DATE
SUBJ.

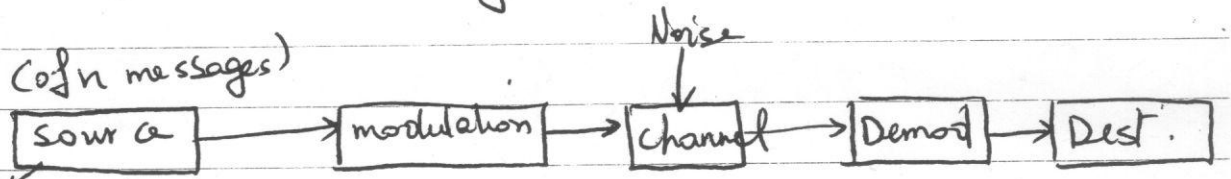
Inf Theory.

1.1 Introduction :

Information theory is the field that studies the properties of a communication system. It gives a quantitative measurement of a communication source and its ability and the system to handle information.

* The purpose of a communication system, is to carry information-bearing base-band signals from one place to another over a com channel.

1.2 Basic com system :



In the context of communications, information theory deals with mathematical modeling and analysis of a com. system rather than with physical sources & physical channels. In particular, it provides answers to two fundamental questions:

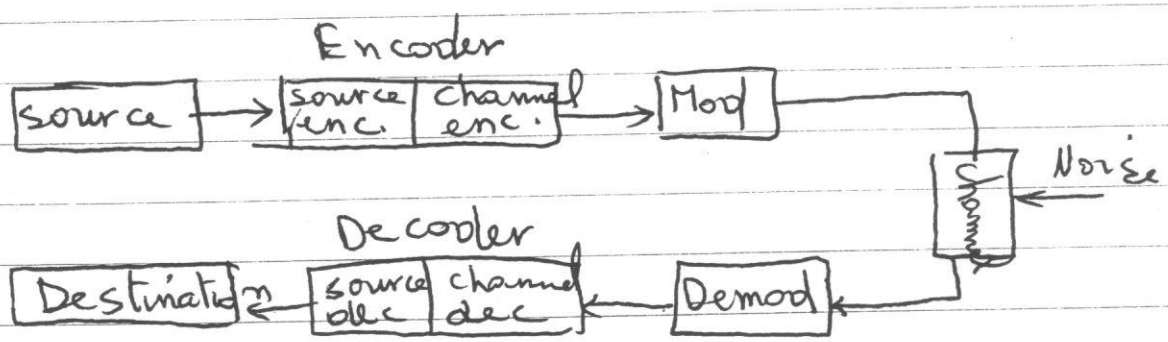
a) What is the irreducible complexity below which a signal cannot be compressed? (Entropy of a source). → e.g music & speech

DATE
SUBJ

b) What is the ultimate transmission rate for reliable communication over a noisy channel? ($R < C$).

1.3 A modified com. system :
According to "Shannon", for a given channel if the rate of transmission $R <$ the channel capacity, $\longrightarrow$ error free transmission can be achieved or Pe can be reduced (greatly) using encoding techniques.

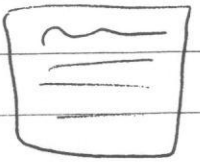
(Pe : it is the measure of performance in a digital environment Pe ↓ Perf ↑.



Source coding : Data compression (space required to store data to be transmitted is reduced and also reduce time of trans.)

channel coding (error control before entering the channel to reduce Pe).

* Ex of a com. system :
A Fax : a) scans tx paper
b) compresses the data
c) modulates d) transmits.
At RX : vice-versa.



DATE
SUBJ.

1.4. Basic Definitions: (Information, uncertainty, Entropy).

The amount of inf. is related to the prob. of occurrence of the event. (We mentioned that the source o/p is probabilistic in nature, so it can be modeled as a discrete r.v., S, which takes on symbols from a fixed finite alphabet:

$$S : \{s_0, s_1, \dots s_{k-1}\}$$

with prob $P(S = s_k) = p_k \quad k = 0, 1, \dots k-1$)

(We assume that the symbols emitted by the source during successive signaling intervals (unit of time) are statistically independent (discrete memoryless source)).

