SINCE X_i NOT DEPENDENT ON OTHER OUTPUTS $\Rightarrow$ CALLED DISCRETE MEMORYLESS SOURCE

$\uparrow$ IN TIME AND AMPLITUDE

$\uparrow$ INDEPENDENCE

OR DMS

EXAMPLE : ENGLISH ALPHABET

$\boxed{\text{DMS}} \longrightarrow$ ... HELP ...

$N = 26$ POSSIBLE CHARACTERS OR SYMBOLS

$$\mathcal{a} = \{A, B, \ldots, Z\} = \{a_1, a_2, \ldots, a_N\}$$

WITH $P[X_i = a_n] = p_n$ ALL i

$\mathcal{a}$ IS CALLED THE ALPHABET

EXAMPLE : COIN TOSSING

$\boxed{\text{DMS}} \longrightarrow$ ... 01101 ...

1 $\Rightarrow$ HEADS
0 $\Rightarrow$ TAILS

HERE $\mathcal{a} = \{0, 1\}$

$$P[X_i = 1] = p = p_1$$
$$P[X_i = 0] = 1 - p = p_2$$

COIN TOSSES ARE INDEPENDENT

ALSO, CALLED A BINARY SOURCE

(JUST A BERNOULLI RANDOM PROCESS!)

MEASURE OF INFORMATION

NEED QUANTITATIVE MEASURE (NOT VALUE OF INFORMATION)

EXAMPLE: EVENT = "IT RAINED YESTERDAY"
⇒ SAME INFORMATION FOR FARMER WHO MUST DECIDE IF CROPS SHOULD BE WATERED OR FOR FACTORY WORKER WHO MUST DECIDE WHETHER TO EAT HIS LUNCH OUTSIDE

INTUITIVE PROPERTIES - INFORMATION SHOULD DEPEND ON HOW PROBABLE EVENT IS

EXAMPLE - 1) "CONGRESSMAN INDICTED FOR FRAUD" VS
2) "CANCER CURE DISCOVERED"

$INFO._2 >> INFO._1$

ONCE REVEALED LESS LIKELY EVENTS CONVEY MORE INFORMATION

ANOTHER PROPERTY IS THAT REVEALING TWO INDEPENDENT EVENT OUTCOMES YIELDS SUM OF INFORMATIONS.

EXAMPLE - 1) "LOC NESS MONSTER FOUND"
+ "URI GOES TO NCAA FINALS"
VS 2) "CURE FOR PROSTRATE CANCER DISCOVERED" + "CURE FOR COLON CANCER DISCOVERED"

$INFO_1$ = SUM BUT NOT $INFO_2$

ALSO, INFORMATION SHOULD BE POSITIVE (CAN'T CANCEL OUT FOR INDEPENDENT EVENTS)

NOW RETURN TO DMS - HOW MUCH INFORMATION IS BEING GENERATED?

$$\mathcal{A} = \{a_1, a_2, \dots, a_N\} \qquad P[X_i = a_n] = p_n$$

A MEASURE OF INFORMATION THAT HAS ALL THE INTUITIVE AND SOME IMPORTANT MATHEMATICAL PROPERTIES IS

$$I(p_j) = \text{LOG } \frac{1}{p_j}$$

DEPENDS ON PROBS. OF EVENTS ONLY, NOT VALUES, i.e. NOT DEPENDENT ON a_j

FOR $0 < p_j \leq 1$

1) $I(p_j) \geq 0$

2) AS p_j DECREASES, $I(p_j)$ INCREASES

3) FOR JOINT EVENTS (TWO OUTCOMES) THE j^{th} OUTCOME IS $x_j = (x_{j_1}, x_{j_2})$ WITH $P(x_j = a_j) = P(x_{j_1} = a_{j_1}, x_{j_2} = a_{j_2})$. IF INDEPENDENT

$$P(x_j = a_j) = P(x_{j_1} = a_{j_1})\, P(x_{j_2} = a_{j_2})$$

$$p_j = p_{j_1} p_{j_2}$$

$$\Rightarrow I(p_j) = \text{LOG} \frac{1}{p_j} = \text{LOG} \frac{1}{p_{j_1} p_{j_2}}$$

$$= \text{LOG} \frac{1}{p_{j_1}} + \text{LOG} \frac{1}{p_{j_2}}$$

$$= I(p_{j_1}) + I(p_{j_2})$$

$\Rightarrow$ INFO. IS ADDITIVE FOR INDEPENDENT EVENTS

$I(p_j)$ CALLED THE <u>SELF-INFORMATION</u> USUALLY MEASURED IN BITS PER SOURCE SYMBOL

$$I(p_j) = \text{LOG}_2 \frac{1}{p_j}$$

= NUMBER OF BITS OF INFO. OBTAINED WHEN

WE OBSERVE THAT $X = a_j$

EXAMPLE (*) $a = \{A, B, C, D\}$ EQUALLY LIKELY
$p_j = 1/4$

$$I(p_j) = \log_2 \frac{1}{p_j} = 2 \text{ BITS/SYMBOL}$$
$$j = 1, 2, 3, 4$$

EXAMPLE (**) $a = \{A, B, C, D\}$

$$p_1 = \frac{1}{2}, \quad p_2 = 1/4, \quad p_3 = 1/8, \quad p_4 = 1/8$$

$$I(p_j) = \begin{cases} 1 & j = 1 \\ 2 & j = 2 \\ 3 & j = 3, 4 \end{cases}$$

SINCE A DMS WILL EMIT ALL SYMBOLS AN AVERAGE MEASURE OF INFO. IS

$$E[I(p)] = \sum_{i=1}^{N} I(p_i)\, p_i$$
$$= \sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i}$$
$$= - \sum_{i=1}^{N} p_i \log_2 p_i \quad \text{BITS/SYMBOL}$$
$$= H(X)$$

