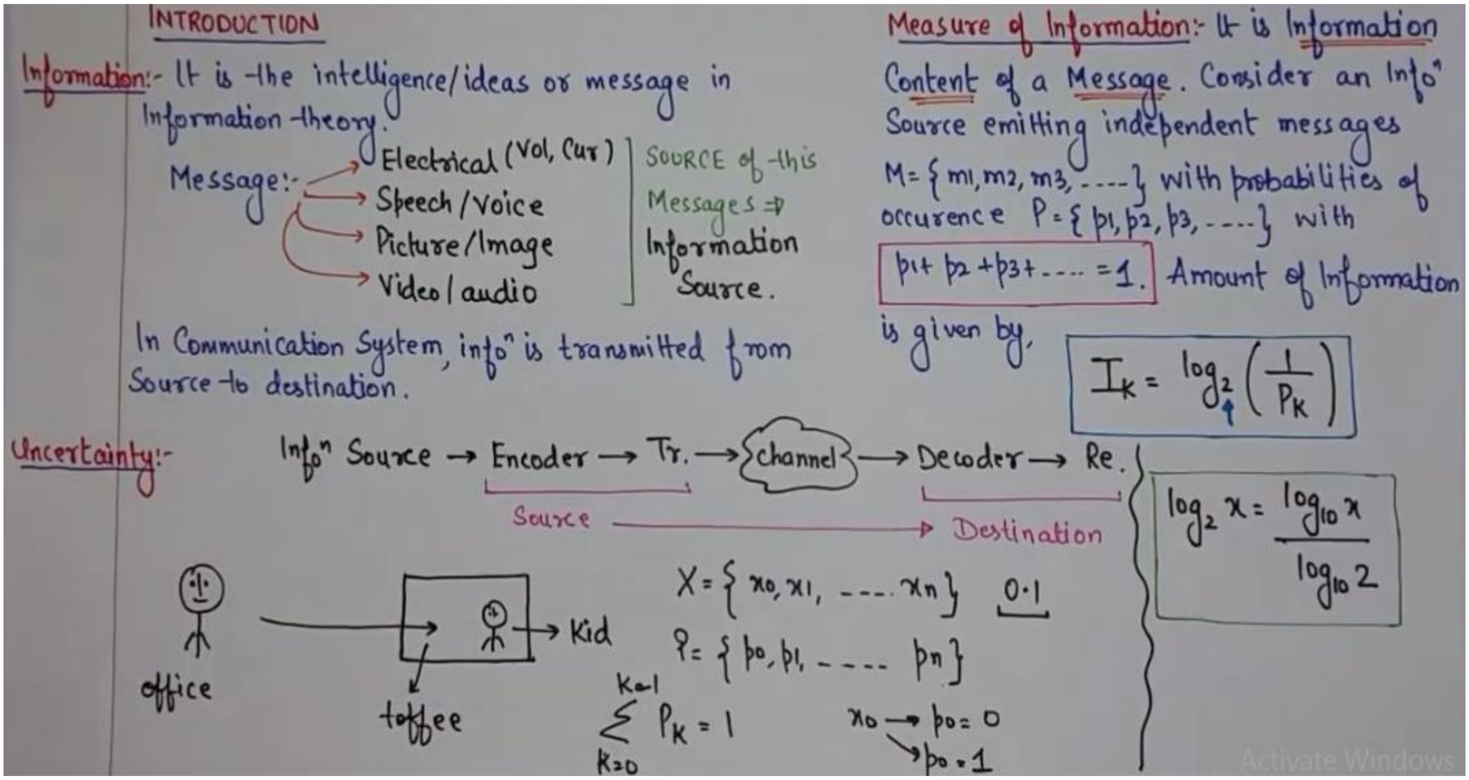
5CS3-01: Information Theory & Coding

UNIT-1 & 2



INTRODUCTION

Information:- It is the intelligence/ideas or message in Information theory.

Message:- → Electrical (Vol, Cur) → Speech / Voice → Picture/Image → Video / audio] SOURCE of this Messages ⇒ Information Source.

In Communication System, info^n is transmitted from Source to destination.

Uncertainty:-

$$\text{Info}^n\text{ Source} \rightarrow \underbrace{\text{Encoder} \rightarrow \text{Tr.}}_{\text{Source}} \rightarrow \text{channel} \rightarrow \underbrace{\text{Decoder} \rightarrow \text{Re.}}_{\text{Destination}}$$

office → Kid → toffee

$X = \{x_0, x_1, \dots x_n\}$ 0·1

$P = \{p_0, p_1, \dots p_n\}$

$$\sum_{K=0}^{K-1} P_K = 1$$

$x_0 \rightarrow p_0 = 0$

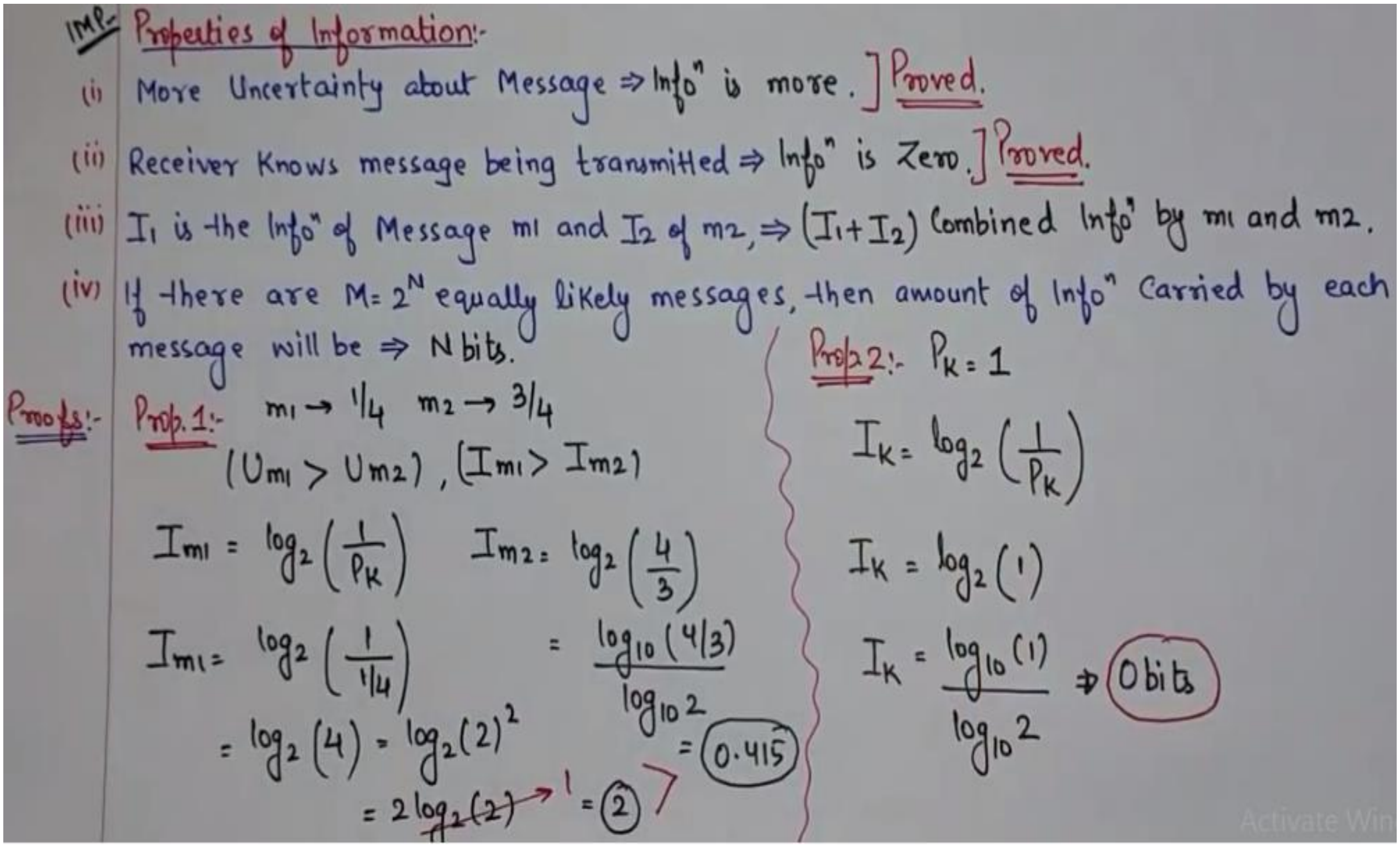
$\searrow p_0 = 1$

Measure of Information:- It is Information Content of a Message. Consider an Info^n Source emitting independent messages $M = \{m_1, m_2, m_3, \dots\}$ with probabilities of occurence $P = \{p_1, p_2, p_3, \dots\}$ with $\boxed{p_1 + p_2 + p_3 + \dots = 1.}$ Amount of Information is given by,

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

$$\boxed{\log_2 x = \frac{\log_{10} x}{\log_{10} 2}}$$

Activate Windows



IMP- **Properties of Information:-**

(i) More Uncertainty about Message ⇒ Info^n is more.] Proved.

(ii) Receiver Knows message being transmitted ⇒ Info^n is Zero.] Proved.

(iii) I_1 is the Info^n of Message m_1 and I_2 of m_2, ⇒ $(I_1 + I_2)$ Combined Info^n by m_1 and m_2.

(iv) If there are $M = 2^N$ equally likely messages, then amount of Info^n Carried by each message will be ⇒ N bits.

Proofs:- **Prob. 1:-** $m_1 \rightarrow 1/4 \quad m_2 \rightarrow 3/4$

$(U_{m1} > U_{m2}), (I_{m1} > I_{m2})$

$$I_{m1} = \log_2\left(\frac{1}{P_K}\right) \qquad I_{m2} = \log_2\left(\frac{4}{3}\right)$$

$$I_{m1} = \log_2\left(\frac{1}{1/4}\right) \qquad = \frac{\log_{10}(4/3)}{\log_{10} 2}$$

$$= \log_2(4) = \log_2(2)^2 \qquad = \boxed{0.415}$$

$$= 2\cancel{\log_2(2)}^{1} = \boxed{2} >$$

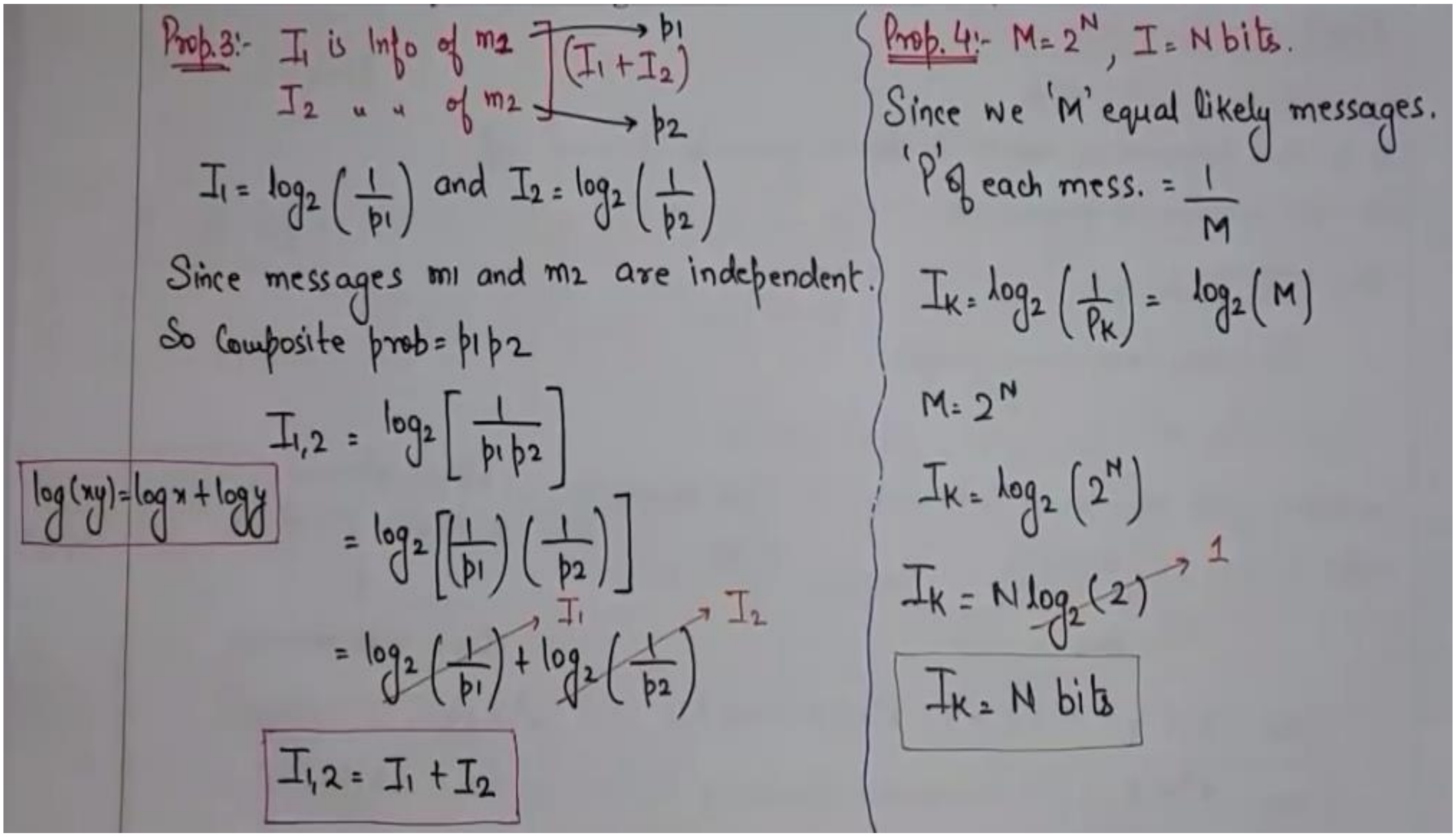
Prob 2:- $P_K = 1$

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

$$I_K = \log_2(1)$$

$$I_K = \frac{\log_{10}(1)}{\log_{10} 2} \Rightarrow \boxed{0\text{ bits}}$$

Activate Win



Prob. 3:- I_1 is Info of m_2 ⎤ $\xrightarrow{\quad} p_1$
I_2 " " of m_2 ⎦ $\xrightarrow{\quad} p_2$
$(I_1 + I_2)$

$I_1 = \log_2\left(\frac{1}{p_1}\right)$ and $I_2 = \log_2\left(\frac{1}{p_2}\right)$

Since messages m_1 and m_2 are independent.
So Composite prob $= p_1 p_2$

$\boxed{\log(xy) = \log x + \log y}$

$$I_{1,2} = \log_2\left[\frac{1}{p_1 p_2}\right]$$

$$= \log_2\left[\left(\frac{1}{p_1}\right)\left(\frac{1}{p_2}\right)\right]$$

$$= \overset{I_1}{\cancel{\log_2\left(\frac{1}{p_1}\right)}} + \overset{I_2}{\cancel{\log_2\left(\frac{1}{p_2}\right)}}$$

$$\boxed{I_{1,2} = I_1 + I_2}$$

Prob. 4:- $M = 2^N$, $I = N$ bits.

Since we 'M' equal likely messages.
'P' of each mess. $= \frac{1}{M}$

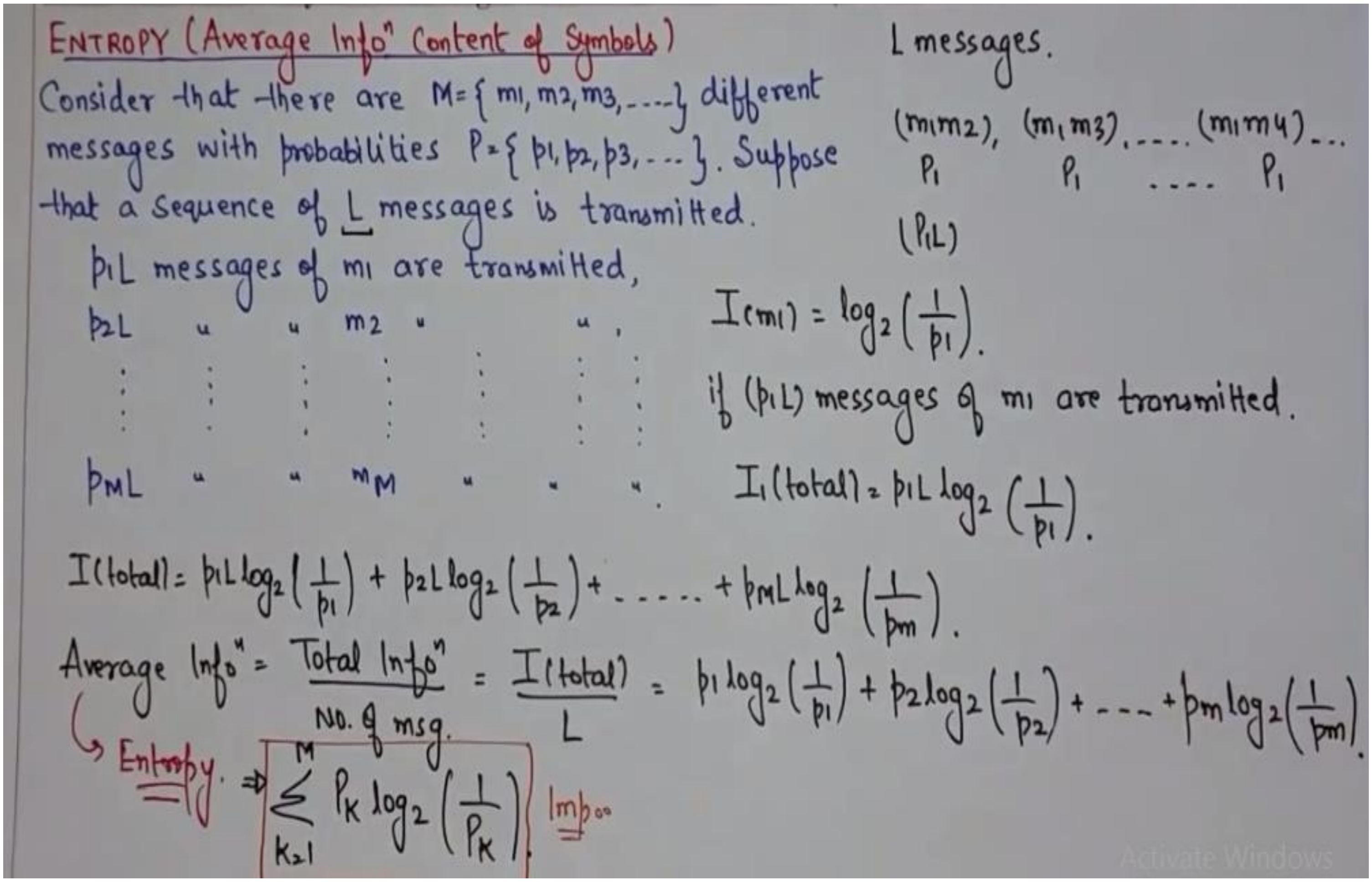
$$I_K = \log_2\left(\frac{1}{P_K}\right) = \log_2(M)$$

$$M = 2^N$$

$$I_K = \log_2\left(2^N\right)$$

$$I_K = N \overset{1}{\cancel{\log_2(2)}}$$

$$\boxed{I_K = N \text{ bits}}$$



ENTROPY (Average Info^n Content of Symbols)

Consider that there are $M = \{m_1, m_2, m_3, \ldots\}$ different messages with probabilities $P = \{p_1, p_2, p_3, \ldots\}$. Suppose that a sequence of $\underline{L}$ messages is transmitted.