* The idea of inf. is closely related to the "uncertainty" or "surprise". It is related to the probability of occurrence of the event.
The amount of inf. received from a message is directly related to the uncertainty (unexpectness) or inversly related to the prob. of its occurrence.
When $P \longrightarrow 1$, $I \longrightarrow 0$, when $P \longrightarrow 0$, $I \longrightarrow \infty$
Hence, we can say (Tomorrow sun will rise in the east $P = 1 \quad I = 0$)

$I(m_i) \propto$ Uncertainty
As $P(m_i) \uparrow \quad I(m_i) \downarrow$
$P(m_i) \downarrow \quad I(m_i) \uparrow$,

$\therefore I(m_i) \propto \frac{1}{P_i}$

DATE
SUBJ.

$$I(m_i) = \log_2\left(\frac{1}{P_i}\right) = -\log_2 P_i$$

(in binary domain, $I = -\log_2 P_i$ bits = number of bits measuring the inf. content in this message.)

Note: If $P_i = 1 \longrightarrow I(m_i) = 0 \rightarrow$ no information is conveyed.

If $P(m_i) = \frac{1}{4}$. $I(m_i) = -\log_2 \frac{1}{4} = +\log_2 4 =$ 2 bits.

* As I increases number of bits required to measure the information increases.

* The information (content) I in a message can be interpreted as the min number of bits required to encode the message. (we will represent messages in codewords)

Information source [S] $\begin{cases} s_1 \rightarrow p_1 \\ s_2 \rightarrow p_2 \\ s_n \rightarrow p_n \end{cases}$

$I(s_1) = \log_2 \frac{1}{p_1}$

$I(s_n) = \log_2 \frac{1}{p_n}$

$$\left(\log_2 x = \frac{\log_e x}{\log_e 2} = \frac{\log_{10} x}{\log_{10} 2}\right) \qquad \sum_{i=1}^{n} P_i = 1$$

<u>Entropy H(X)</u> Entropy of the source;

$$H(X) = \sum_{i=1}^{n} P_i \log_2\left(\frac{1}{P_i}\right) \quad \left(\text{bits / message symbol}\right)$$

DATE
SUBJ.

The amount of inf. produced by the source during an arbitrary signaling interval depends on the symbol m_i emitted by the source at that time (and its prob. of occurrence.) Indeed, $I(m_i)$ (or $I(s_i)$) is a discrete r.v. that takes on the values $I(m_0), I(m_1) - - -$ with prob $p(m_0), p(m_1), --$ respectively. The mean of $I(m_i)$ over the source alphabet is given by:

$$H(X) = \left(E[I(m_i)]\right) = \sum_{i=1}^{n} P_i \log_2 \frac{1}{P_i}$$

bits / message.

= average amount of information associated with the source.
= average amount of ~~inf~~ uncertainty per message
= Least number of bits required to represent a message from the source.

$H(X) = \max$?

$H(X)$ is max when the uncertainty is max. This is when $p_1 = p_2 = - - = p_n = \frac{1}{n}$ (equiprobable events occur),

$$\underset{\max}{H(X)} = p_1 \log_2 \frac{1}{p_1} + p_2 \log \frac{1}{p_2} + - -$$

$$= \frac{1}{n} \times n \log_2 n = \log_2 n.$$

DATE
SUBJ.

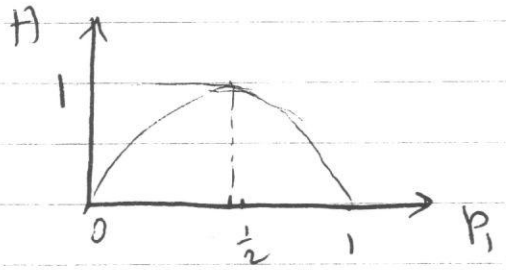
$$H(\max(X) = \log_2 n \quad \text{bits / symbol} \quad \text{message}.$$

Ex: source $x_1 \rightarrow p_1$ $\quad p_1 = p_2 = \frac{1}{2}$ $\quad H(X)_{\max} = \log_2 2 = 1$