SLIGHT ABUSE OF NOTATION

THIS INFO. CONTENT OF SOURCE IS CALLED THE ENTROPY. (UNCERTAINTY OR EQUIVOCATION)

EXAMPLE (*): $I(p_j) = 2 \quad j = 1, 2, 3, 4$

$\Rightarrow H(X) = E\,[I(p)] = 2$

EXAMPLE (**): $I(p_j) = \begin{cases} 1 & j=1 \\ 2 & j=2 \\ 3 & j=3 \\ 3 & j=4 \end{cases}$

$$H(X) = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} + 3 \cdot \tfrac{1}{8} + 3 \cdot \tfrac{1}{8}$$
$$= 1\,3/4 < 2 \qquad \text{WHY ?}$$

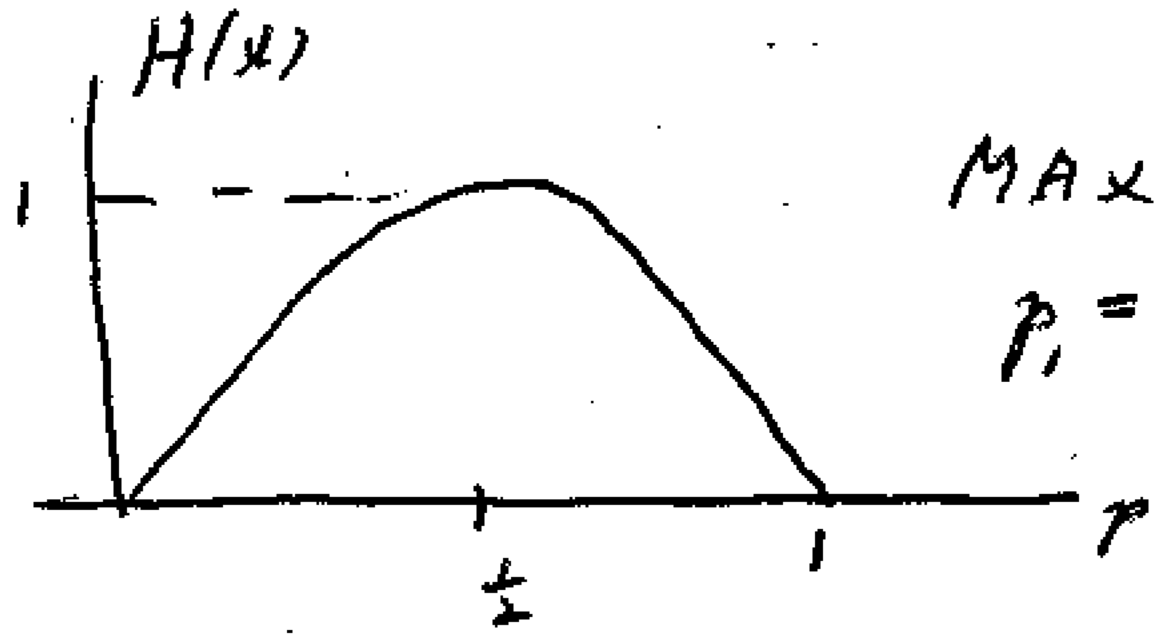
$$H(X) = -\sum_{i=1}^{N} p_i \,\mathrm{LOG}_2\, p_i \geq 0 \qquad \text{WHY ?}$$

(DEFINE $0 \,\mathrm{LOG}_2\, 0 = 0$, EVENT HAS PROB. $= 0$)

EXAMPLE: BINARY DMS $(N=2)$

$p_1 = p, \quad p_2 = 1-p$

$$H(X) = -p\,\mathrm{LOG}_2\, p - (1-p)\,\mathrm{LOG}_2\,(1-p)$$



MAX AT $p_1 = p_2 = \frac{1}{2}$ ..

NOTE: $H(X)\big|_{p=0} = H(X)\big|_{p=1} = 0 \qquad \text{WHY ?}$

EXAMPLE: ANALOG SOURCE HAS $W = 4000$ Hz

SAMPLE AT $f_s = 2W$

QUANTIZE TO $a = \{-2, -1, 0, 1, 2\}$

WITH $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$

$p_4 = \frac{1}{16}$, $p_5 = \frac{1}{16}$

WHAT IS INFORMATION RATE OF SOURCE?

$$H(X) = \tfrac{1}{2}\log_2 2 + \tfrac{1}{4}\log_2 4 + \tfrac{1}{8}\log_2 8 + \tfrac{1}{16}\log_2 16 + \tfrac{1}{16}\log_2 16 = 15/8$$

BITS / SAMPLE

INFO. RATE = (BITS/SAMPLE) × (SAMPLES/SEC)