$p_1 L$ messages of m_1 are transmitted,
$p_2 L$ " " m_2 " " ,
⋮
$p_M L$ " " m_M " " " .

L messages.

$(m_1 m_2)$, $(m_1 m_3)$, ... $(m_1 m_4)$ ...
P_1 P_1 .... P_1

$(P_1 L)$

$$I(m_1) = \log_2\left(\frac{1}{p_1}\right).$$

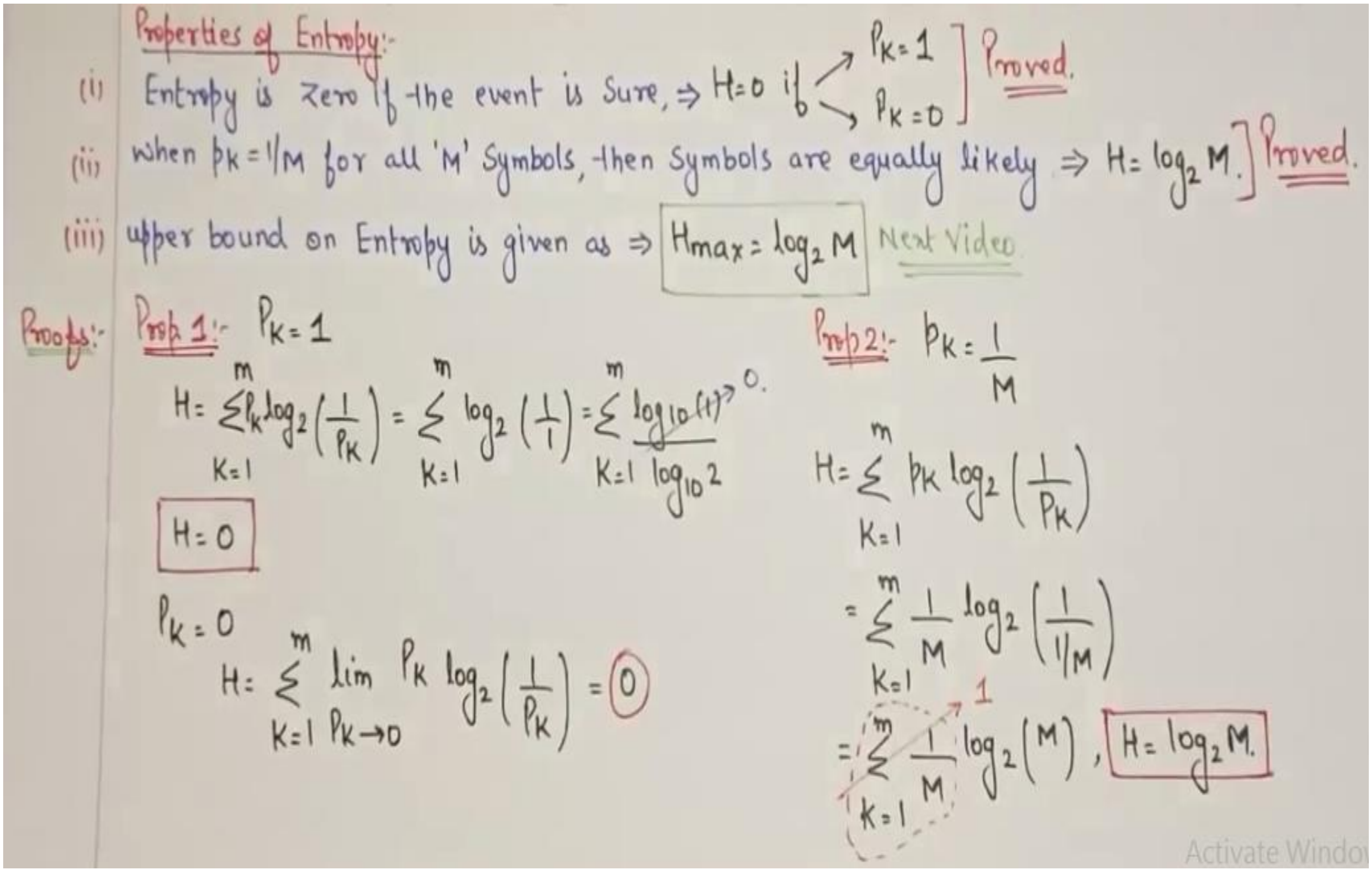
if $(p_1 L)$ messages of m_1 are transmitted.

$$I_1(\text{total}) = p_1 L \log_2\left(\frac{1}{p_1}\right).$$

$$I(\text{total}) = p_1 L \log_2\left(\frac{1}{p_1}\right) + p_2 L \log_2\left(\frac{1}{p_2}\right) + \ldots\ldots + p_M L \log_2\left(\frac{1}{p_m}\right).$$

$$\text{Average Info}^n = \frac{\text{Total Info}^n}{\text{No. of msg.}} = \frac{I(\text{total})}{L} = p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) + \ldots + p_m \log_2\left(\frac{1}{p_m}\right).$$

↳ Entropy. $\Rightarrow \boxed{\sum_{K=1}^{M} P_K \log_2\left(\frac{1}{P_K}\right)}$ Impo



Properties of Entropy:-

(i) Entropy is zero if the event is Sure, $\Rightarrow H=0$ if $\begin{cases} P_K=1 \\ P_K=0 \end{cases}$] Proved.

(ii) when $p_K = 1/M$ for all 'M' Symbols, then Symbols are equally likely $\Rightarrow H=\log_2 M.$] Proved.

(iii) upper bound on Entropy is given as $\Rightarrow$ $\boxed{H_{max}=\log_2 M}$ Next Video

Proofs:- Prop 1:- $P_K=1$

$$H=\sum_{K=1}^{m}P_K\log_2\left(\frac{1}{P_K}\right)=\sum_{K=1}^{m}\log_2\left(\frac{1}{1}\right)=\sum_{K=1}^{m}\frac{\log_{10}(1)\rightarrow 0.}{\log_{10}2}$$

$$\boxed{H=0}$$

$P_K=0$

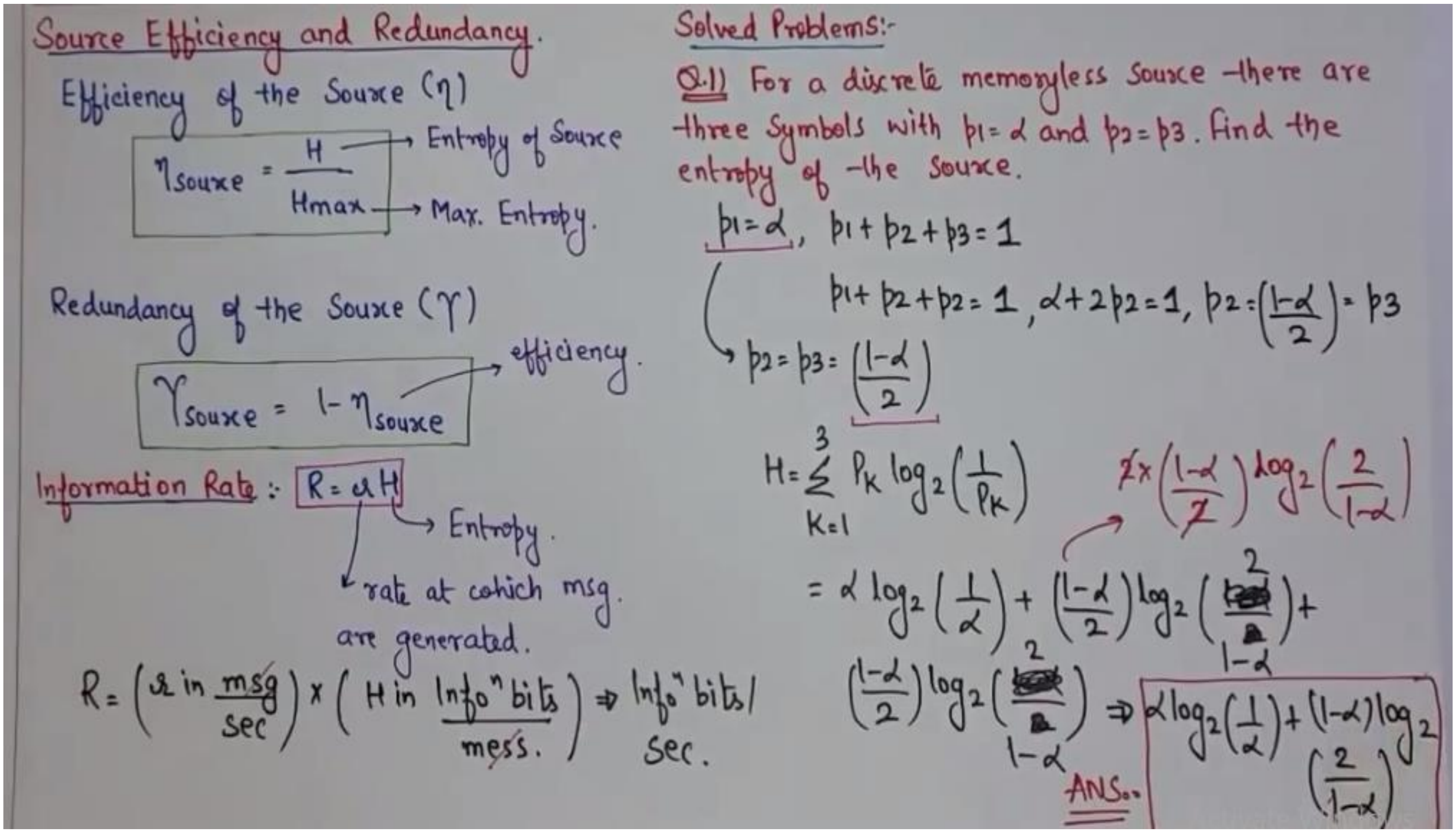
$$H=\sum_{K=1}^{m}\lim_{P_K\to 0}P_K\log_2\left(\frac{1}{P_K}\right)=\textcircled{0}$$

Prop 2:- $p_K=\frac{1}{M}$

$$H=\sum_{K=1}^{m}p_K\log_2\left(\frac{1}{P_K}\right)$$

$$=\sum_{K=1}^{m}\frac{1}{M}\log_2\left(\frac{1}{1/M}\right)$$

$$=\overset{1}{\cancel{\sum_{K=1}^{m}\frac{1}{M}}}\log_2(M),\ \boxed{H=\log_2 M.}$$



Source Efficiency and Redundancy.

Efficiency of the Source (η)

$$\boxed{\eta_{source}=\frac{H}{H_{max}}}$$

H → Entropy of Source

H_{max} → Max. Entropy.

Redundancy of the Source (γ)

$$\boxed{\gamma_{source}=1-\eta_{source}}$$

η_{source} → efficiency.

Information Rate :- $\boxed{R=r H}$

H → Entropy.

r → rate at which msg. are generated.

$$R=\left(r \text{ in } \frac{msg}{sec}\right)\times\left(H \text{ in } \frac{\text{Info}^n\text{ bits}}{\text{mess.}}\right)\Rightarrow \text{Info}^n\text{ bits/sec.}$$

Solved Problems:-

Q.1) For a discrete memoryless source there are three Symbols with $p_1=\alpha$ and $p_2=p_3$. Find the entropy of the source.

$p_1=\alpha$, $p_1+p_2+p_3=1$

$p_1+p_2+p_2=1$, $\alpha+2p_2=1$, $p_2=\left(\frac{1-\alpha}{2}\right)=p_3$

$p_2=p_3=\left(\frac{1-\alpha}{2}\right)$

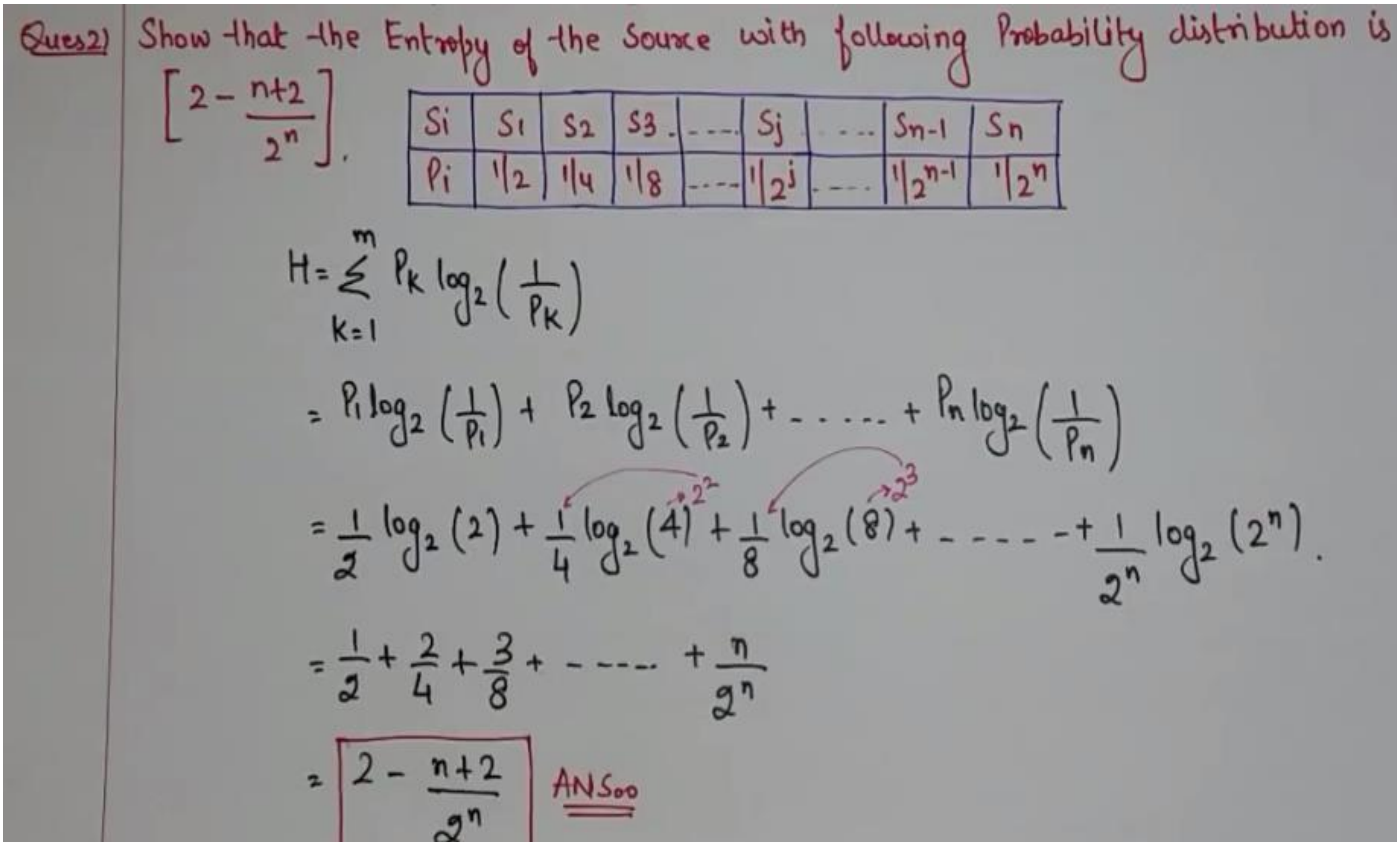
$$H=\sum_{K=1}^{3}P_K\log_2\left(\frac{1}{P_K}\right)$$

$$\cancel{2}\times\left(\frac{1-\alpha}{\cancel{2}}\right)\log_2\left(\frac{2}{1-\alpha}\right)$$

$$=\alpha\log_2\left(\frac{1}{\alpha}\right)+\left(\frac{1-\alpha}{2}\right)\log_2\left(\frac{\overset{2}{[illegible]}}{\underset{1-\alpha}{[illegible]}}\right)+$$

$$\left(\frac{1-\alpha}{2}\right)\log_2\left(\frac{\overset{2}{[illegible]}}{\underset{1-\alpha}{[illegible]}}\right)\Rightarrow\boxed{\alpha\log_2\left(\frac{1}{\alpha}\right)+(1-\alpha)\log_2\left(\frac{2}{1-\alpha}\right)}$$

ANS..



Ques2) Show that the Entropy of the Source with following Probability distribution is $\left[2-\frac{n+2}{2^n}\right]$.