$x_2 \rightarrow p_2$

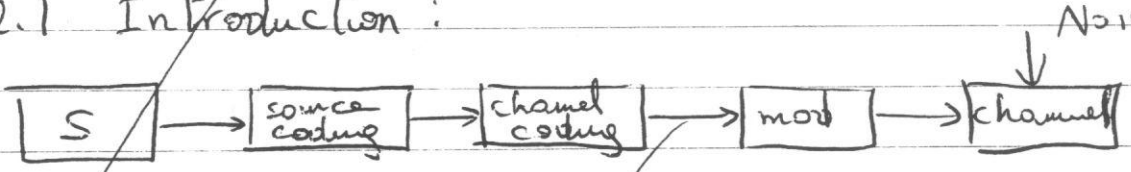
if $p_1 = 1$ $p_2 = 0$ $\quad H(X) = 0$

if $p_1 = 0$ $p_2 = 1$ $\quad H(X) = 0$



2. Source Coding

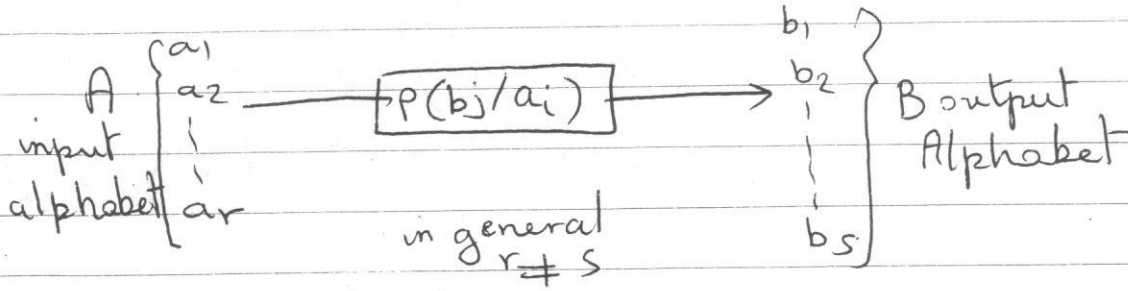
2.1 Introduction:



An important problem in communications is the efficient representation of data generated by a discrete source. The process by which this representation is accomplished is called source encoding. The device that performs this rep. is called a source encoder. For the source encoder to be efficient, we require knowledge of the statistics of the source. In particular if some source symbols are known to be more probable than others, then we may exploit this feature in the generation of a source code by assigning short codewords to frequent

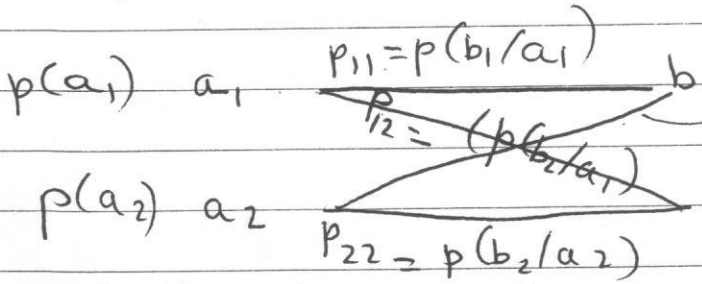
DATE
SUBJ.

3. Information Channels.



Def: An information channel is described by giving an input alphabet: $A = \{a_i\}$, $i = 1, 2, \ldots r$, an O/P alphabet $B = \{b_j\}$, $j = 1, 2, \ldots s$, and a set of conditional prob $p(b_j/a_i)$ for all i & j, ~~and a set of conditional prob $p(b_j/a_i)$ for all i & j~~.
(Channel transition prob. set — prob that b_j is received given that a_i was sent).

e.g. Binary channel:



$$P(b_1) = p(b_1/a_1)\,P(a_1) + p(b_1/a_2)\,p(a_2)$$

$$P(b_2) = p(b_2/a_1)\,P(a_1) + P(b_2/a_2)\,p(a_2)$$

Transition matrix:

$$\begin{array}{c} \\ a_1 \\ a_2 \end{array} \begin{array}{c} \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{bmatrix} p(b_1/a_1) & p(b_2/a_1) \\ p(b_1/a_2) & p(b_2/a_2) \end{bmatrix} \end{array}$$

each column corresponds to a channel O/P ; $\sum$ any row = 1

DATE
SUBJ.

in general

channel matrix $P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & & & \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix}$ each row corresponds to a fixed input

expressions for the probs of the various o/p symbols

$$P(b_1) = P(a_1)p_{11} + P(a_2)p_{21} + \cdots + P(a_r)p_{r1}$$
$$P(b_2) = P(a_1)p_{12} + P(a_2)p_{22} + \cdots + P(a_r)p_{r2}$$
$$\vdots$$
$$P(b_s) = P(a_1)p_{1s} + P(a_2)p_{2s} + \cdots + P(a_r)p_{rs}$$

conditional input probabilities:

