum of the information contents of the messages "spade" and "ace"?

(10) The output of an information source consists OF 128 symbols, 16 of which occurs with probability of 1/32 and remaining 112 occur with a probability of 1/224. The source emits 1000 symbols/sec. assuming that the symbols are chosen independently; find the rate of information of the source.