Si	S1	S2	S3	----	Sj	----	Sn-1	Sn
Pi	1/2	1/4	1/8	----	$1/2^j$	----	$1/2^{n-1}$	$1/2^n$

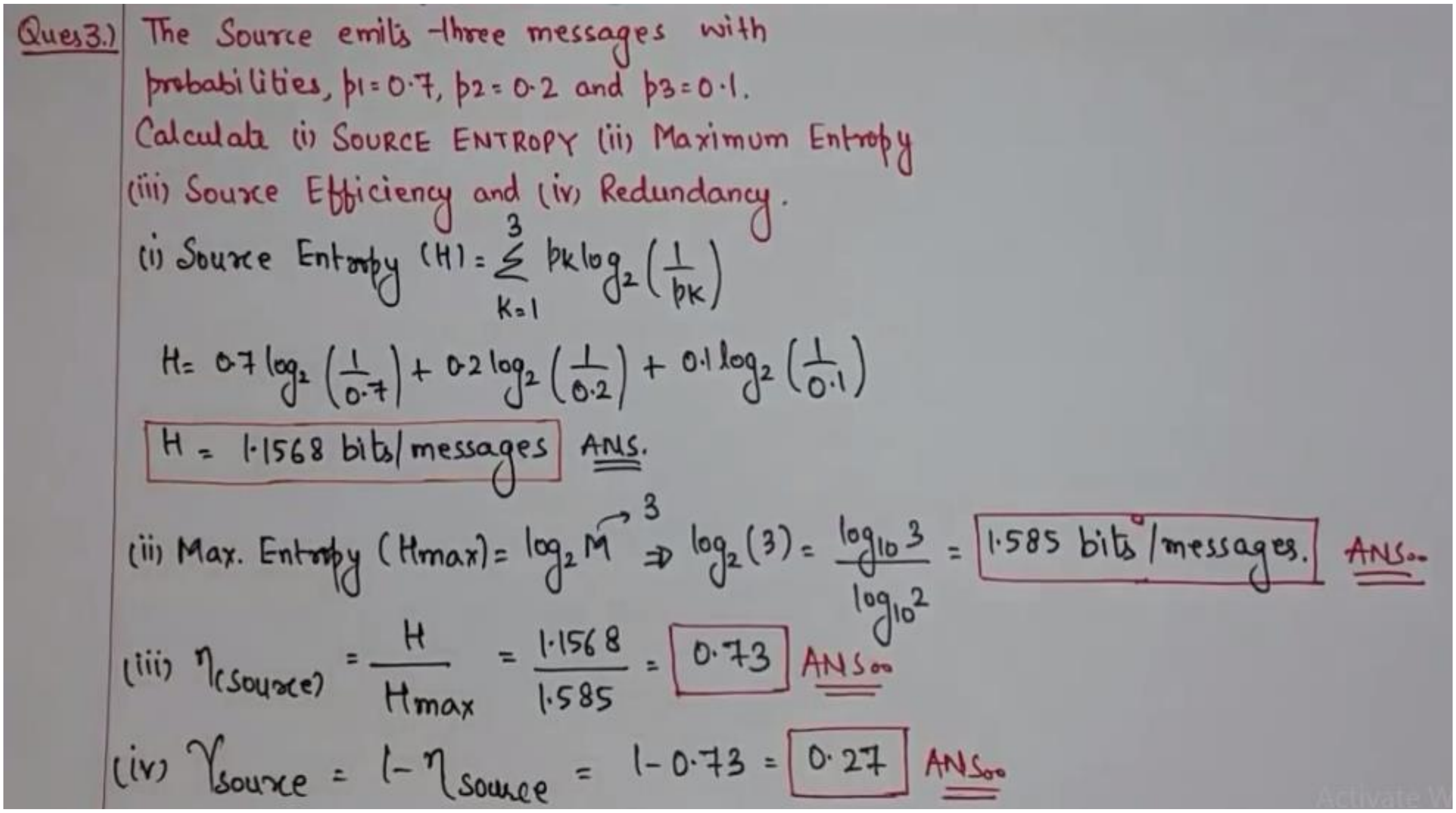
$$H=\sum_{K=1}^{m} P_K \log_2\left(\frac{1}{P_K}\right)$$

$$= P_1\log_2\left(\frac{1}{P_1}\right) + P_2\log_2\left(\frac{1}{P_2}\right) + \text{-----} + P_n\log_2\left(\frac{1}{P_n}\right)$$

$$= \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(\overset{2^2}{4}) + \frac{1}{8}\log_2(\overset{2^3}{8}) + \text{-----} + \frac{1}{2^n}\log_2(2^n).$$

$$= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \text{-----} + \frac{n}{2^n}$$

$$= \boxed{2-\frac{n+2}{2^n}}$$ ANS..



Ques3.) The Source emits three messages with probabilities, $p_1 = 0.7$, $p_2 = 0.2$ and $p_3 = 0.1$.
Calculate (i) SOURCE ENTROPY (ii) Maximum Entropy (iii) Source Efficiency and (iv) Redundancy.

(i) Source Entropy $(H) = \sum_{K=1}^{3} p_K \log_2\left(\frac{1}{p_K}\right)$

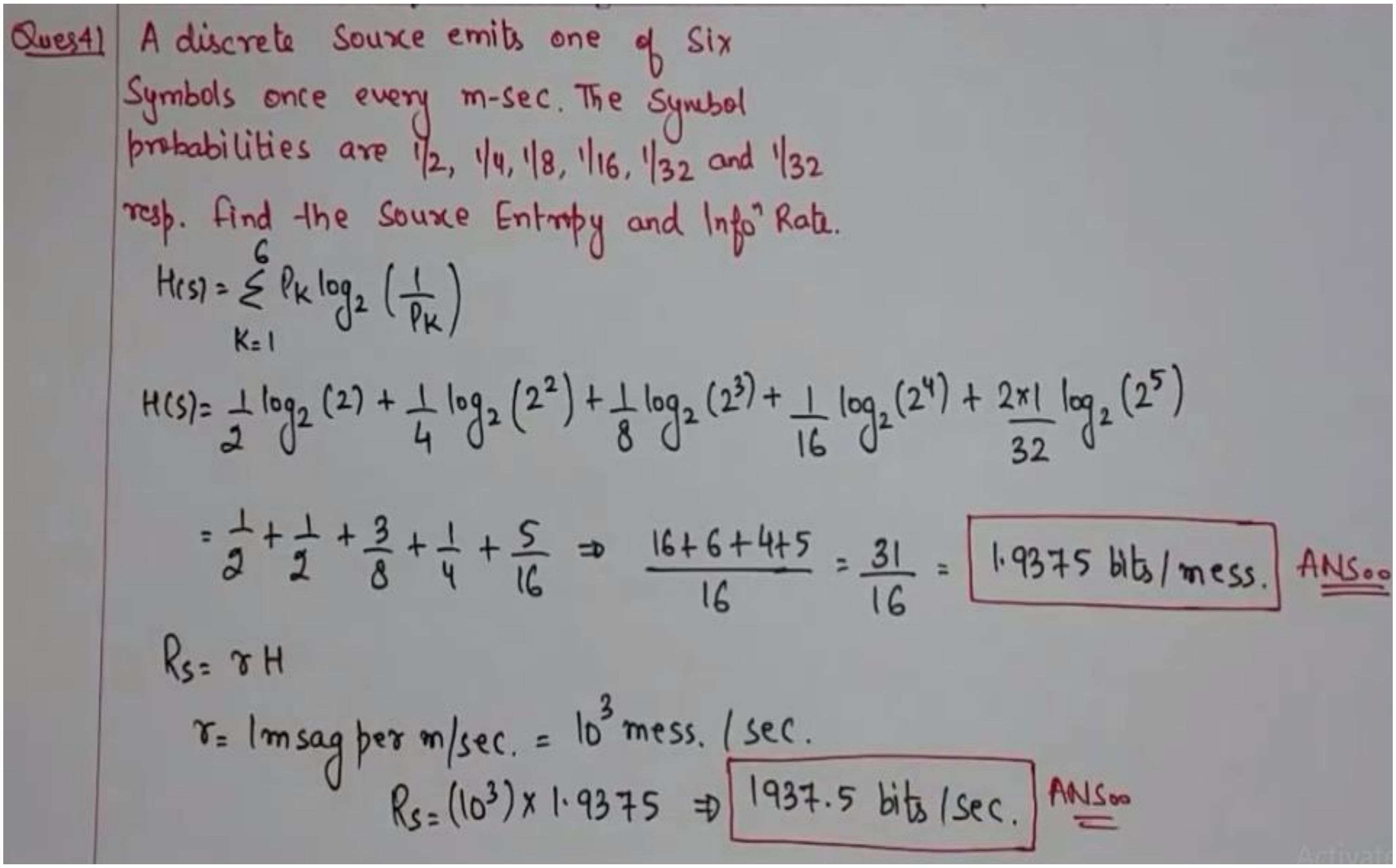
$$H = 0.7\log_2\left(\frac{1}{0.7}\right) + 0.2\log_2\left(\frac{1}{0.2}\right) + 0.1\log_2\left(\frac{1}{0.1}\right)$$

$\boxed{H = 1.1568 \text{ bits/messages}}$ ANS.

(ii) Max. Entropy $(H_{max}) = \log_2 \overset{3}{M} \Rightarrow \log_2(3) = \frac{\log_{10}3}{\log_{10}2} = \boxed{1.585 \text{ bits/messages.}}$ ANS..

(iii) $\eta_{(source)} = \frac{H}{H_{max}} = \frac{1.1568}{1.585} = \boxed{0.73}$ ANS..

(iv) $\gamma_{source} = 1-\eta_{source} = 1-0.73 = \boxed{0.27}$ ANS..



Ques41 A discrete Source emits one of Six Symbols once every m-sec. The Symbol probabilities are 1/2, 1/4, 1/8, 1/16, 1/32 and 1/32 resp. find the Source Entropy and Info^n Rate.

$$H(S) = \sum_{K=1}^{6} P_K \log_2\left(\frac{1}{P_K}\right)$$

$$H(S) = \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(2^2) + \frac{1}{8}\log_2(2^3) + \frac{1}{16}\log_2(2^4) + \frac{2\times1}{32}\log_2(2^5)$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \frac{5}{16} \Rightarrow \frac{16+6+4+5}{16} = \frac{31}{16} = \boxed{1.9375 \text{ bits/mess.}}$$ ANS..

$$R_s = r H$$

r = 1msag per m/sec. = 10^3 mess. / sec.

$$R_s = (10^3) \times 1.9375 \Rightarrow \boxed{1937.5 \text{ bits/sec.}}$$ ANS..

The Source Coding

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. Figure 1.3 shows a source encoder.

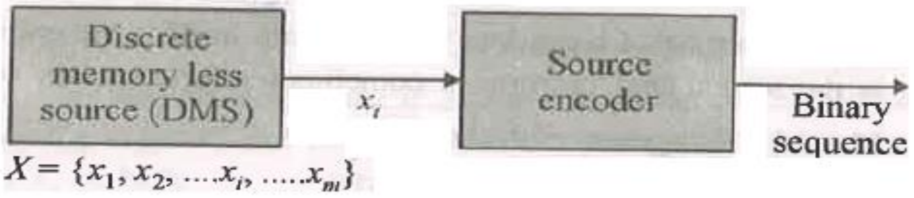


Figure 1.3 Block diagram for source coding

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

In this sub section, let us study the following terms which are related to source coding process:

1. Codeword length
2. Average codeword length
3. Code efficiency
4. Code redundancy

1. Codeword Length

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

2. Average Codeword Length

The average codeword length L, per source symbol is given by

$$L = \sum_{i=1}^{m} p(x_i) n_i$$

The parameter L represents the average number of bits per source symbol used in the source coding process.

3. Code Efficiency

The code efficiency η is defined as under:

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

where $L_{\min}$ is the minimum possible value of L. When η approaches unity, the code is said to be efficient.

4. Code Redundancy

The code redundancy γ is defined as

$$\gamma = 1 - \eta$$

Classification of codes

Classification of codes is best illustrated by an example. Let us consider table 1.1. Where a source of size 4 has encoded in binary codes with symbols 0 and 1.

* Fixed Length Code: → Here symbols are equally probable.
Suppose a DMS outputs a symbol selected from a finite set of symbols x_i, $i = 1, 2, --, L$. The no. of binary digits (bits) R required for unique coding

$$R = \begin{cases} \lceil \log_2 L \rceil & : \text{when } L \text{ is a power of '2'} \\ \lfloor \log_2 L \rfloor + 1 & : \text{when } L \text{ is not power of '2'} \end{cases}$$

eg: To encode 26 alphabets of English. we need

$$R = \lfloor \log_2 26 \rfloor + 1$$

$$= \lfloor 4.6 \rfloor + 1 \Rightarrow 4 + 1 \Rightarrow 5 \text{ bits.}$$

* Variable Length Code: A variable length code is one whose codeword length is not fixed. Here in general we do not have equally probable symbols.

*. Prefix Code: → It is a variable length code.
Set of binary sequence such that no sequence in set 'P' is a prefix of any other binary sequence in P.

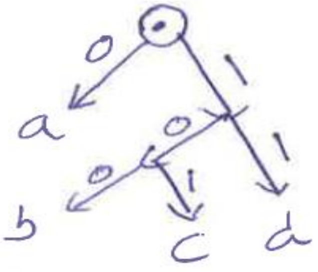
eg. $P = 1. \{01, 010, 110, 0100\}$

Here ~~of~~ '01' comes in 010
& 0100
so it is not prefix code

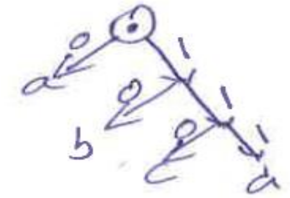
2\. $\{011, 101, 0011\}$

↳ This is prefix code.

eg $a = 0$
$b = 100$
$c = 101$
$d = 11$



eg 0, 10, 110, 111



* Uniquely Decodable Code: A distance code is uniquely decodable if the original source sequence can be reconstructed from the binary sequence.

eg
$a_1 \rightarrow 0$
$a_2 \rightarrow 1$
$a_3 \rightarrow 10$
$a_4 \rightarrow 11$

$\Rightarrow 10 \rightarrow a_2 a_1$
$\rightarrow a_3$

Here two way of decoding is there so it is not U.D.C.

Note: [All prefix code is U.D.C but all UDC is not prefix code]

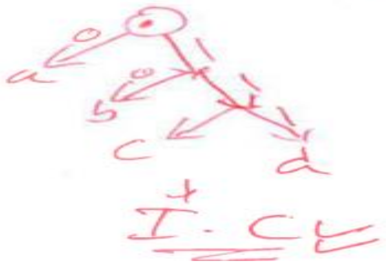
eg.
$a_1 \rightarrow 00$
$a_2 \rightarrow 01$
$a_3 \rightarrow 10$
$a_4 \rightarrow 11$

$\underbrace{00}_{a_1}\underbrace{10}_{a_3}\underbrace{01}_{a_2}\underbrace{11}_{a_4} \rightarrow a_1 a_3 a_2 a_4$

It is Uniquely dec. code

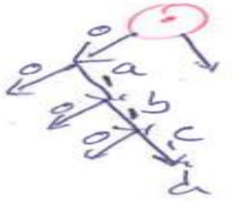
* Instantaneous Code: A uniquely decodable code is called I.C. if the end of any codeword is recognizable without examining subsequent code symbols.
Every prefix code is I.C.

{0, 10, 110, 111}
a b c d



I.C ✓

{0, 01, 011, 0111}



Not I.C.

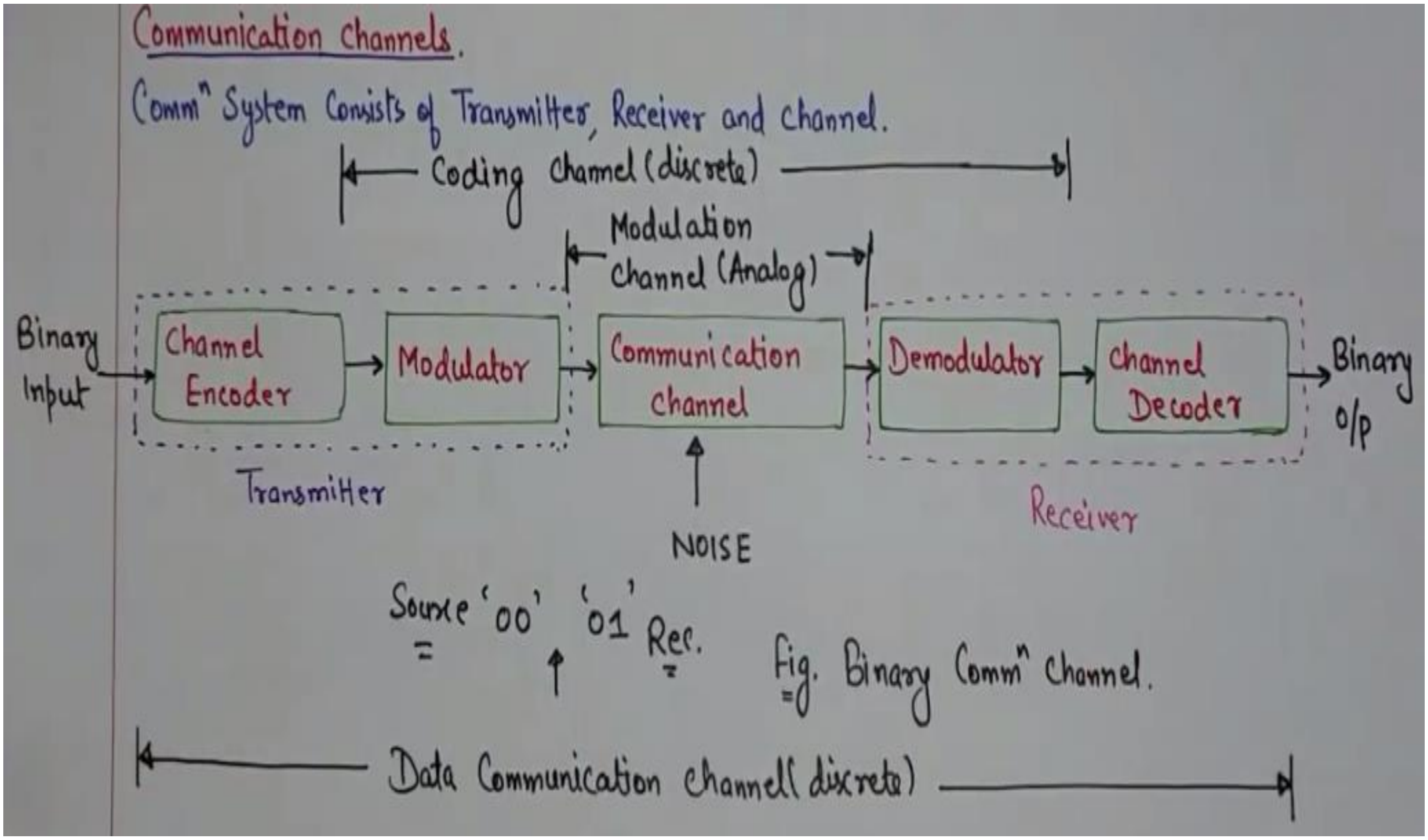
The Kraft Inequality

Let X be a DMS with alphabet $\{x_i\}$ (i = 1,2,...m). Assume that the length of the assigned binary codeward corresponding to x_i is n_i.