$$P(a_i/b_j) = \frac{P(b_j/a_i)\,P(a_i)}{P(b_j)}$$

$$p(a_i/b_j) = \frac{p(b_j/a_i)\,P(a_i)}{\sum_{i=1}^{r} p(b_j/a_i)P(a_i)}$$

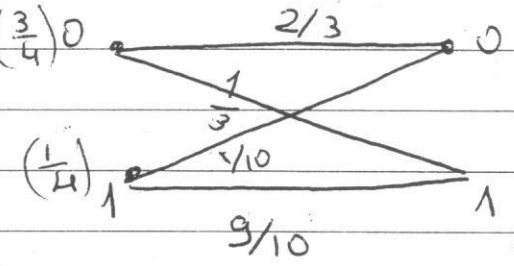
joint prob $P(a_i, b_j) = P(b_j/a_i)\,P(a_i)$
$= P(a_i/b_j)\,P(b_j)$

example :

DATE
SUBJ.

Example :

To illustrate the calculation of the various probabilities associated with an inf. channel, consider a binary channel; that is $A = \{0, 1\}$ & $B = \{0, 1\}$.



A noisy information channel.

The channel matrix $P = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$

The probabilities of the output symbols are;

$$P_r(b=0) = \left(\frac{3}{4}\right)\left(\frac{2}{3}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{10}\right) = \frac{21}{40}$$

$$P_r(b=1) = \left(\frac{3}{4}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{4}\right)\left(\frac{9}{10}\right) = \frac{19}{40}$$

$\Sigma = 1$

The conditional input probabilities:

$$P_r(a=0/b=0) = \frac{20}{21} \qquad P_r(a=1/b=0) = \frac{1}{21}$$

$$P_r(a=0/b=1) = \frac{10}{19} \qquad P_r(a=1/b=1) = \frac{9}{19}$$

The prob of various joint events are calculated as:

$$P_r(a=0, b=0) = P_r(b=0/a=0)\,p(a=0) = \left(\frac{2}{3}\right)\left(\frac{3}{4}\right) = \frac{2}{4} = \frac{1}{2}$$

Similarly, the other joint probs can be calculated.

DATE
SUBJ.

Channel Entropies:

$$H(A) = -\sum_{i=1}^{r} p(a_i) \log_2 p(a_i)$$ = source Entropy (a priori Entropy)

$$H(B) = -\sum_{j=1}^{s} p(b_j) \log_2 p(b_j)$$ = Destination entropy

$$H(A,B) = -\sum_{i=1}^{r} \sum_{j=1}^{s} p(b_j, a_i) \log_2 p(a_i, b_j).$$

joint entropy of the channel inputs & outputs.

$$H(B/A) = -\sum_{i=1}^{r} \sum_{j=1}^{s} p(b_j, a_i) \log_2 p(b_j/a_i)$$

(conditional entropy). Average amount of uncertainty at received symbols given that A was transmitted.

$$H(A/B) = -\sum_{i=1}^{r} \sum_{j=1}^{s} p(b_j, a_i) \log_2 P(a_i/b_j)$$

(a posteriori probabilities). Average amount of uncertainty of A (uncertainty remaining about the input) given that B was received).

* $H(A/B)$ = measure of the amount of inf. lost in the channel due to channel C/C's.

If $H(A/B) = 0 \longrightarrow$ No loss in the channel (error (noise) free channel).

$H(A/B) \leq H(A)$.

. If $H(A/B) = H(A) \longrightarrow$ All information is lost.

DATE
SUBJ.

* By Shannon's first theorem we need an average of H(A) bits to specify one input symbol a_i. An average of H(A/B) bits/message are lost in the channel. Therefore, on the average, observation of a single output symbol provides us with (H(A) - H(A/B)) bits of information. (mutual information of A & B) or mutual inf. of the channel.

$$I(A;B) = H(A) - H(A/B)$$

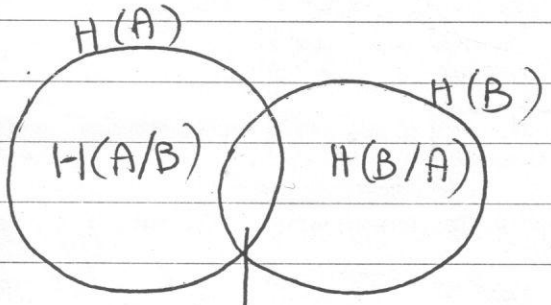


Fig. Illustrating the relations among various channel entropies.