$= H(X) f_s = ((15)/8)(8000)$

$= 15{,}000$ BITS/SEC

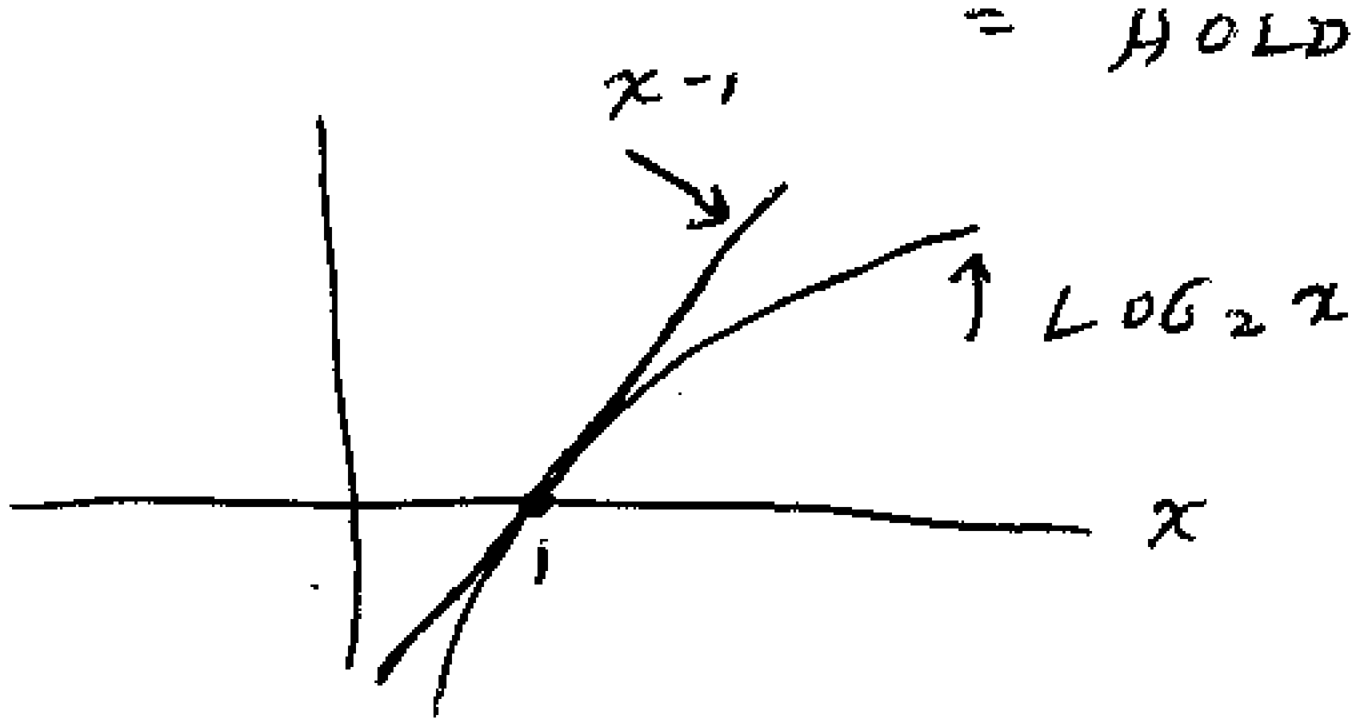
ENTROPY IS MAXIMUM FOR $p_j = 1/N$

$$H(X) = \sum_{j=1}^{N} p_j \log_2 \frac{1}{1/N} = \log_2 N \sum_{j=1}^{N} p_j$$

$= \log_2 N$ BITS/SYMBOL

PROOF: NEED $\log_2 x \le x - 1 \quad x > 0$

= HOLDS IFF $x = 1$



$$H(X) - \log_2 N = \sum_{j=1}^{N} p_j \log_2 \frac{1}{p_j} - \log_2 N$$

$$= \sum_{j=1}^{N} p_j \left(\log_2 \frac{1}{p_j} - \log_2 N\right)$$

$$= \sum_{j=1}^{N} p_j \log_2 \frac{1}{Np_j} \leq \sum_{j=1}^{N} p_j \left(\frac{1}{Np_j} - 1\right)$$

$$= \sum_{j=1}^{N} \left(\frac{1}{N} - p_j\right) = 1 - \sum_{j=1}^{N} p_j = 0$$

$$\left(= \text{ for } \frac{1}{Np_j} = 1 \Rightarrow p_j = 1/N \right)$$

SOURCE CODING THEOREM

DUE TO SHANNON

FOR SOURCE WITH ENTROPY $H(X)$ WE CAN ENCODE SYMBOLS SO THAT ON THE AVERAGE THE CODE WORD LENGTH IS $H(X)$ BITS/SYMBOL. (BUT HOW?)

THUS, WE CAN COMPRESS SOURCE OUTPUT WITH ALPHABET SIZE 2^k (k BITS NORMALLY REQUIRED) DOWN TO $H(X)$.

OBVIOUSLY SINCE FOR $p_j = \frac{1}{2^k}$ $\quad j = 1, 2, \ldots, 2^k$ (EQUIPROBABLE OUTPUTS)

$$H(X) = \sum_{j=1}^{2^k} p_j \log_2 \frac{1}{p_j} = \log_2 \frac{1}{p_j} = k \text{ BITS/SYMBOL}$$

NO COMPRESSION POSSIBLE.