A necessary and sufficent condition for the existence of an instantaneous binary code is

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

Which is known as the Kraft inequality.

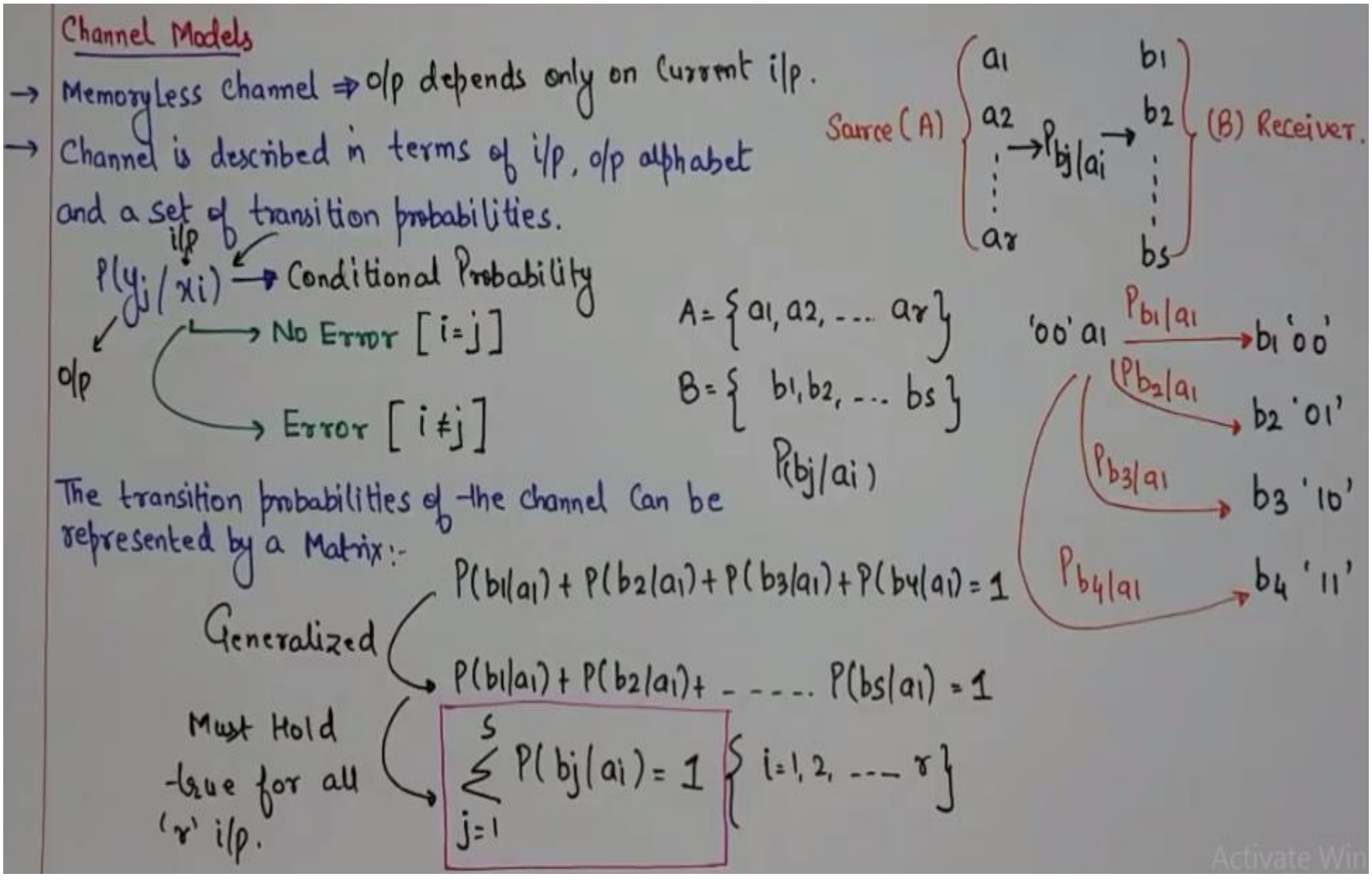


Communication Channels.

Commⁿ System Consists of Transmitter, Receiver and Channel.

Source '00' '01' Rec.

Fig. Binary Commⁿ Channel.



Channel Models

→ Memoryless Channel ⇒ o/p depends only on Current i/p.

→ Channel is described in terms of i/p, o/p alphabet and a set of transition probabilities.

$\underset{o/p}{P(y_j} / \overset{i/p}{x_i)}$ → Conditional Probability

→ No Error $[i=j]$

→ Error $[i \neq j]$

The transition probabilities of the Channel Can be represented by a Matrix:-

Generalized:

$P(b_1|a_1) + P(b_2|a_1) + P(b_3|a_1) + P(b_4|a_1) = 1$

$P(b_1|a_1) + P(b_2|a_1) + \cdots\cdots P(b_s|a_1) = 1$

Must Hold true for all 'r' i/p.

$$\boxed{\sum_{j=1}^{s} P(b_j|a_i) = 1} \quad \{i = 1, 2, \cdots r\}$$

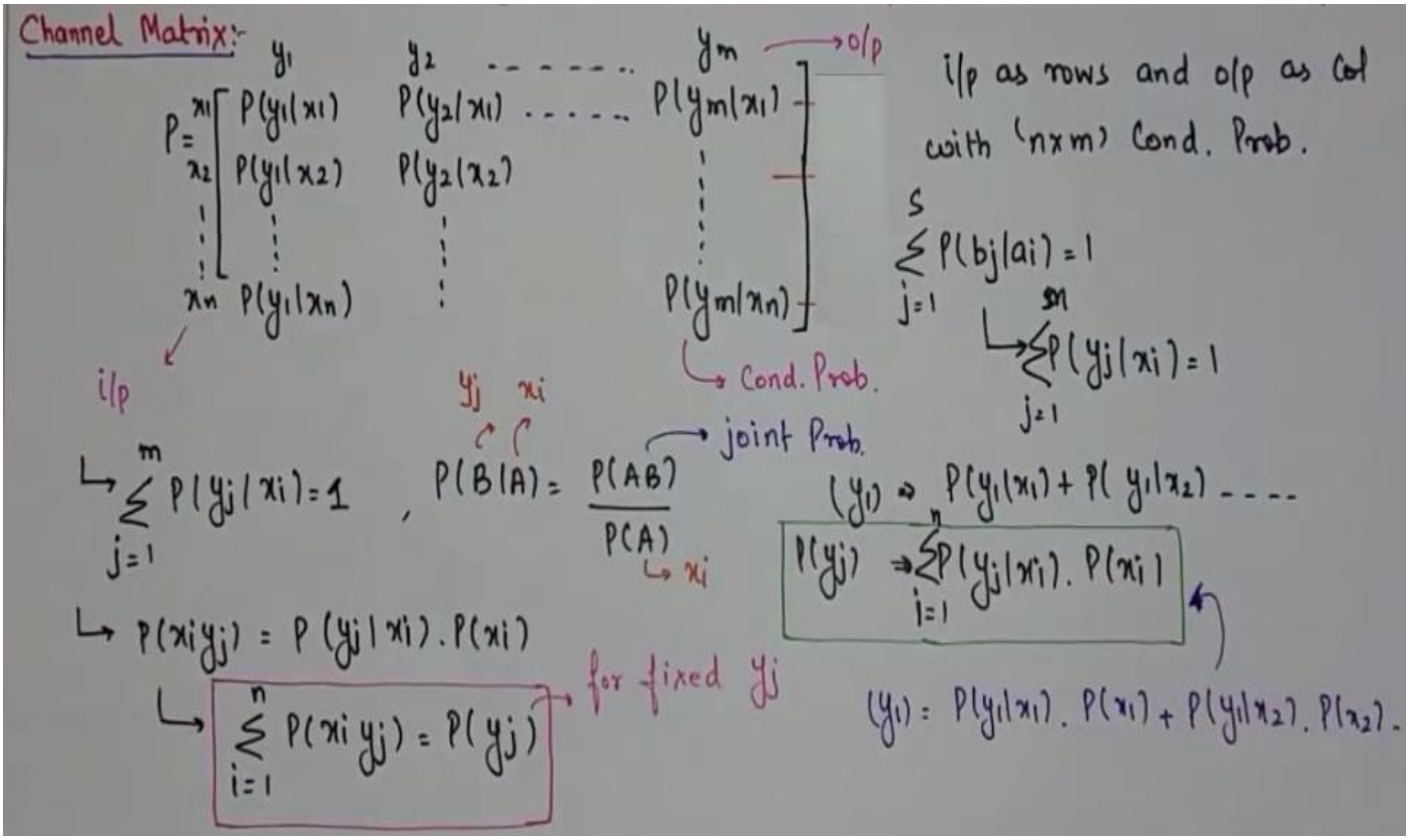
Source (A) $\begin{Bmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{Bmatrix} \rightarrow P_{b_j|a_i} \rightarrow \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{Bmatrix}$ (B) Receiver.

$A = \{a_1, a_2, \dots a_r\}$

$B = \{b_1, b_2, \dots b_s\}$

$P(b_j/a_i)$

$$\begin{tikzcd} \text{'00'}\ a_1 \arrow[r, "P_{b_1|a_1}"] \arrow[dr, bend right, "P_{b_2|a_1}"] \arrow[ddr, bend right, "P_{b_3|a_1}"] \arrow[dddr, bend right, "P_{b_4|a_1}"] & b_1\ \text{'00'} \\ & b_2\ \text{'01'} \\ & b_3\ \text{'10'} \\ & b_4\ \text{'11'} \end{tikzcd}$$



Channel Matrix:-

$$P=\begin{array}{c|cccc} & y_1 & y_2 & \cdots & y_m \\ \hline x_1 & P(y_1|x_1) & P(y_2|x_1) & \cdots & P(y_m|x_1) \\ x_2 & P(y_1|x_2) & P(y_2|x_2) & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ x_n & P(y_1|x_n) & \vdots & & P(y_m|x_n) \end{array} \xrightarrow{} \text{o/p}$$

i/p

Cond. Prob.

i/p as rows and o/p as col with (n×m) Cond. Prob.

$$\sum_{j=1}^{s} P(b_j|a_i)=1$$

$$\hookrightarrow \sum_{j=1}^{m} P(y_j|x_i)=1$$

$$\hookrightarrow \sum_{j=1}^{m} P(y_j|x_i)=1 \quad , \quad P(\overset{y_j}{B}|\overset{x_i}{A})=\frac{\overset{\text{joint Prob.}}{P(AB)}}{\underset{\hookrightarrow x_i}{P(A)}}$$

$$\hookrightarrow P(x_iy_j)=P(y_j|x_i).P(x_i)$$

$$\hookrightarrow \boxed{\sum_{i=1}^{n} P(x_i y_j)=P(y_j)} \rightarrow \text{for fixed } y_j$$

$$(y_1) \Rightarrow P(y_1|x_1)+P(y_1|x_2)----$$

$$\boxed{P(y_j) \Rightarrow \sum_{i=1}^{n} P(y_j|x_i).P(x_i)}$$

$$(y_1): P(y_1|x_1).P(x_1)+P(y_1|x_2).P(x_2).$$

Lossless Channel

A channel described by a channel matrix with only one non zero element in each column is called lossless channel. An example of a lossless channel has been shown in figure 1.6, and the corresponding channel matrix is given in equation (1.61) as under.

$$(P(Y|X))=\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad .....(1.61)$$

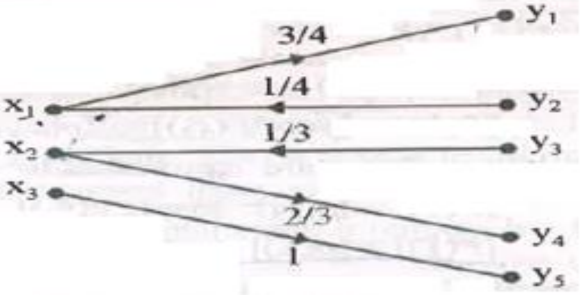


Figure 1.6 Lossless channel

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

Deterministic Channel

$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (1.62)$$

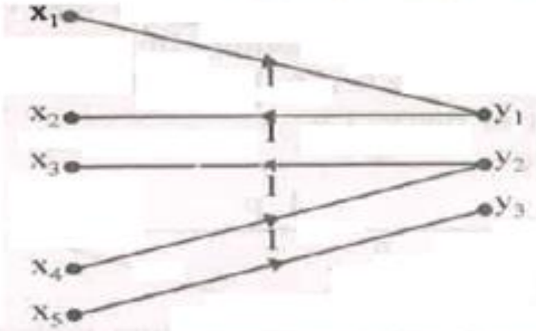


Figure 1.7 Deterministic Channel

Note: It may be noted that since each row has only one non zero element, therefore, this element must be unity. Thus, when a given source symbol is sent in the deterministic channel, it is clear which output symbol will be received.

Noiseless Channel

A channel is called noiseless if it is both lossless and deterministic. A noise less channel has been shown in figure 1.8. The channel matrix has any 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.

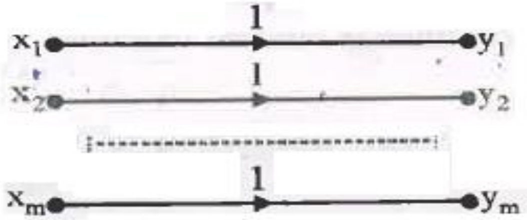
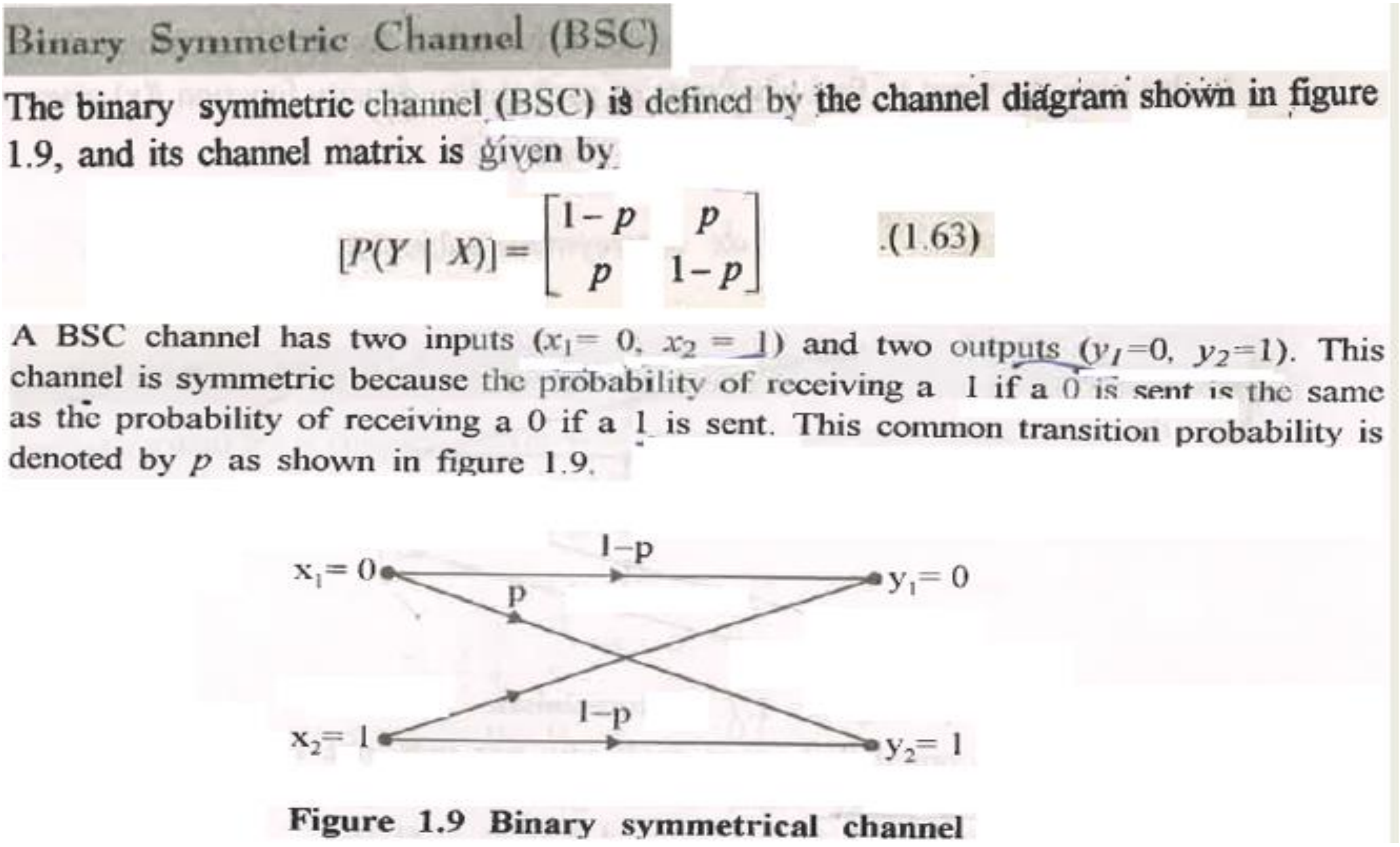