$$H(A) - H(A/B) = H(B) - H(B/A)$$
$$I(A;B) \qquad I(B;A)$$

$$\therefore I(A;B) = H(A) - H(A/B) = H(B) - H(B/A)$$

* relation bet mutual inf & joint entropy of the channel inputs & outputs:

$$H(A,B) = H(A) + H(B/A) = H(B) + H(A/B)$$

$$H(A,B) = H(A) + H(B) - I(A;B).$$

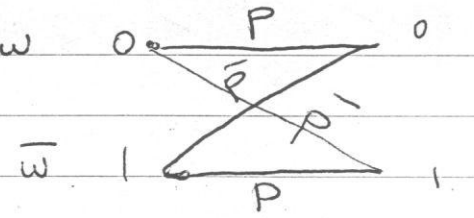
Properties of mut inf: $I(A;B) \geq 0$ (not negative)
$I(A;B) = I(B;A)$.

DATE
SUBJ.

ex: Compute the mutual information for a BSC.
The channel matrix of the BSC is:

$$\begin{bmatrix} p & \bar{p} \\ \bar{p} & p \end{bmatrix}$$

$\bar{p} = 1 - p$



p_0 = prob of sending x_0 (0)
$p_1 = 1 - p_0$ x_1 (1)

p = transition prob of the channel

$$I(x;y) = H(z) - H(P)$$

where $H(z) = z \log_2\left(\frac{1}{z}\right) + (1-z)\log_2\left(\frac{1}{1-z}\right)$

$$z = p_0 P + (1-p_0)(1-p)$$

&

$$H(P) = p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right).$$

DATE
SUBJ.

Channel Capacity:

From entropy relations we can show that:

$$I[A;B] = \sum_i \sum_j p(a_i, b_j) \log_2 \left[\frac{p(b_j/a_i)}{p(b_j)} \right]$$

$$= I[B;A] \qquad (1)$$

but $p(a_i, b_j) = p(b_j/a_i) p(a_i)$

Also: $p(b_j) = \sum_{i=1}^{r} p(b_j/a_i) p(a_i)$

From the above 3 equations, we see that it is necessary for us to know the input prob distributions (a priori probs) $p\{a_i = 1, 2, \ldots r\}$ so that we may calculate the mutual inf. The mutual information of a channel therefore depends not only on the channel, but also on the way in which channel is used. The input prob distribution $p\{a_i\}$ is obviously independent of the channel. Hence, we can maximize the ~~average~~ mutual information $I(A;B)$ of the channel with respect to $p(a_i)$.

- We define the channel capacity of a discrete memoryless channel as the max ~~average~~ mutual information $I(A;B)$ in any single use of the channel (i.e signaling interval), where the maximization is ~~all~~ over all possible ~~input~~ input prob distributions $\{p(a_i)\}$ on A.

DATE
SUBJ.

Thus, $C = \max [I(A;B)]$ bits/message (transmission)

Note that the channel capacity C is a function only of the transition probabilities $p(b_j / a_i)$ which define the channel.

The calculation of C involves maximization of the ~~average~~ mutual inf. $I(A;B)$ over r variables {i.e. the ~~prob~~ input probabilities $p(a_1) = \dots = p(a_r)$} subject to 2 constraints:

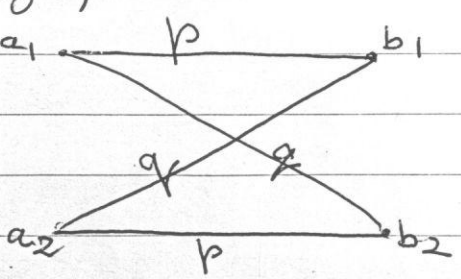
$$p(a_i) \geq 0 \quad \text{for all } a_i$$

$$\sum_{i=1}^{r} p(a_i) = 1$$

* Binary Symmetric channel:

For a Binary channel the entropy $H(m)$ (source entropy) is maximized when $p(a_1) = p(a_2) = \frac{1}{2}$ (a_1 or a_2 are each 0 or 1), the mutual inf. is ($p(a_1)$ & $p(a_2)$ are the channel input probabilities).

similarly maximized, so that we can write;



p = conditional prob. of a correct reception
q = conditional prob of error reception

symmetric ($p_{11} = p_{22} = p$, $p_{12} = p_{21} = q$).

DATE
SUBJ.

$$C = I(A;B) / p(a_1) = p(a_2) = \frac{1}{2}$$

& we have, $p(b_2/a_1) = p(b_1/a_2) = q = (1-p)$.

$$p(b_2/a_2) = p(b_1/a_1) = 1-q = p$$

substitute in (1) $\longrightarrow$ $C = 1 + p\log_2 p +$

$(1-p)\log_2(1-p)$.