EXAMPLE: $N = 5$ SYMBOLS

$\Rightarrow$ NEED $\log_2 N = \log_2 5$ BITS/SYMBOL
NEED 3 BITS/SYMBOL

ASSUME $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$
$p_4 = \frac{1}{16}$, $p_5 = \frac{1}{16}$

$H(X) = 15/8$ BITS/SYMBOL

HOW DO WE ATTAIN THIS (ON THE AVERAGE)?

		ℓ
a_1 -	0	1
a_2 -	10	2
a_3 -	110	3
a_4 -	1110	4
a_5 -	1111	4

VARIABLE LENGTH CODE
ℓ = CODE WORD LENGTH
(RECALL MORSE CODE
USE "$\cdot$" FOR "E"
"$\cdot\cdot\cdot$" FOR "S")

$$E[L] = \sum_{j=1}^{5} \ell_j p_j = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} + 3 \cdot \tfrac{1}{8} + 4 \cdot \tfrac{1}{16} + 4 \cdot \tfrac{1}{16} = 15/8$$

CHOSE $l_j = \log_2 1/p_j$!!

EXPLANATION OF WHY SOURCE CODING THEOREM IS VALID

CONSIDER DMS FOR n OUTPUTS AND ASK

THE QUESTION: DO WE REALLY NEED TO CONSIDER TRANSMITTING ONE OF N^n POSSIBLE SEQUENCES (N = ALPHABET SIZE)?

ANSWER: NO, IF SOME ALPHABET OUTPUTS ARE MORE PROBABLE THAN OTHERS.

ASSUME n IS LARGE $\Rightarrow$ OUTPUT SEQUENCE HAS (np_1) a_1's, (np_2) a_2's, ..., (np_N) a_N's (LAW OF LARGE NUMBERS - RECALL COIN TOSS WITH $p = P(\text{HEADS}) = 3/4 \Rightarrow$ ... HTHHTHHHTHHHHT... $\approx 3/4\, n$ HEADS AND $\frac{1}{4} n$ TAILS)

A TYPICAL SEQUENCE $\underline{x}$ ($3/4\, n$ HEADS, $\frac{1}{4} n$ TAILS FOR EXAMPLE) HAS PROB. APPROXIMATELY

$$P(\underline{X} = \underline{x}) \approx \prod_{i=1}^{N} p_i^{np_i}$$

$$\left(\begin{array}{ll} \text{HHHTHHHT} \Rightarrow P = p^6 (1-p)^2 & \\ N=2 & \\ n=8 & = p^{8(3/4)} (1-p)^{8(1/4)} \end{array} \right)$$

$$= \prod_{i=1}^{N} 2^{\log_2 p_i^{np_i}} = \prod_{i=1}^{N} 2^{np_i \log_2 p_i}$$

$$= 2^{\sum_{i=1}^{N} n p_i \log_2 p_i} = 2^{-n(-\sum p_i \log_2 p_i)}$$

$$= 2^{-nH(X)}$$

SINCE ALL TYPICAL SEQUENCES ARE EQUALLY LIKELY AND ATYPICAL SEQUENCES HAVE PROB. $\to 0$ AS $n \to \infty$, THERE MUST BE

$$2^{nH(X)} \text{ TYPICAL SEQUENCES}$$

$\Rightarrow$ NEED $nH(X)$ BITS TO ENCODE THE TYPICAL n-LENGTH SEQUENCES $\Rightarrow$ NEED

$H(X)$ BITS / SYMBOL !

THIS IS AN <u>EXISTENCE</u> PROOF - DOESN'T SAY HOW TO DO IT. PROVIDES THE HINT THAT WE MUST CODE NOT INDIVIDUAL SYMBOLS BUT INSTEAD LONG BLOCKS OF SYMBOLS.

<u>HUFFMAN CODING</u>

ATTAINS $H(X)$ BITS/SYMBOL FOR n LENGTH SEQUENCE AS $n \to \infty$, USES VARIABLE LENGTH CODE WORDS

TYPES OF VARIABLE LENGTH CODES

Letter	Probability	Codewords			
		Code 1	Code 2	Code 3	Code 4
a_1	$p_1 = \frac{1}{2}$	1	1	0	00
a_2	$p_2 = \frac{1}{4}$	01	10	10	01
a_3	$p_3 = \frac{1}{8}$	001	100	110	10
a_4	$p_4 = \frac{1}{16}$	0001	1000	1110	11
a_5	$p_5 = \frac{1}{16}$	00001	10000	1111	110

CODE 4 IS NO GOOD SINCE

$110110 \rightarrow 110, 110 \Rightarrow a_5, a_5$

OR $11, 01, 10 \Rightarrow a_4, a_2, a_3$

NOT <u>UNIQUELY</u> <u>DECIPHERABLE</u> (DON'T WANT TO INSERT BLANKS, SLOWS TRANSMISSION RATE)

CODE 1, 2 ARE O.K. BUT

CODE 1 $\ldots \underbrace{001}_{a_3}\underbrace{001}_{a_3}\underbrace{01}_{a_2} \ldots$