Figure 1.8 Noiseless channel

The matrix for a noiseless channel is given by

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



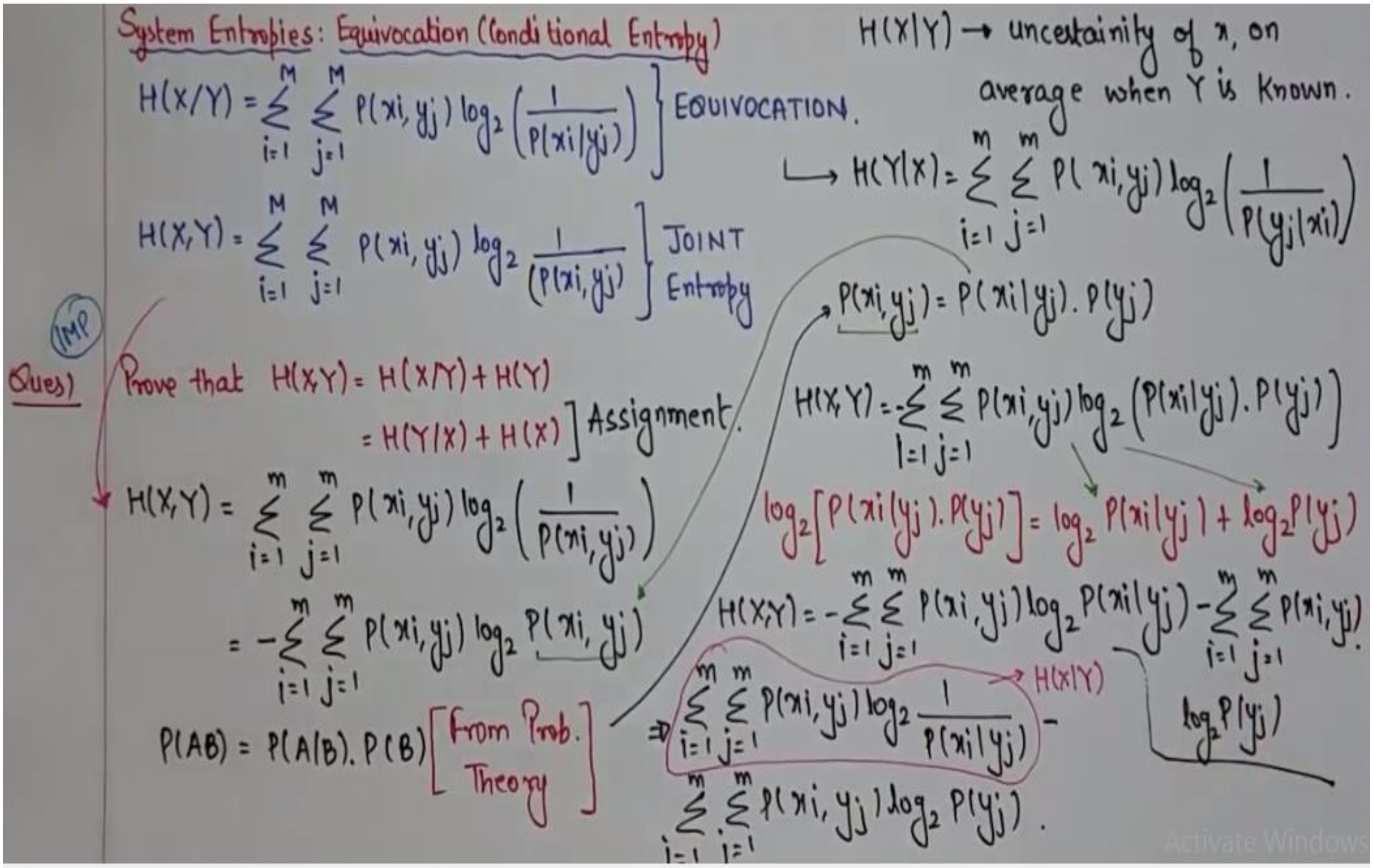
Binary Symmetric Channel (BSC)

The binary symmetric channel (BSC) is defined by the channel diagram shown in figure 1.9, and its channel matrix is given by

$$[P(Y \mid X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \qquad (1.63)$$

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 1.9.

Figure 1.9 Binary symmetrical channel



System Entropies: Equivocation (Conditional Entropy)

$$H(X/Y) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(xi, yj) \log_2 \left(\frac{1}{P(xi/yj)}\right) \Big] \text{EQUIVOCATION.}$$

$$H(X,Y) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(xi, yj) \log_2 \frac{1}{(P(xi, yj)} \Big] \text{JOINT Entropy}$$

$H(X|Y) \rightarrow$ uncertainity of x, on average when Y is Known.

$$\hookrightarrow H(Y|X) = \sum_{i=1}^{m} \sum_{j=1}^{m} P(xi, yj) \log_2 \left(\frac{1}{P(yj|xi)}\right)$$

(IMP)

Ques) Prove that $H(X,Y) = H(X/Y) + H(Y)$

$= H(Y/X) + H(X) \Big]$ Assignment.

$$H(X,Y) = \sum_{i=1}^{m} \sum_{j=1}^{m} P(xi, yj) \log_2 \left(\frac{1}{P(xi, yj)}\right)$$

$$= -\sum_{i=1}^{m} \sum_{j=1}^{m} P(xi, yj) \log_2 \underline{P(xi, yj)}$$

$P(AB) = P(A/B).P(B)$ [From Prob. Theory]

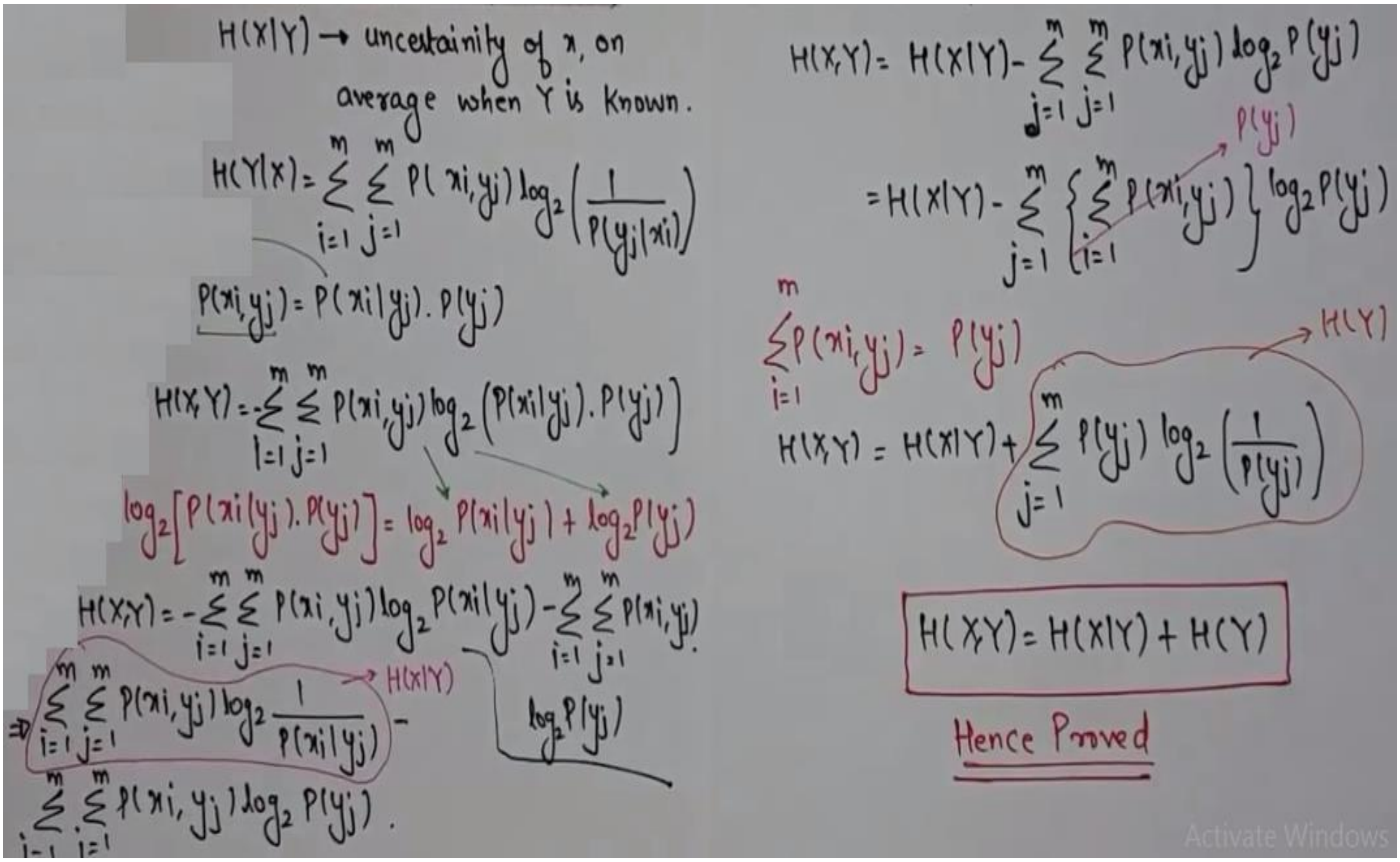
$$\underline{P(xi, yj)} = P(xi|yj).P(yj)$$

$$H(X,Y) = -\sum_{1=1}^{m} \sum_{j=1}^{m} P(xi, yj) \log_2 (P(xi|yj).P(yj))$$

$$\log_2[P(xi|yj).P(yj)] = \log_2 P(xi|yj) + \log_2 P(yj)$$

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

$$\Rightarrow \underbrace{\sum_{i=1}^{m} \sum_{j=1}^{m} P(xi, yj) \log_2 \frac{1}{P(xi|yj)}}_{\rightarrow H(X|Y)} - \sum_{i=1}^{m} \sum_{j=1}^{m} P(xi, yj) \log_2 P(yj).$$



H(X|Y) → uncertainity of x, on average when Y is Known.

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

$$P(x_i,y_j) = P(x_i|y_j).P(y_j)$$

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

$$\log_2\left[P(x_i|y_j).P(y_j)\right] = \log_2 P(x_i|y_j) + \log_2 P(y_j)$$

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

$$\Rightarrow \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i,y_j)\log_2\frac{1}{P(x_i|y_j)}}_{H(X|Y)} - \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i,y_j)\log_2 P(y_j).$$

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

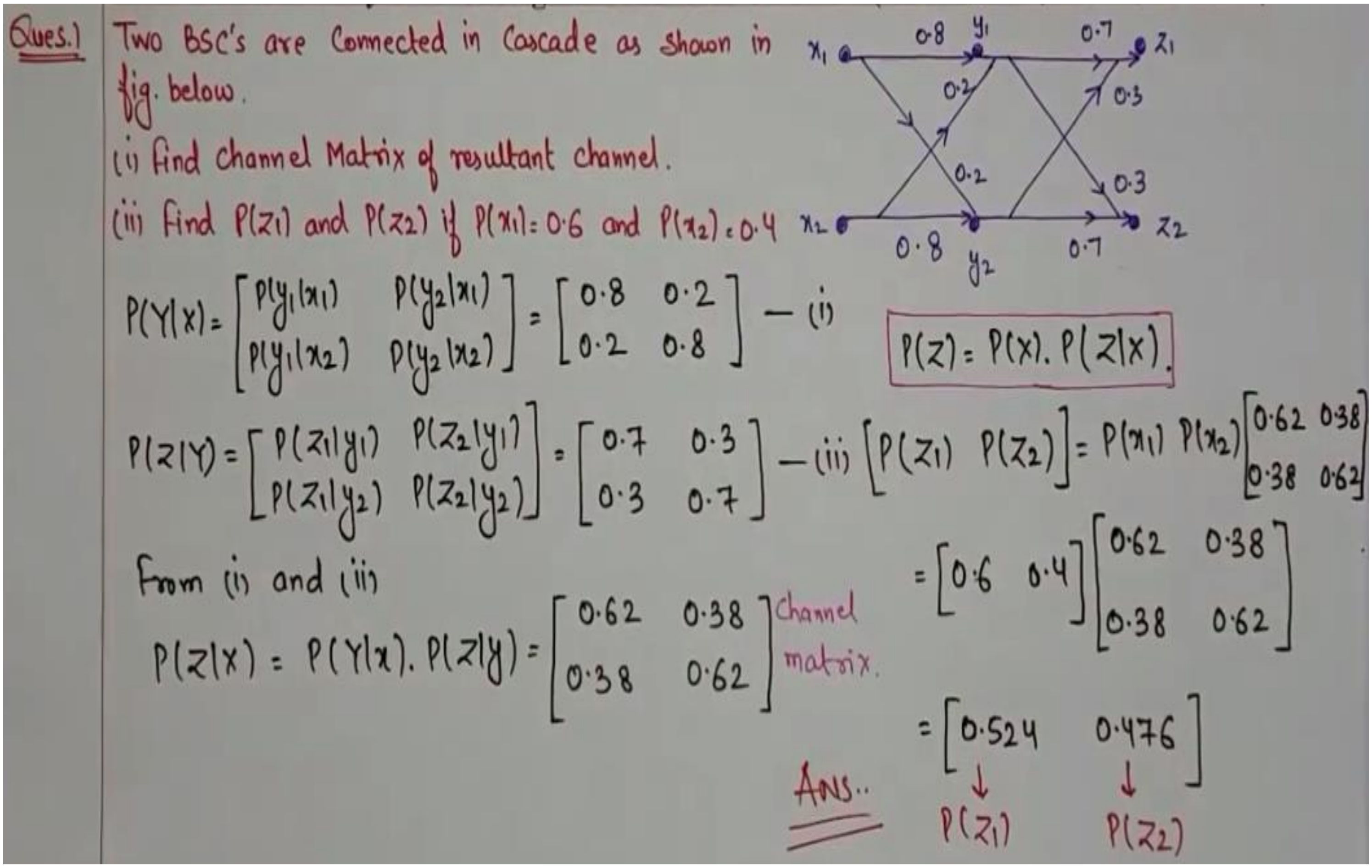
$$= H(X|Y) - \sum_{j=1}^{m}\left\{\overbrace{\sum_{i=1}^{m} P(x_i,y_j)}^{P(y_j)}\right\}\log_2 P(y_j)$$

$$\sum_{i=1}^{m} P(x_i,y_j) = P(y_j)$$

$$H(X,Y) = H(X|Y) + \underbrace{\sum_{j=1}^{m} P(y_j)\log_2\left(\frac{1}{P(y_j)}\right)}_{H(Y)}$$

$$\boxed{H(X,Y) = H(X|Y) + H(Y)}$$

Hence Proved



Ques.) Two BSC's are Connected in Cascade as shown in fig. below.

(i) find Channel Matrix of resultant channel.

(ii) find P(z1) and P(z2) if P(x1) = 0.6 and P(x2) = 0.4

$$P(Y|X) = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) \\ P(y_1|x_2) & P(y_2|x_2) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \quad \text{— (i)}$$

$$P(Z|Y) = \begin{bmatrix} P(z_1|y_1) & P(z_2|y_1) \\ P(z_1|y_2) & P(z_2|y_2) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad \text{— (ii)}$$

From (i) and (ii)

$$P(Z|X) = P(Y|x).P(Z|y) = \begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix} \text{Channel matrix.}$$

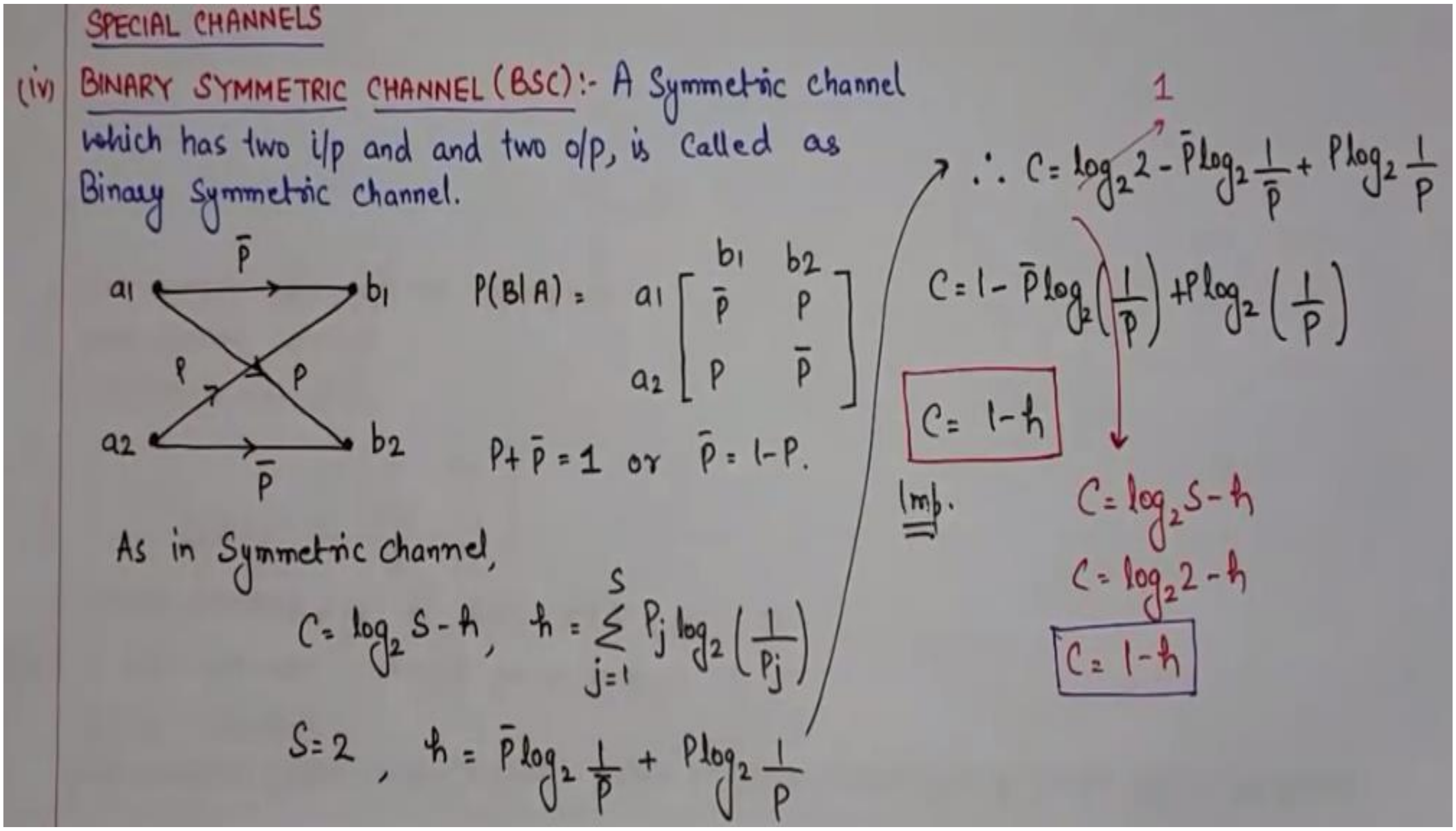
$$\boxed{P(Z) = P(X).P(Z|X)}$$

$$\begin{bmatrix} P(z_1) & P(z_2) \end{bmatrix} = P(x_1)\ P(x_2)\begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix}$$

$$= \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}\begin{bmatrix} 0.62 & 0.38 \\ 0.38 & 0.62 \end{bmatrix}$$

$$= \begin{bmatrix} \underset{P(z_1)}{\underset{\downarrow}{0.524}} & \underset{P(z_2)}{\underset{\downarrow}{0.476}} \end{bmatrix}$$

Ans..