DATE
SUBJ.

Channel - Coding
Coding for Reliable Rigital Transmission.

* Messages of a source entropy $H(S)$ can be encoded by using an average of $H(S)$ digits per message. This encoding has zero redundancy. Hence, if we transmit these coded messages over a noisy channel, some of the information will be received erroneously.

→ Use of redundancy, in general, helps combating noise. This can be seen from a simple example of a single parity-check code, in which an extra bit is added to each code to ensure that the number of 1's in the resulting codeword is always even (or odd). If a single error occurs in the received codeword, the parity is violated, and the receiver requests retransmission. (This example demonstrates the utility of redundancy).

ex: 1100 0 even parity:
if received 11100 → parity is odd
∴ there is error.

1) * Why channel coding:

Prob of error of a particular signalling scheme is function of the S/N at the receiver input:

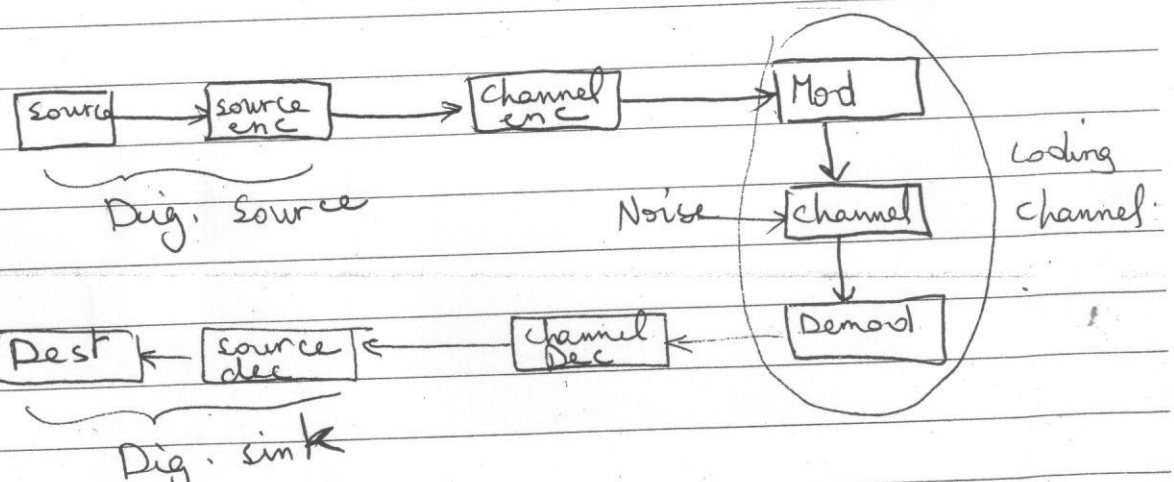
$$P_E = F(SNR)$$

DATE
SUBJ.

Power of signal & BW are restricted to rules; we cannot increase them freely - Else, $\eta/2$ (power spectral density of noise) is fixed for a certain environment + parameters of signalling schemes such as the number and the type of signals used are chosen to minimize the complexity & cost of the equipment. With all these constraints, it is not possible to arrive at a signalling scheme which will yield an acceptable prob of error ⟶ use of channel coding.

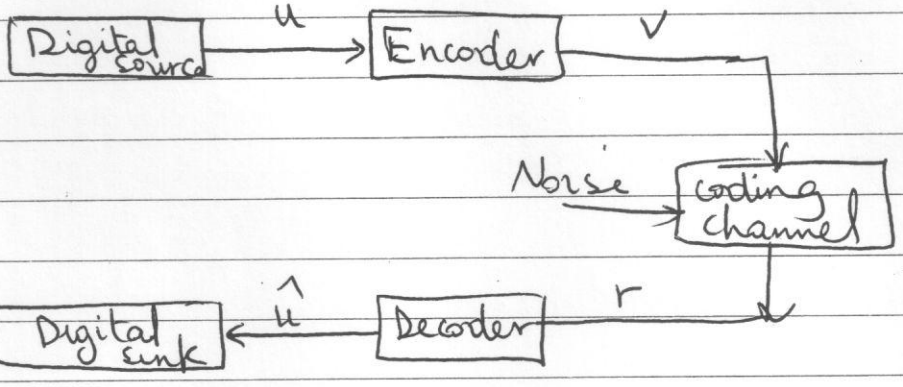
* In 1948, Shannon demonstrated that by proper encoding of the information, errors induced in the channel or storage medium can be reduced to any desired level. {Now the use of coding for error control has become an integral part in the design of modern com. systems & digital computers).