MAKE DECISION

CODE 2 $\ldots \underbrace{100}_{a_3}1000100 \ldots$

MAKE DECISION - MUST WAIT ONE BIT TO SEE IF IT IS A 1

CODE 1 IS <u>INSTANTANEOUS</u> BUT NOT CODE 2

CODES 1, 3 ARE BOTH UNIQUELY DECIPHERABLE AND INSTANTANEOUS BUT

$E(L) = 31/16$ CODE 1

$30/16$ CODE 3

CODE 3 IS HUFFMAN CODE - LEAST AVERAGE CODEWORD LENGTH (AVERAGE CAN BE NONINTEGER NUMBER OF BITS) - FOR LONG STRINGS OF SYMBOLS IT REQUIRES LEAST NUMBER OF BITS

EXAMPLE

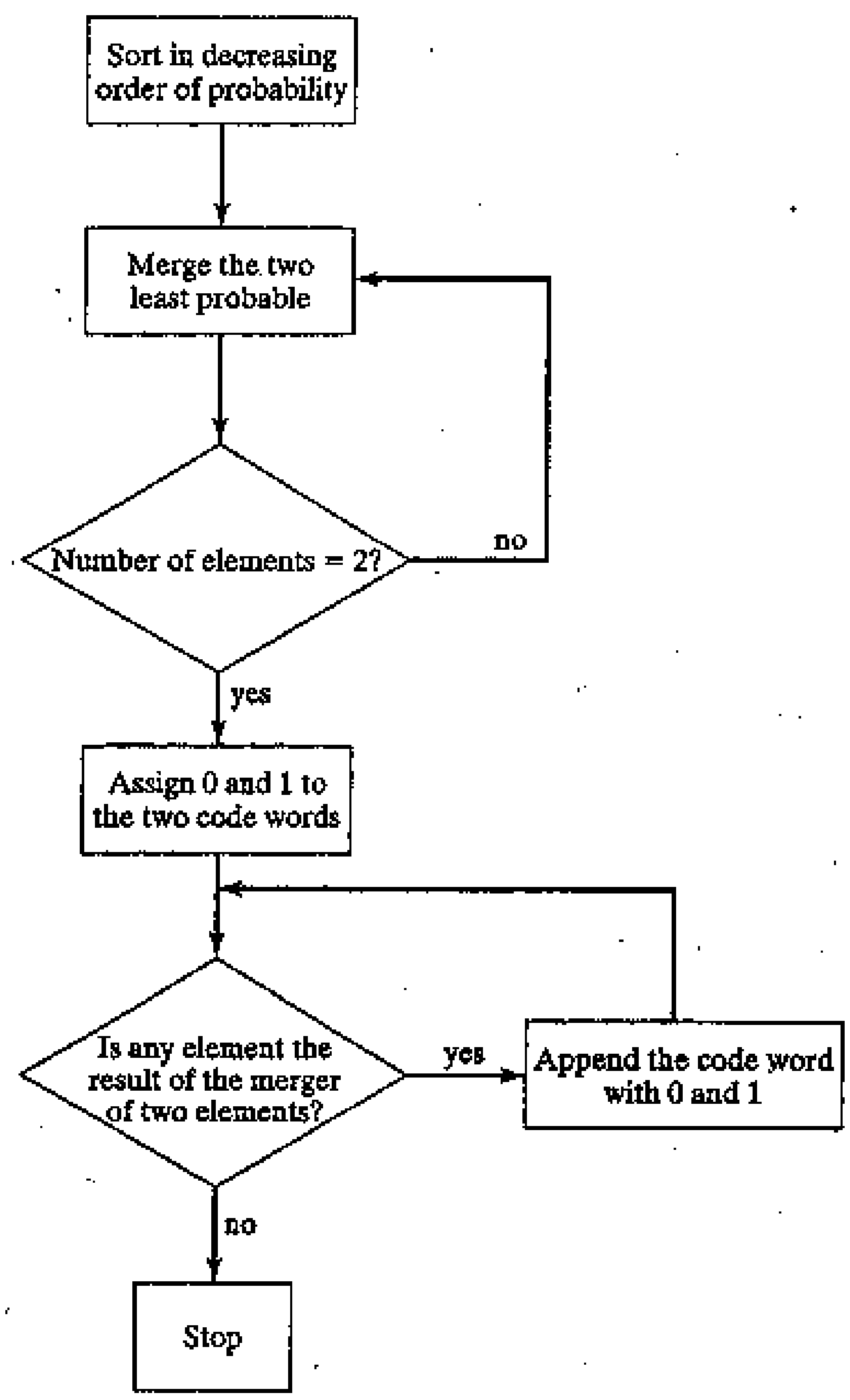
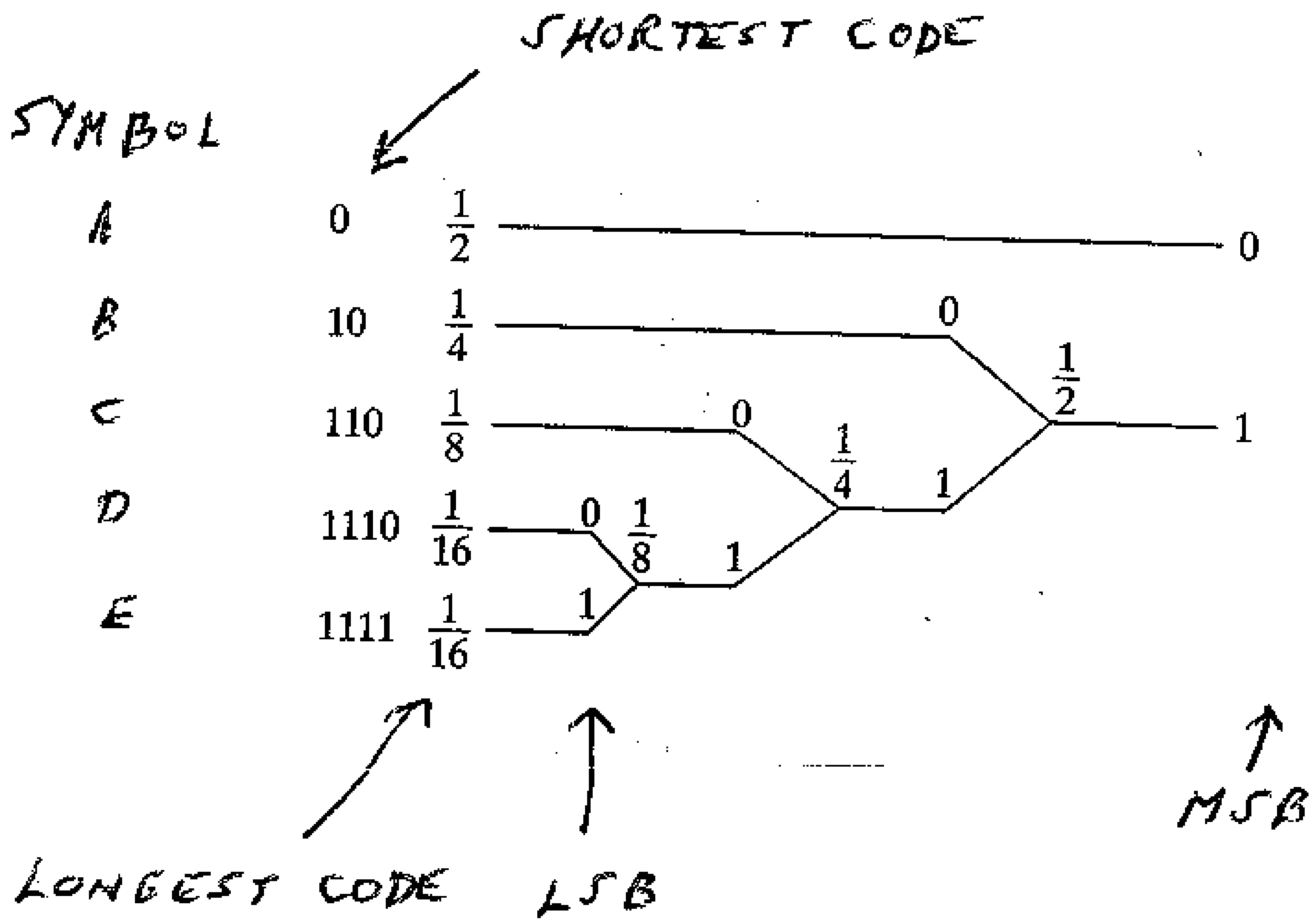