SPECIAL CHANNELS

(iv) **BINARY SYMMETRIC CHANNEL (BSC)**:- A Symmetric channel which has two i/p and and two o/p, is Called as Binary Symmetric channel.

$$P(B|A) = \begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & \bar{P} & P \\ a_2 & P & \bar{P} \end{array}$$

$P + \bar{P} = 1$ or $\bar{P} = 1 - P$.

As in Symmetric channel,

$$C = \log_2 S - h, \quad h = \sum_{j=1}^{S} P_j \log_2\left(\frac{1}{P_j}\right)$$

$$S = 2, \quad h = \bar{P}\log_2\frac{1}{\bar{P}} + P\log_2\frac{1}{P}$$

$$\therefore C = \overset{1}{\cancel{\log_2 2}} - \bar{P}\log_2\frac{1}{\bar{P}} + P\log_2\frac{1}{P}$$

$$C = 1 - \bar{P}\log_2\left(\frac{1}{\bar{P}}\right) + P\log_2\left(\frac{1}{P}\right)$$

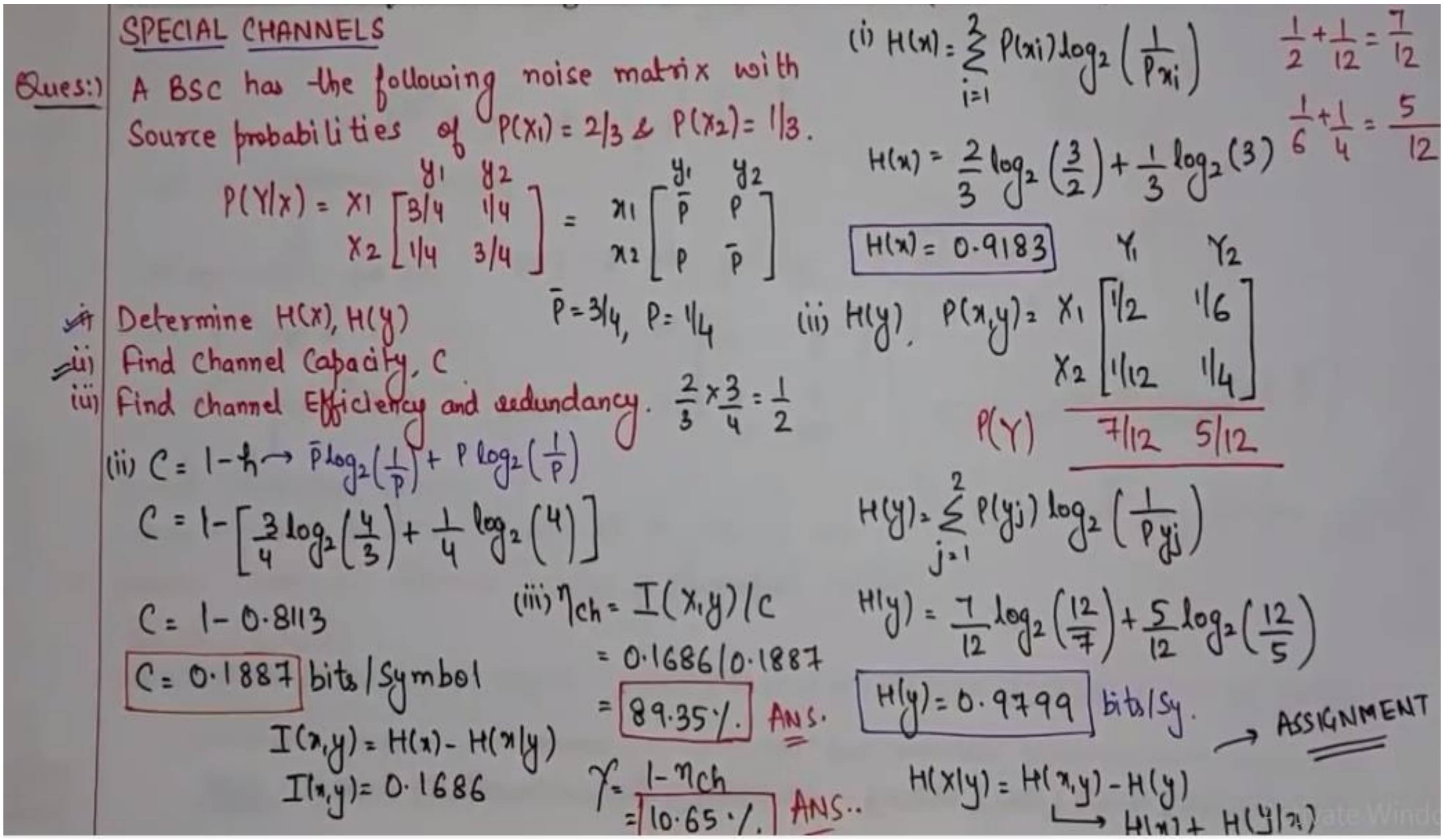
$$\boxed{C = 1 - h}$$

Imp.

$$C = \log_2 S - h$$

$$C = \log_2 2 - h$$

$$\boxed{C = 1 - h}$$



SPECIAL CHANNELS

Ques:) A BSC has the following noise matrix with Source probabilities of $P(X_1) = 2/3$ & $P(X_2) = 1/3$.

$$P(Y|X) = \begin{array}{c|cc} & y_1 & y_2 \\ \hline X_1 & 3/4 & 1/4 \\ X_2 & 1/4 & 3/4 \end{array} = \begin{array}{c|cc} & y_1 & y_2 \\ \hline x_1 & \bar{P} & P \\ x_2 & P & \bar{P} \end{array}$$

$\bar{P} = 3/4, \; P = 1/4$

i) Determine H(X), H(y)

ii) Find channel Capacity, C

iii) Find channel Efficiency and redundancy.

$$\frac{2}{3} \times \frac{3}{4} = \frac{1}{2}$$

(ii) $C = 1 - h \rightarrow \bar{P}\log_2\left(\frac{1}{\bar{P}}\right) + P\log_2\left(\frac{1}{P}\right)$

$$C = 1 - \left[\frac{3}{4}\log_2\left(\frac{4}{3}\right) + \frac{1}{4}\log_2(4)\right]$$

$$C = 1 - 0.8113$$

$\boxed{C = 0.1887}$ bits/Symbol

$$I(x,y) = H(x) - H(x|y)$$

$$I(x,y) = 0.1686$$

(iii) $\eta_{ch} = I(x,y)/C$

$= 0.1686/0.1887$

$= \boxed{89.35\%}$ ANS.

$\gamma = 1 - \eta_{ch}$

$= \boxed{10.65\%}$ ANS..

(i) $H(x) = \sum_{i=1}^{2} P(x_i)\log_2\left(\frac{1}{P_{x_i}}\right)$

$$\frac{1}{2} + \frac{1}{12} = \frac{7}{12}$$

$$\frac{1}{6} + \frac{1}{4} = \frac{5}{12}$$

$$H(x) = \frac{2}{3}\log_2\left(\frac{3}{2}\right) + \frac{1}{3}\log_2(3)$$

$$\boxed{H(x) = 0.9183}$$

(ii) H(y). $P(x,y) = \begin{array}{c|cc} & Y_1 & Y_2 \\ \hline X_1 & 1/2 & 1/6 \\ X_2 & 1/12 & 1/4 \\ \hline P(Y) & 7/12 & 5/12 \end{array}$

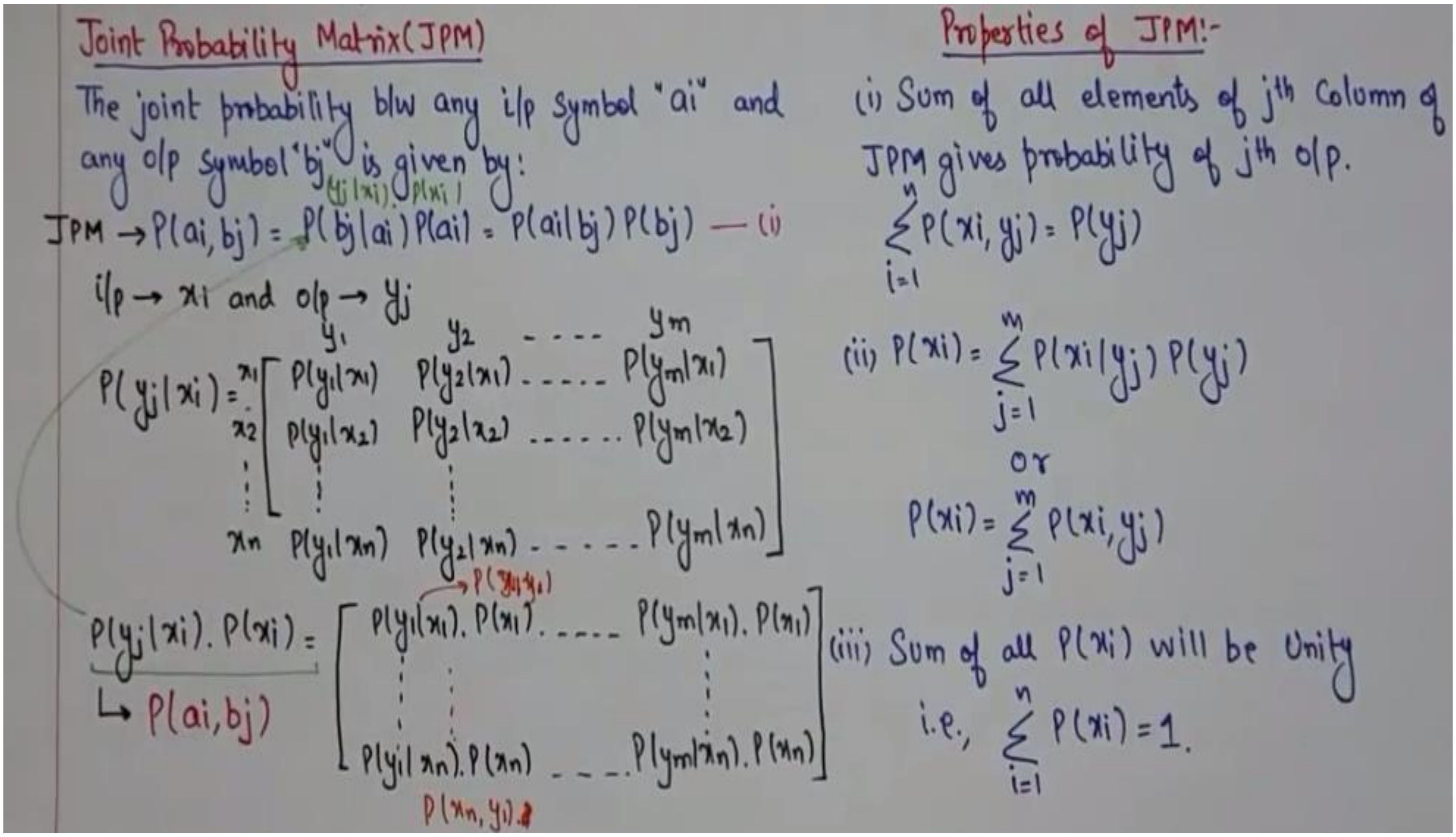
$$H(y) = \sum_{j=1}^{2} P(y_j)\log_2\left(\frac{1}{P_{y_j}}\right)$$

$$H(y) = \frac{7}{12}\log_2\left(\frac{12}{7}\right) + \frac{5}{12}\log_2\left(\frac{12}{5}\right)$$

$\boxed{H(y) = 0.9799}$ bits/Sy.

$H(X|y) = H(x,y) - H(y)$ → ASSIGNMENT

$\hookrightarrow H(x) + H(y/x)$



Joint Probability Matrix(JPM)

The joint probability b/w any i/p Symbol "ai" and any o/p symbol "bj" is given by:

$$\text{JPM} \rightarrow P(ai,bj) = \overset{(yj|xi).}{P}(\overset{p(xi)}{bj|ai})P(ai) = P(ai|bj)P(bj) \quad \text{— (i)}$$

i/p → x_i and o/p → y_j

$$P(y_j|x_i) = \begin{array}{c|cccc} & y_1 & y_2 & \cdots & y_m \\ \hline x_1 & P(y_1|x_1) & P(y_2|x_1) & \cdots & P(y_m|x_1) \\ x_2 & P(y_1|x_2) & P(y_2|x_2) & \cdots & P(y_m|x_2) \\ \vdots & \vdots & \vdots & & \\ x_n & P(y_1|x_n) & P(y_2|x_n) & \cdots & P(y_m|x_n) \end{array}$$

$$\underbrace{P(y_j|x_i).P(x_i)}_{\hookrightarrow P(ai,bj)} = \begin{bmatrix} \overset{\rightarrow P(x_1,y_1)}{P(y_1|x_1).P(x_1)} & \cdots & P(y_m|x_1).P(x_1) \\ \vdots & & \vdots \\ \underset{P(x_n,y_1)}{P(y_1|x_n).P(x_n)} & \cdots & P(y_m|x_n).P(x_n) \end{bmatrix}$$

Properties of JPM:-

(i) Sum of all elements of j^{th} Column of JPM gives probability of j^{th} o/p.

$$\sum_{i=1}^{n} P(x_i,y_j) = P(y_j)$$

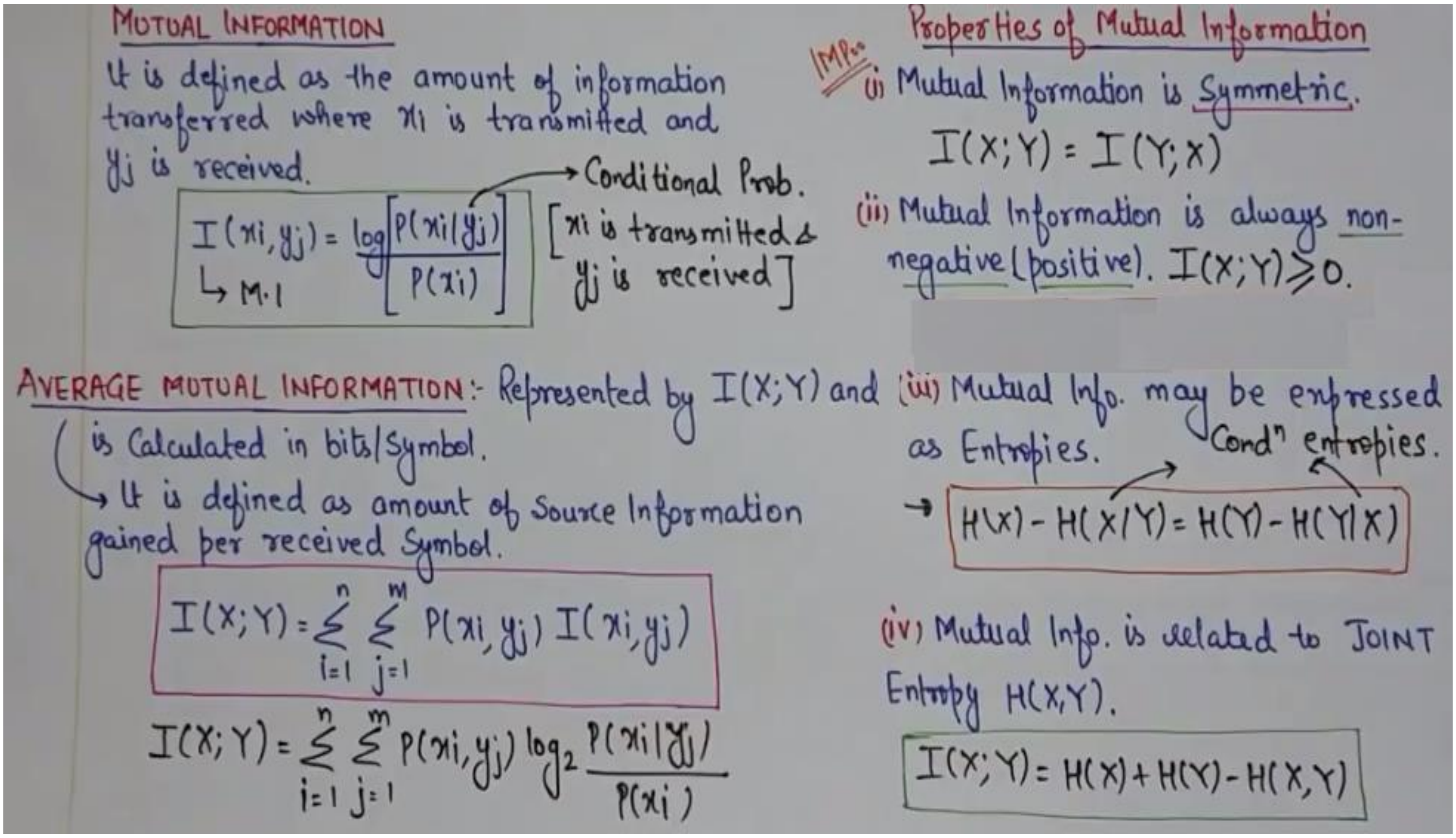
(ii) $$P(x_i) = \sum_{j=1}^{m} P(x_i|y_j)P(y_j)$$

or

$$P(x_i) = \sum_{j=1}^{m} P(x_i,y_j)$$

(iii) Sum of all $P(x_i)$ will be Unity

i.e., $$\sum_{i=1}^{n} P(x_i) = 1.$$



MUTUAL INFORMATION

It is defined as the amount of information transferred where x_i is transmitted and y_j is received.

$$\boxed{\underset{\hookrightarrow \text{M.I}}{I(x_i,y_j)} = \log\left[\frac{P(x_i|y_j)}{P(x_i)}\right]} \quad \left[\begin{array}{l} x_i \text{ is transmitted \&} \\ y_j \text{ is received} \end{array}\right]$$

→ Conditional Prob.