A typical transmission system B.D:



DATE
SUBJ.

simplified to :



Now the problem is to design an enc & a decoder such that :

1) inf can be transmitted in a noisy environment as fast as possible

2) reliable reproduction of the inf. can be obtained at the o/p of the decoder

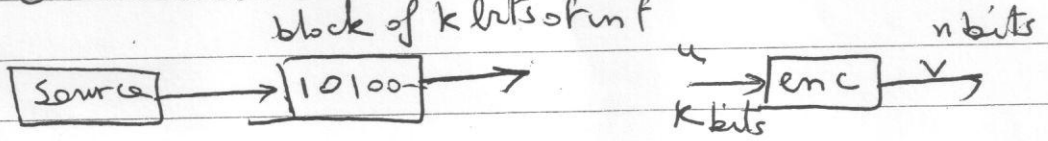
& 3) the implementation cost of enc. & dec. falls into acceptable limits.

DATE
SUBJ.

2) Types of Codes :

There are two types of codes in common use today; Block and Convolutional codes :

Block codes :



The encoder divides the information source output into message blocks of length k inf bits each ($u = u_1, u_2, \dots u_k$) called a k-tuple message (2^k different possible messages are available). The encoder transforms $u \longrightarrow$ codeword $v = (v_1, v_2, \dots v_n)$ of length n discrete symbols (2^k codewords are available) $\Longrightarrow$ This set is called an (n, k) block code of rate k/n. The encoder is memoryless [The output v depends only on the corresponding k-bit message at the same time unit].

* n - k redundant bits are added to each message to provide the code with the capability of combating channel noise. How to use these redundant bits to achieve reliable transmission over a noisy channel is the major problem in designing an encoder.

example : A Binary block code with $k = 4$ & $n = 7$,

DATE
SUBJ.

Messages	codewords
0000	0000000
1000	1101000
0100	0110100
1100	1011100
⋮	⋮
1111	1111111

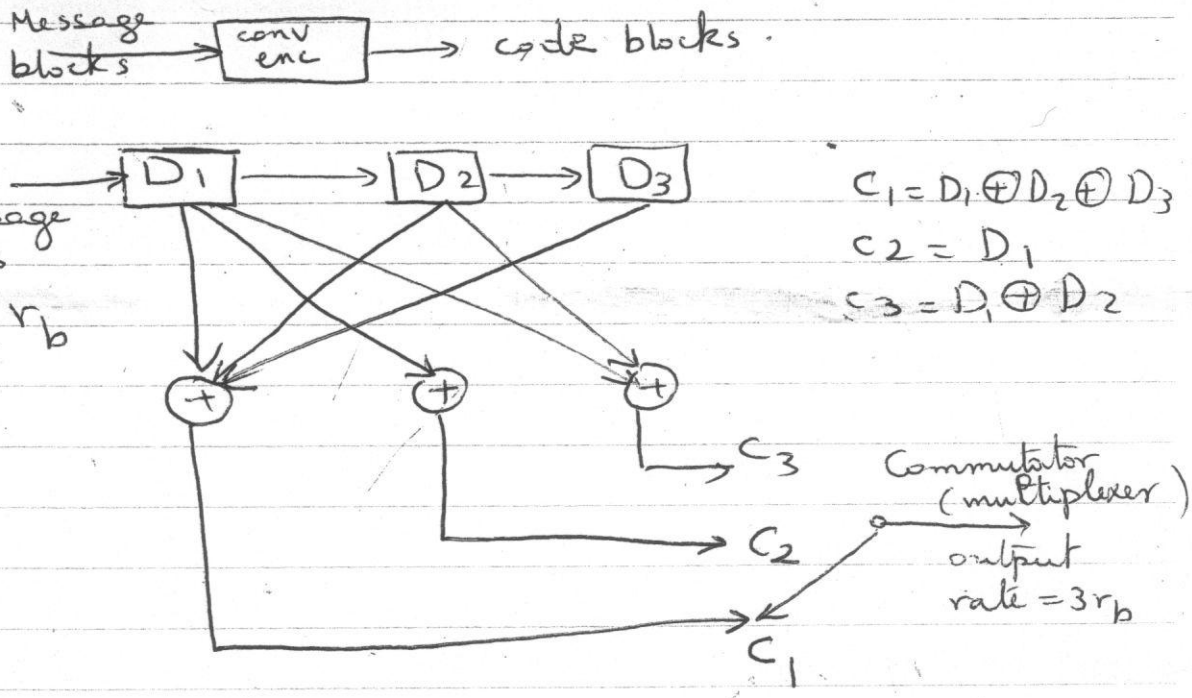
Table 1.