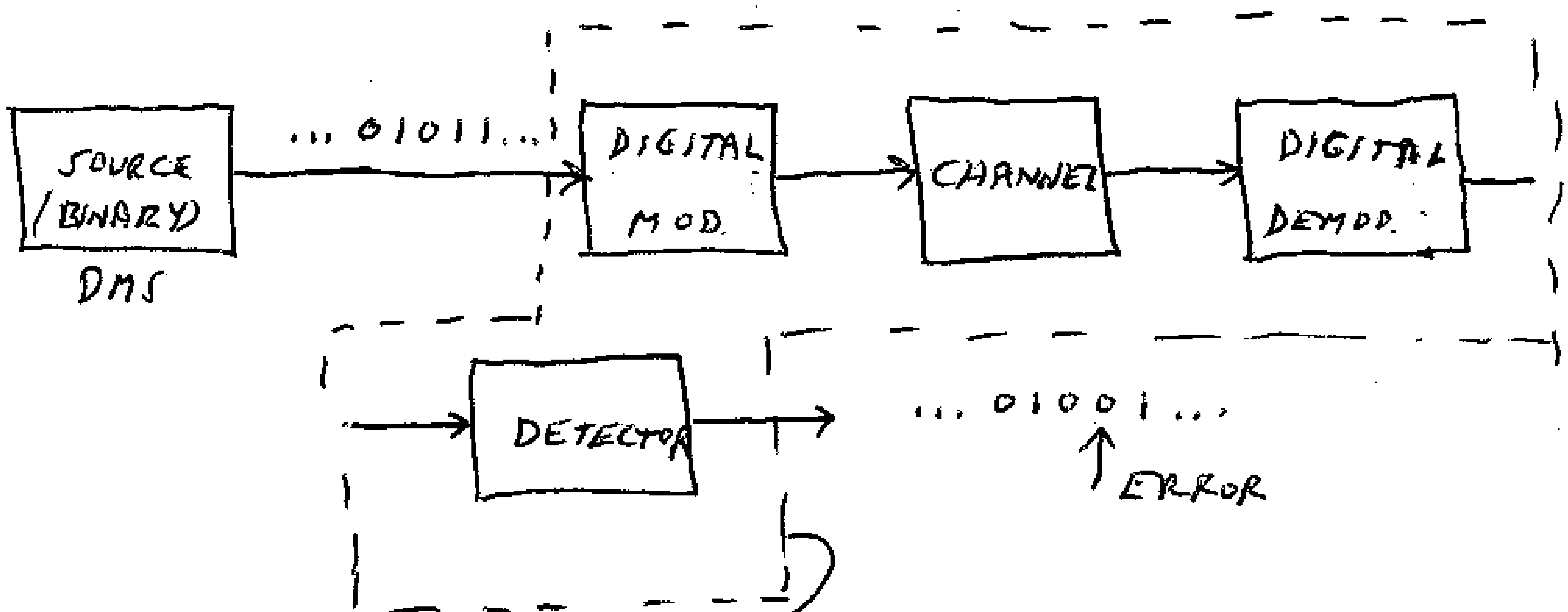
Figure 12.6 Huffman coding algorithm.