AVERAGE MUTUAL INFORMATION:- Represented by $I(X;Y)$ and is Calculated in bits/Symbol.

→ It is defined as amount of Source Information gained per received Symbol.

$$\boxed{I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\, I(x_i,y_j)}$$

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

Properties of Mutual Information

IMP

(i) Mutual Information is Symmetric.

$$I(X;Y) = I(Y;X)$$

(ii) Mutual Information is always non-negative (positive). $I(X;Y) \geqslant 0$.

(iii) Mutual Info. may be expressed as Entropies.

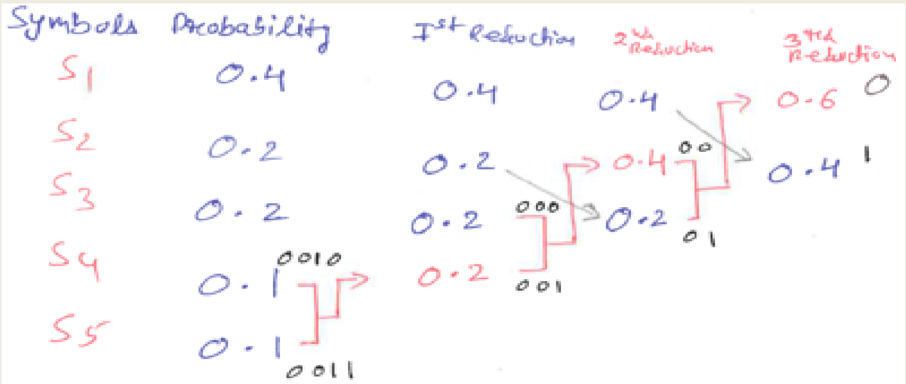
Cond^n entropies.

$$\rightarrow \boxed{H(X) - H(X/Y) = H(Y) - H(Y|X)}$$

(iv) Mutual Info. is related to JOINT Entropy $H(X,Y)$.

$$\boxed{I(X;Y) = H(X) + H(Y) - H(X,Y)}$$

Huffman coding: Example



Symbols	Codewords
S_1	1
S_2	01
S_3	000
S_4	0010
S_5	0011

Now, calculate entropy → $H = \sum_{i=1}^{5} P_i \log_2\left(\frac{1}{P_i}\right)$

Avg. Codeword length = $L_{avg} = \sum_{i=1}^{M} P_i n_i$

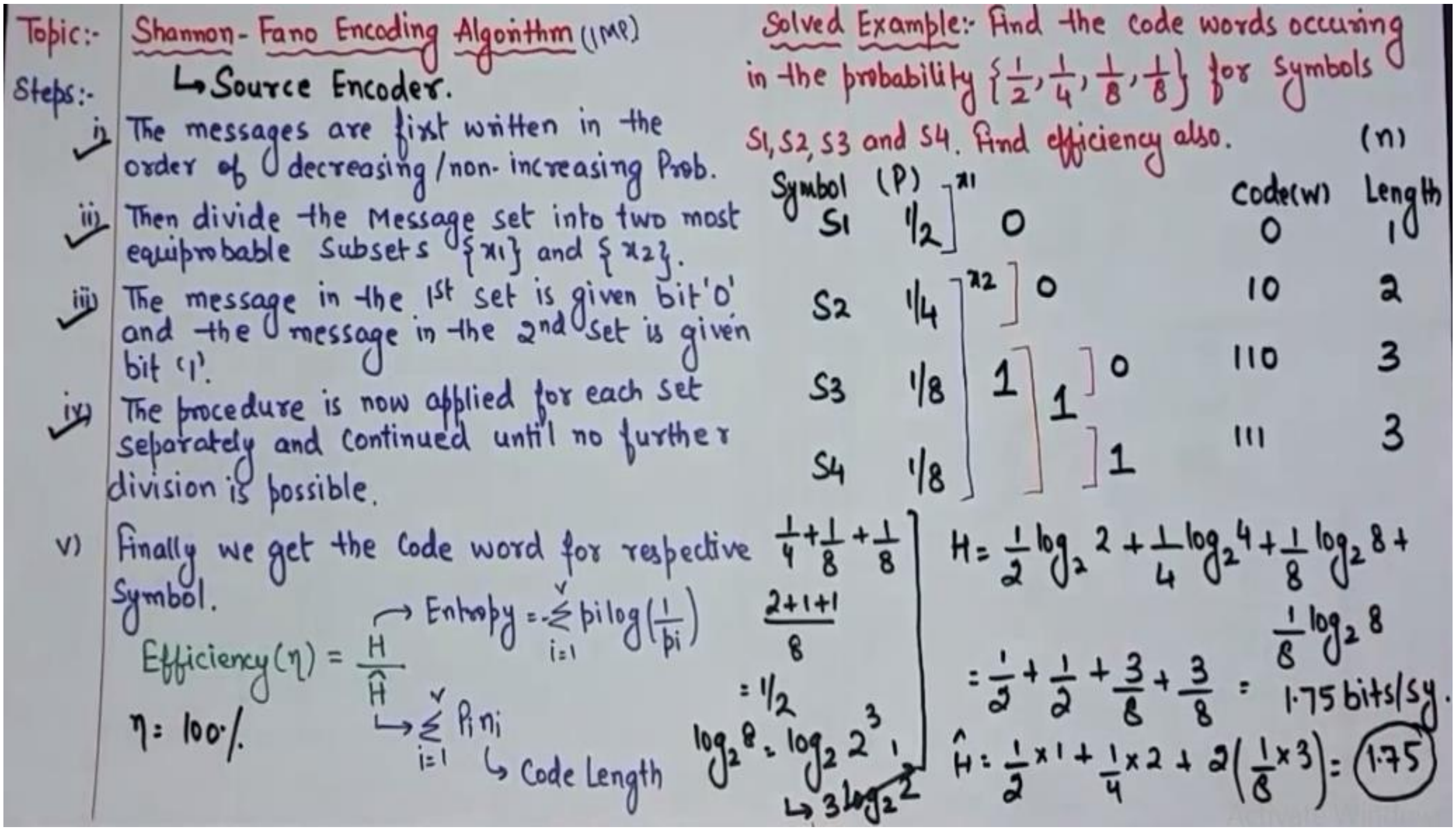
efficiency = $\eta = \frac{H}{L_{avg} \log_2 r}$ → base (2)

$= \frac{H}{L_{avg}}$

$H = 2.1216$ bits/symbol

$L_{avg} = 2.2$ bits/symbol

$\eta = 96.4\%$



Topic:- **Shannon-Fano Encoding Algorithm** (IMP)

↳ Source Encoder.

Steps:-

- i) ✓ The messages are first written in the order of decreasing / non-increasing Prob.
- ii) ✓ Then divide the Message set into two most equiprobable subsets $\{x_1\}$ and $\{x_2\}$.
- iii) ✓ The message in the 1st set is given bit '0' and the message in the 2nd set is given bit '1'.
- iv) ✓ The procedure is now applied for each set separately and continued until no further division is possible.
- v) Finally we get the code word for respective Symbol.

$$\text{Efficiency}(\eta) = \frac{H}{\hat{H}}$$

$H \rightarrow$ Entropy $= -\sum_{i=1}^{v} p_i \log\left(\frac{1}{p_i}\right)$

$\hat{H} \hookrightarrow \sum_{i=1}^{v} P_i n_i$ ↳ Code Length

$\eta = 100\%$

Solved Example:- Find the code words occurring in the probability $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$ for symbols S1, S2, S3 and S4. Find efficiency also.

Symbol	(P)				Code(w)	(n) Length
S1	1/2	x_1 0			0	1
S2	1/4	x_2 1	0		10	2
S3	1/8		1	0	110	3
S4	1/8			1	111	3

$\frac{1}{4}+\frac{1}{8}+\frac{1}{8}$

$\frac{2+1+1}{8}$

$=1/2$

$\log_2 8 = \log_2 2^3$ ↳ $3\log_2 2$

$$H = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \frac{1}{8}\log_2 8$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = 1.75 \text{ bits/sy.}$$

$$\hat{H} = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + 2\left(\frac{1}{8}\times 3\right) = \boxed{1.75}$$

m	p	stage1	stage2	stage3	codeword	Length
m_1	0.30	0	0		00	2
m_2	0.25	0	1		01	2
m_3	0.15	1	0	0	100	3
m_4	0.12	1	0	1	101	3
m_6	0.10	1	1	0	110	3
m_5	0.08	1	1	1	111	3

$$H = \sum_{K=1}^{6} P_K \log_2 \frac{1}{P_K}$$

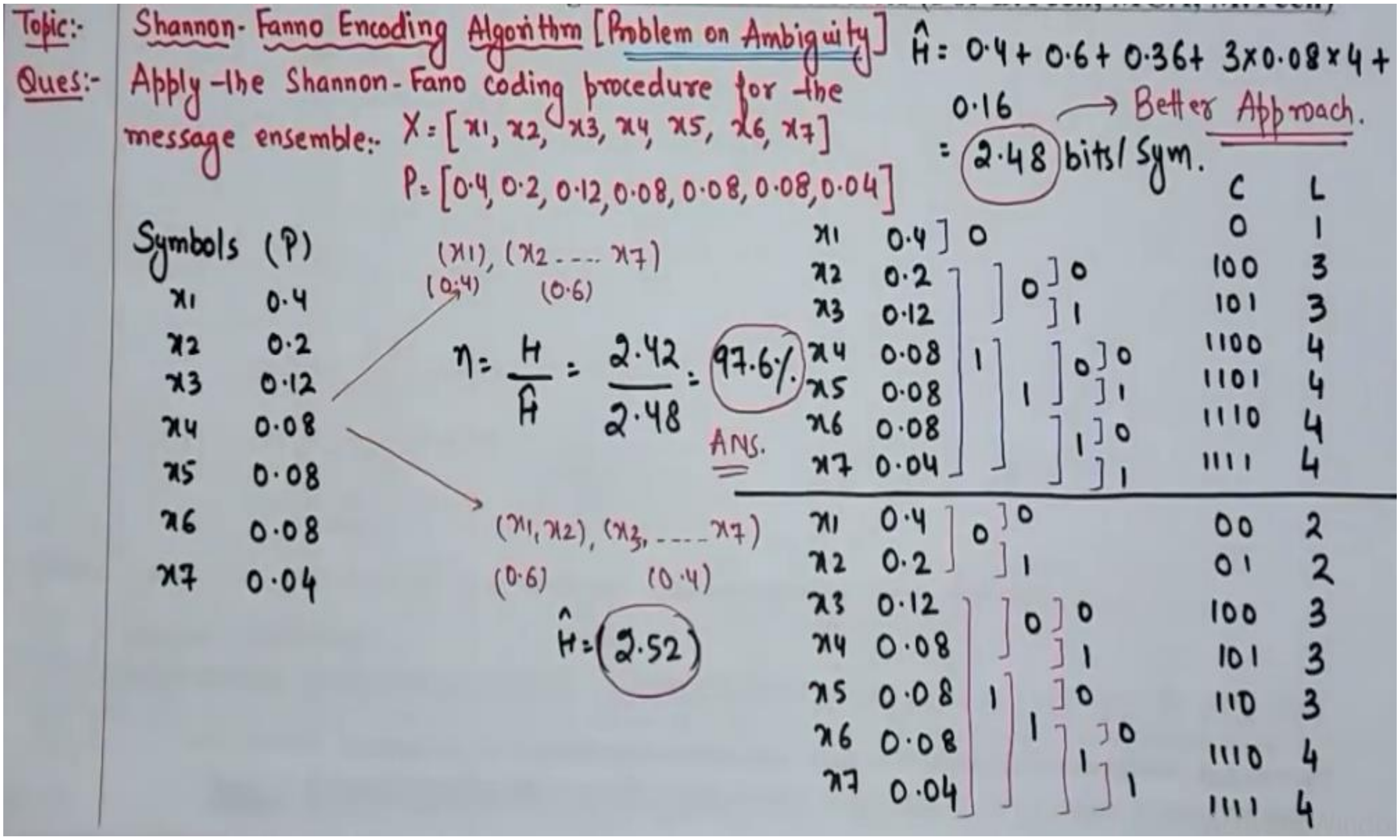
$$= (0.30)\log_2\frac{1}{0.30} + (0.25)\log_2\frac{1}{0.25} + (0.15)\log_2\frac{1}{0.15} + (0.12)\log_2\frac{1}{0.12} + (0.10)\log_2\frac{1}{0.10} + (0.08)\log_2\frac{1}{0.08}$$

$$= 2.418 \text{ bits}$$

$$L = \sum_{K=1}^{6} P_K n_K$$

$$= (0.30\times 2) + (0.25\times 2) + (0.15\times 3) + (0.12\times 3) + (0.10\times 3) + (0.08\times 3)$$

$$= 2.45 \text{ bits}$$



Topic:- Shannon-Fanno Encoding Algorithm [Problem on Ambiguity]

Ques:- Apply the Shannon-Fano coding procedure for the message ensemble:- $X = [x_1, x_2, x_3, x_4, x_5, x_6, x_7]$

$P = [0.4, 0.2, 0.12, 0.08, 0.08, 0.08, 0.04]$

Symbols	(P)
x_1	0.4
x_2	0.2
x_3	0.12
x_4	0.08
x_5	0.08
x_6	0.08
x_7	0.04

$(x_1), (x_2 \dots x_7)$
$(0.4) \quad (0.6)$

$$\eta = \frac{H}{\hat{H}} = \frac{2.42}{2.48} = \boxed{97.6\%}$$

ANS.

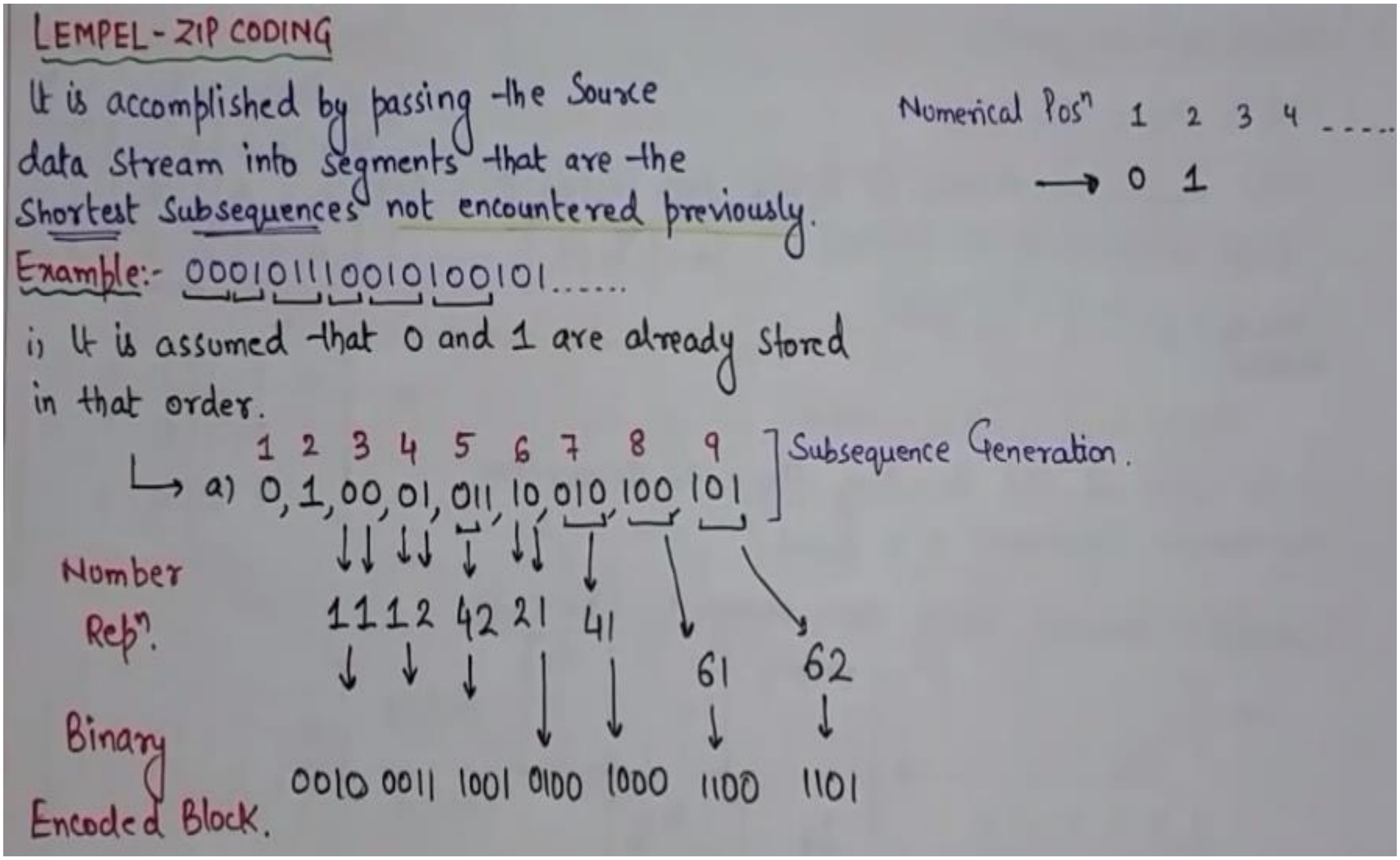
$(x_1, x_2), (x_3, \dots x_7)$
$(0.6) \quad (0.4)$

$$\hat{H} = \boxed{2.52}$$

$\hat{H} = 0.4 + 0.6 + 0.36 + 3 \times 0.08 \times 4 + 0.16$ → Better Approach.