For a convolutional code:
The encoder also accepts k-bit blocks of the inf sequence u & produces an encoded sequence of length n. However, the n-digit code block depends not only on the k-digit message block of the same time unit, but also on (N-1) previous message blocks (~~in previous~~ N-1 ~~message~~ previous time units). Hence the encoder has a memory of order N-1. The set of encoded messages produced by a K-input, n-output encoder of memory N-1 is called an (n, K) convolutional code of constraint length nN digits & rate efficiency K/n.

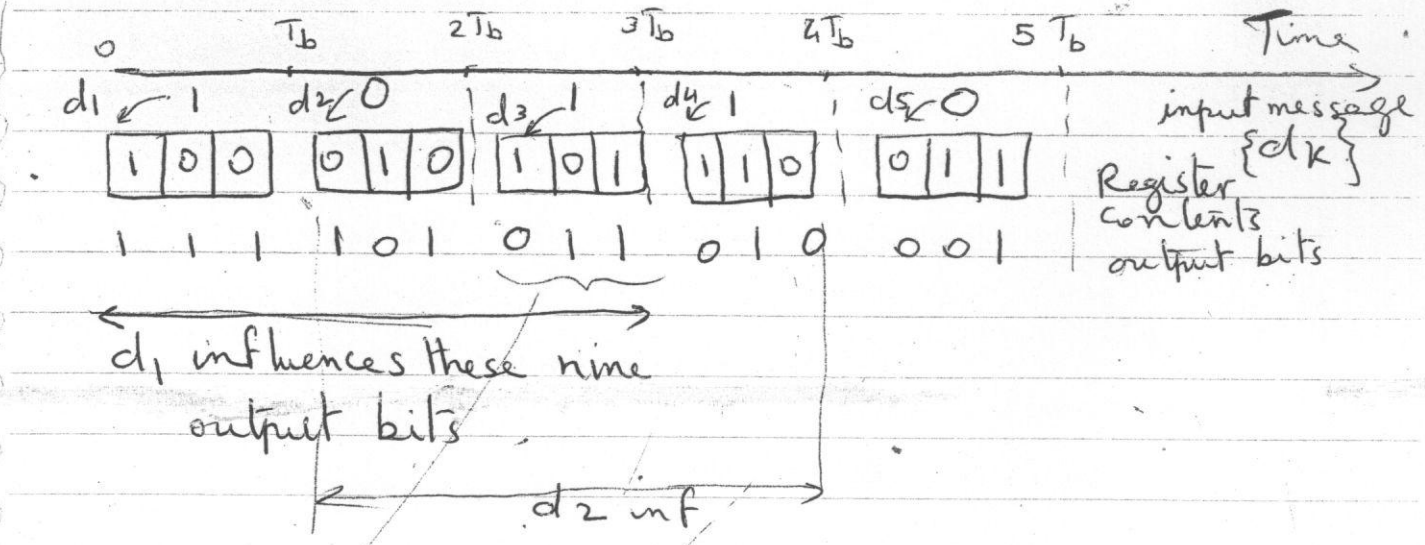
Example: (The encoder consists of shift registers & modulo-2 adders). The parameters of the encoder k, n & N are in general small integers.).

* An encoder for a (3,1) convolutional code with a

constraint length of 9 bits is shown.
$nN = 9 \rightarrow N = 3$ & number of shift register stages $= N = 3$.



The operation is as follows:
We assume the S.R. is clear initially. The first bit of the input data is entered into D_1. During this message bit interval, the commutator samples the mod-2 adder outputs C_1, C_2 & C_3. Thus a single message bit yields 3 output bits. The next message bit now enters D_1 while the content of D_1 is shifted into D_2 & the commutator again samples the 3 adder outputs. The process is repeated until the last message digit is shifted into D_3.



* Encoding operation of the conv enc,

The output for message block d_3 depends not only on d_3 but also on d_1 & d_2 (= N-1 previous inputs).

(constraint length ≡ is defined as the number of shifts over which a single message bit can influence the encoder output).