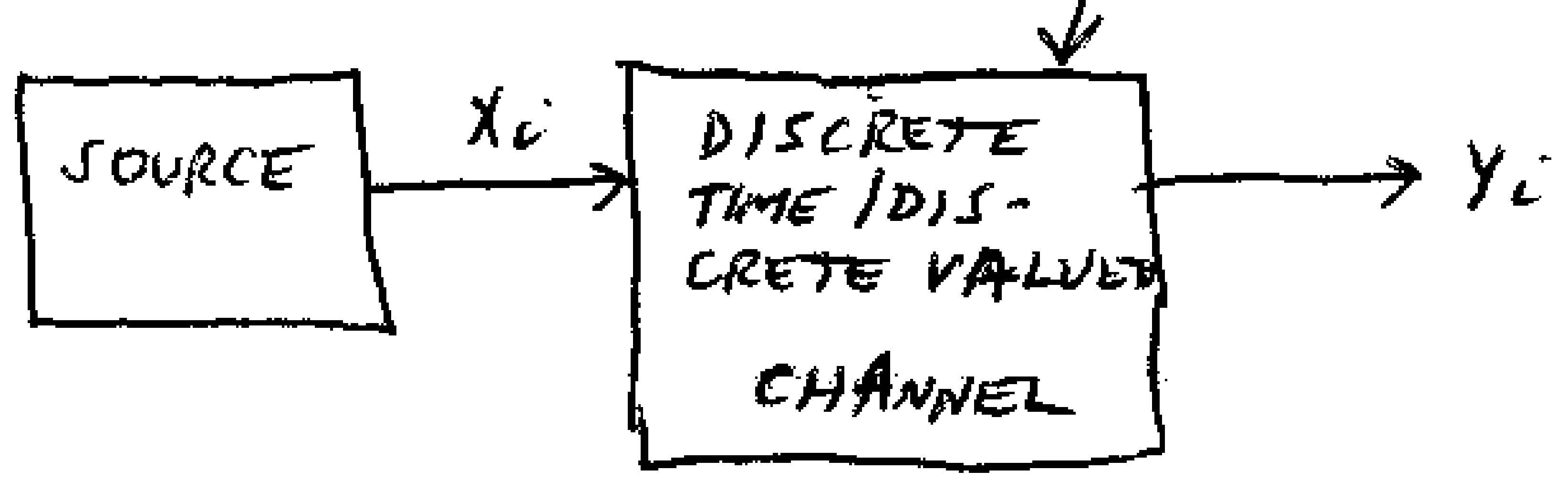
$p_A = \frac{1}{2},\ p_B = \frac{1}{4},\ p_C = \frac{1}{8},\ p_D = \frac{1}{16},\ p_E = \frac{1}{16}$



THIS WORKS PERFECTLY, i.e., $E(L) = H(X)$ SINCE $\log_2 \frac{1}{p}$ = INTEGER. IF THIS IS NOT THE CASE WE MUST ENCODE <u>BLOCKS</u> OF SYMBOLS TO HAVE $E(L)$ NEAR TO $H(X)$. SEE EXAMPLE 12.3.3.