$= \boxed{2.48}$ bits/Sym.

		C	L
x_1	0.4	0	1
x_2	0.2	100	3
x_3	0.12	101	3
x_4	0.08	1100	4
x_5	0.08	1101	4
x_6	0.08	1110	4
x_7	0.04	1111	4
x_1	0.4	00	2
x_2	0.2	01	2
x_3	0.12	100	3
x_4	0.08	101	3
x_5	0.08	110	3
x_6	0.08	1110	4
x_7	0.04	1111	4



LEMPEL-ZIP CODING

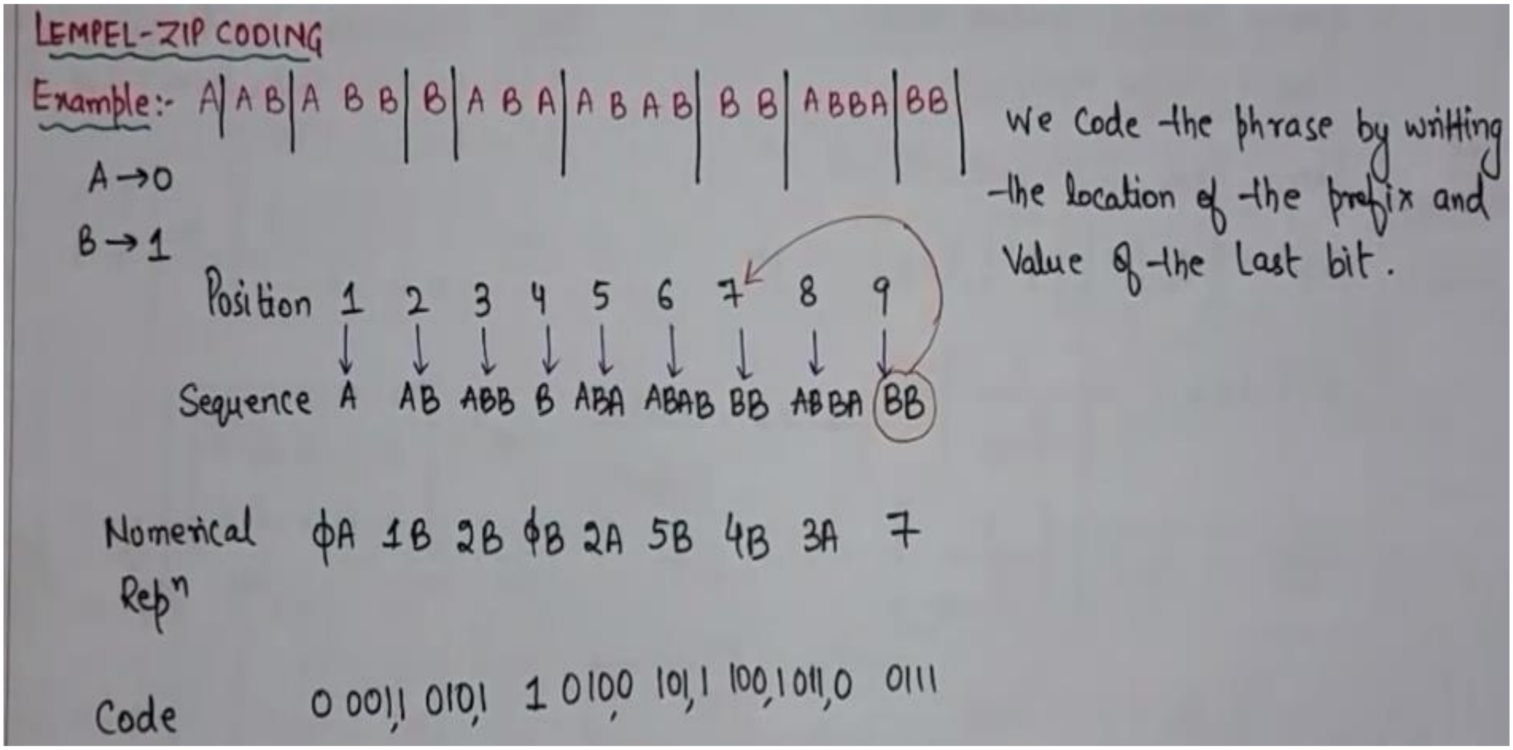
It is accomplished by passing the Source data stream into segments that are the Shortest Subsequences not encountered previously.

Nomerical Posⁿ 1 2 3 4 ....
→ 0 1

Example:- 000101110010100101......

i) It is assumed that 0 and 1 are already stored in that order.

	1	2	3	4	5	6	7	8	9
a) Subsequence Generation.	0	1	00	01	011	10	010	100	101
Nomber Repⁿ.			11	12	42	21	41	61	62
Binary Encoded Block.			0010	0011	1001	0100	1000	1100	1101



LEMPEL-ZIP CODING

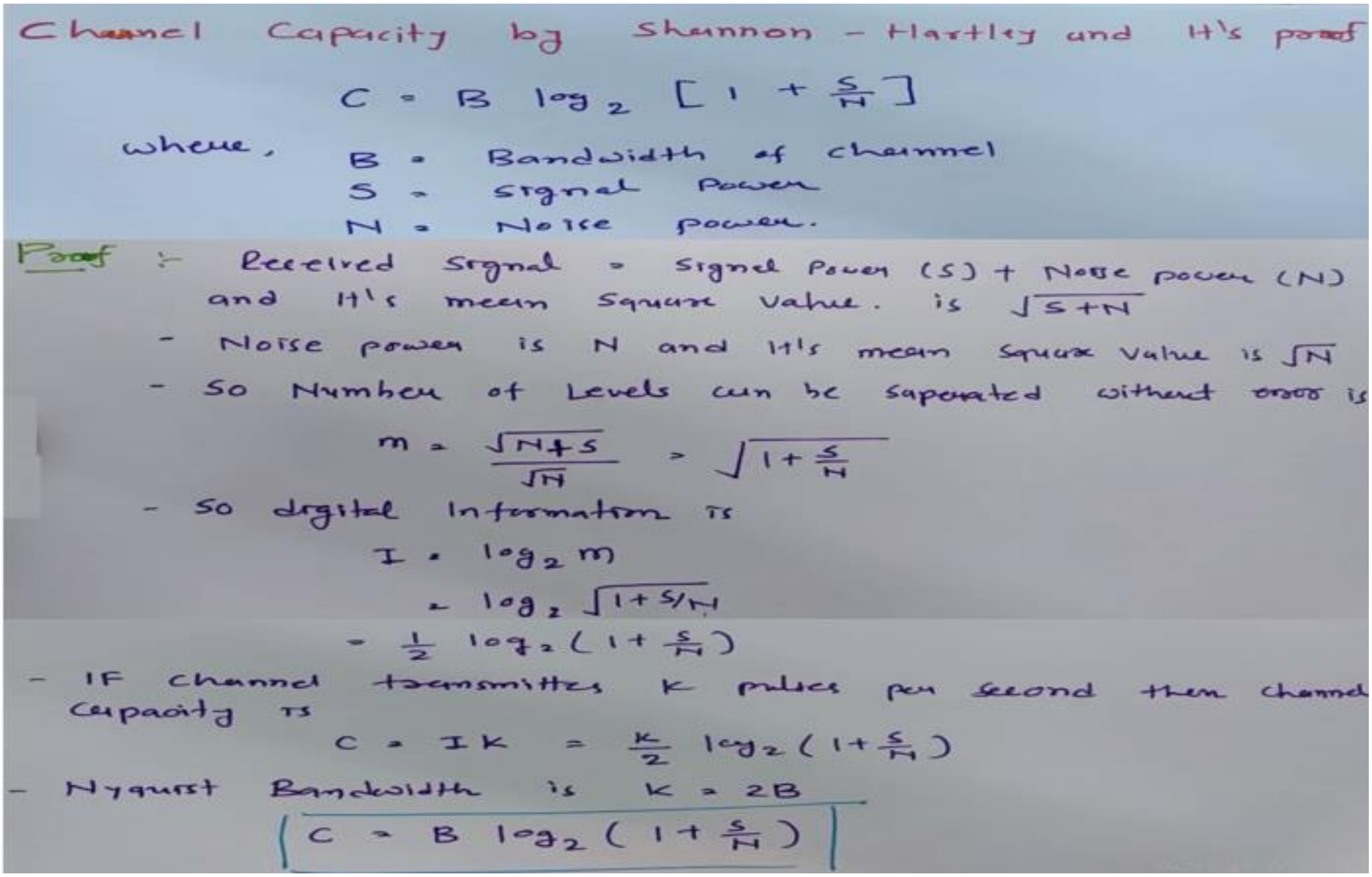
Example:- A|AB|A B B|B|A B A|A B A B|B B|A BBA|BB|

A → 0
B → 1

We code the phrase by writting the location of the prefix and Value of the Last bit.

Position	1	2	3	4	5	6	7	8	9
Sequence	A	AB	ABB	B	ABA	ABAB	BB	ABBA	BB
Nomerical Rep^n	ϕA	1B	2B	ϕB	2A	5B	4B	3A	7

Code 0 001,1 010,1 1 010,0 101,1 100,1011,0 0111



Channel Capacity by Shannon - Hartley and It's proof

$$C = B \log_2 \left[1 + \frac{S}{N} \right]$$

where,
B = Bandwidth of channel
S = Signal Power
N = Noise power.

Proof :- Received signal = Signal Power (S) + Noise power (N) and It's mean square value. is $\sqrt{S+N}$

- Noise power is N and It's mean square value is $\sqrt{N}$
- So Number of Levels can be saperated without error is

$$m = \frac{\sqrt{N+S}}{\sqrt{N}} = \sqrt{1+\frac{S}{N}}$$

- So digital Information is

$$I = \log_2 m$$
$$= \log_2 \sqrt{1+S/N}$$
$$= \frac{1}{2} \log_2 \left(1+\frac{S}{N} \right)$$

- IF channel transmittes K pulses per second then channel capacity is

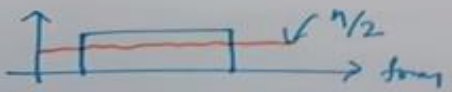
$$C = IK = \frac{K}{2} \log_2 \left(1+\frac{S}{N} \right)$$

- Nyqurst Bandwidth is K = 2B

$$\boxed{C = B \log_2 \left(1+\frac{S}{N} \right)}$$

Practically If we Increase B then noise N will also Increase. So Capacity of channel can not be Infinite.

If $\eta/2$ Is power density then



$$N = \eta B$$

So channel capacity is

$$C = B \log_2 \left(1 + \frac{S}{\eta B}\right)$$

$$= \frac{\eta B}{S}\left(\frac{S}{\eta}\right) \log_2\left(1+\frac{S}{\eta B}\right)$$

$$= \frac{S}{\eta}\left[\frac{\log_2\left(1+\frac{S}{\eta B}\right)}{\left(\frac{S}{\eta B}\right)}\right]$$

- For $C = \frac{S}{\eta} \log_2 e = 1.44\,\frac{S}{\eta}$ $_2e = 1.44$

$$C = \frac{S}{\eta} \log_2 e = 1.44\,\frac{S}{\eta}$$

SHANNON'S Channel Capacity:- "Given that a Source of M equally likely messages with M>>1, which is generating Info^n at a rate 'R': Given that a channel of Capacity 'C' exists.

if $\boxed{R \leq C}$ then there exists a Coding technique such that the o/p of Source may be transmitted over the channel with probability of error in the received message which may be made arbitrarily small."

W/o error in the presence of NOISE.

Shannon's theorem Negative Stmt:- "Given a Source of equally likely messages with M>>1, which is generating information at a rate R, if $\boxed{R>C}$ then the Probability of error is close to unity for every possible set of M transmitted Signals".

$\hookrightarrow$ Complexity of Coding ↑ → ↑ in Error.

According to Shannon's theorem, if (S/N) is Signal to noise ratio, then the maximum data rate is given by,

$$\boxed{R = B\log_2\left[1+\frac{S}{N}\right] \text{ bits/sec.}}$$

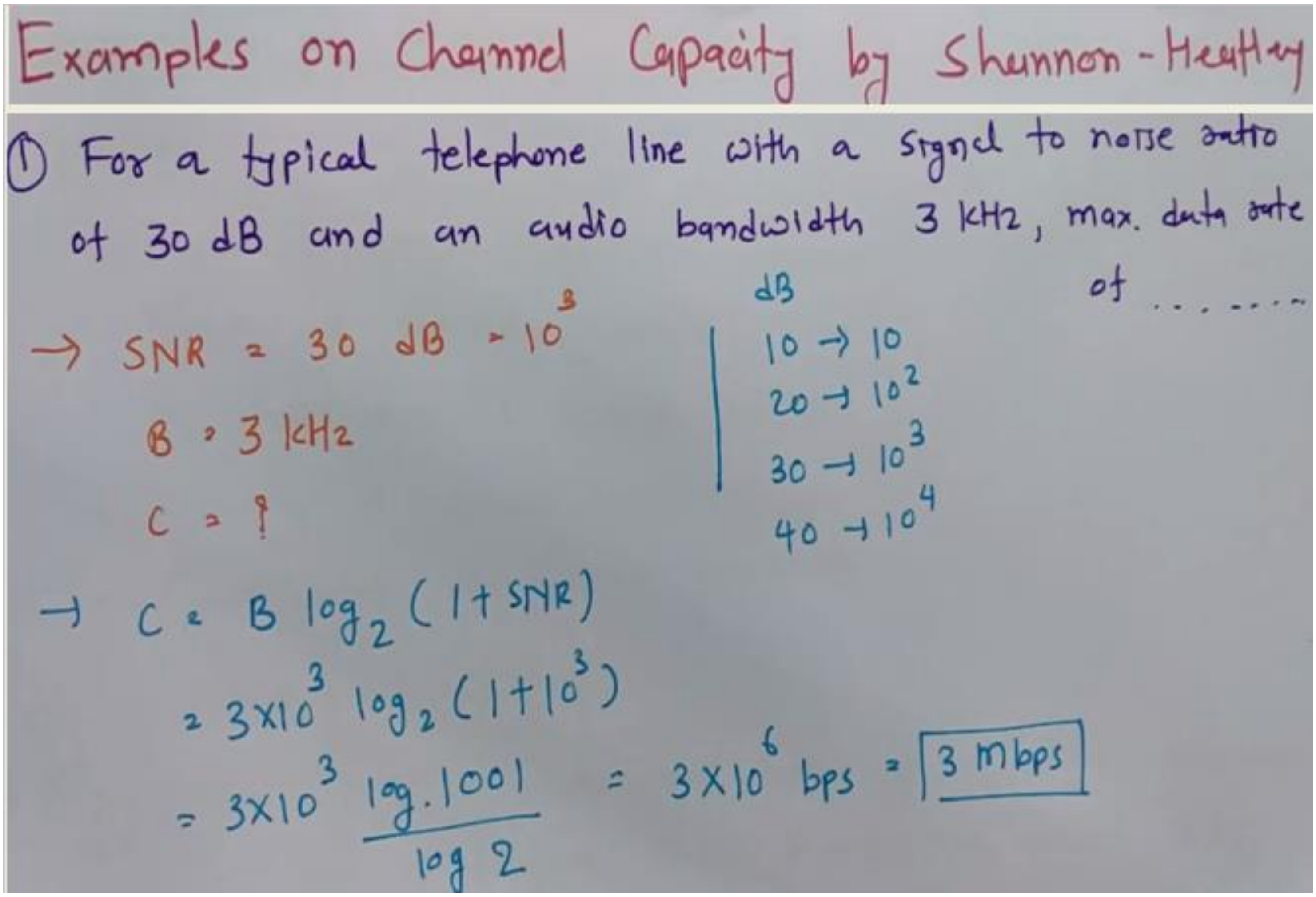
Ques 1) Find Maximum Bit Rate for Channel having bandwidth 3100 Hz and S/N ratio is 20dB.

$B = 3100$ Hz, $\frac{S}{N} = 20$dB, $\left[\left(\frac{S}{N}\right)_{dB} = 10\log_{10}\left(\frac{S}{N}\right)\right] \rightarrow 20\text{dB} = 10\log\left(\frac{S}{N}\right)$

$\frac{S}{N} = 100$

$$R_{max} = 3100\log_2\left[1+100\right] = \frac{3100\log_{10}(101)}{\log_{10}2}$$

$\hookrightarrow$ 20,640 bits/sec. ANS..



Examples on Channel Capacity by Shannon-Hartley

① For a typical telephone line with a signal to noise ratio of 30 dB and an audio bandwidth 3 kHz, max. data rate of ....

→ $SNR = 30\ dB = 10^3$

$B = 3\ kHz$

$C = ?$

dB

$10 \rightarrow 10$

$20 \rightarrow 10^2$

$30 \rightarrow 10^3$

$40 \rightarrow 10^4$

→ $C = B \log_2 (1 + SNR)$

$= 3 \times 10^3 \log_2 (1 + 10^3)$

$= 3 \times 10^3 \frac{\log .1001}{\log 2} = 3 \times 10^6\ bps = \boxed{3\ Mbps}$