MODELING OF COMM. CHANNELS





X_i = i^{th} INPUT BIT

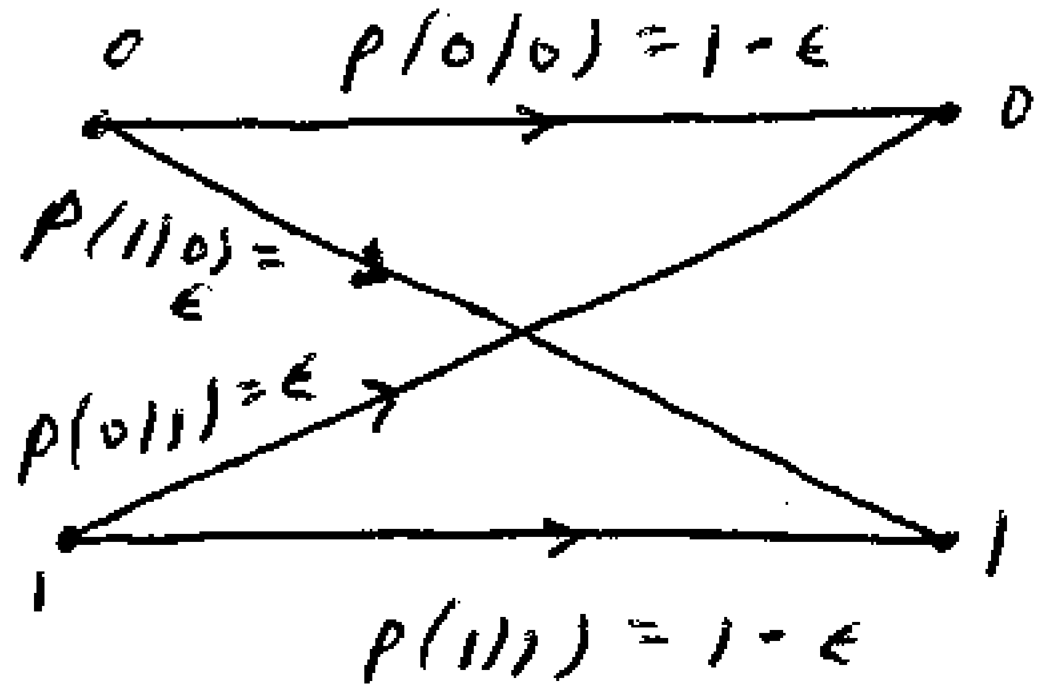
Y_i = i^{th} OUTPUT BIT

TO CHARACTERIZE DTDV CHANNEL JUST NEED $P(y_i | x_i)$ FOR ALL VALUES OF (x_i, y_i). HERE WE HAVE

$$x_i = 0, 1$$
$$y_i = 0, 1$$

$$\Rightarrow \underbrace{P(0|0),\ P(1|0)}_{\text{SUM TO } 1},\ \underbrace{P(0|1),\ P(1|1)}_{\text{SUM TO } 1}$$

EXAMPLE: BINARY SYMMETRIC CHANNEL



SYMMETRIC SINCE
$p(1|0) = p(0|1)$
$= \epsilon$

THIS MODELS A BPSK SYSTEM (ANTIPODAL SIGNALS) WITH

$$\epsilon = p(1|0) = p(0|1) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

NOTE: WE HAVE IMPLICITLY ASSUMED
THAT EFFECT OF CHANNEL ON
BIT i DOES NOT DEPEND ON BIT j.
(NO INTERSYMBOL INTERFERENCE
FOR EXAMPLE) ⇒

DISCRETE MEMORYLESS CHANNEL (DMC)

CHANNEL CAPACITY

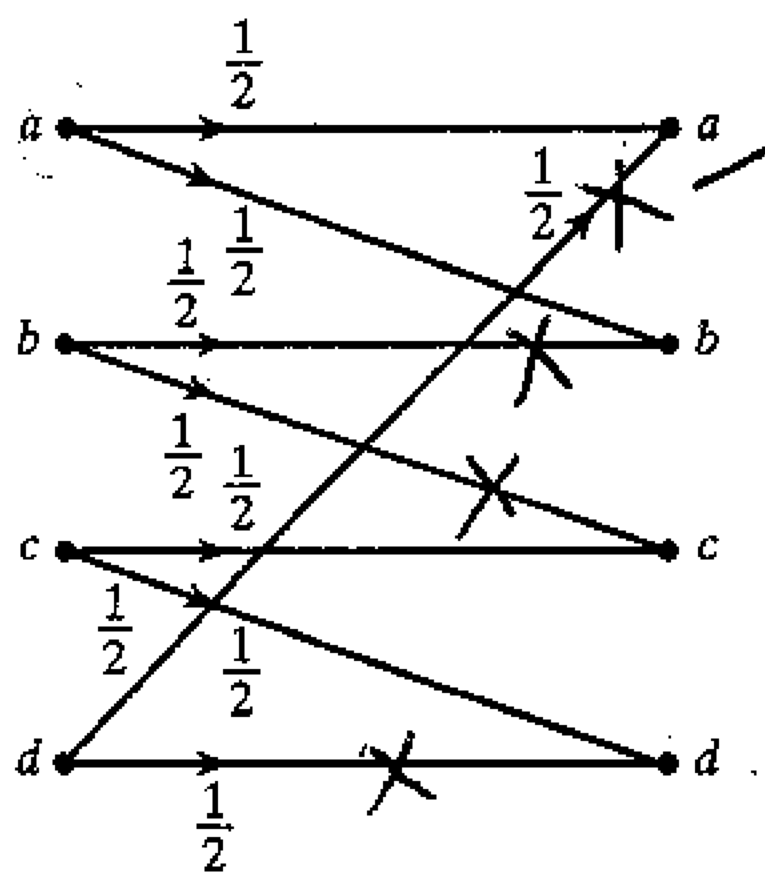
IN 1948, SHANNON PROVED
THAT ERRORLESS TRANSMISSION IS
POSSIBLE AS LONG AS TRANSMISSION
RATE R BPS IS LESS THAN THE
CHANNEL CAPACITY C.

THIS IS NOISY CHANNEL CODING THEOREM

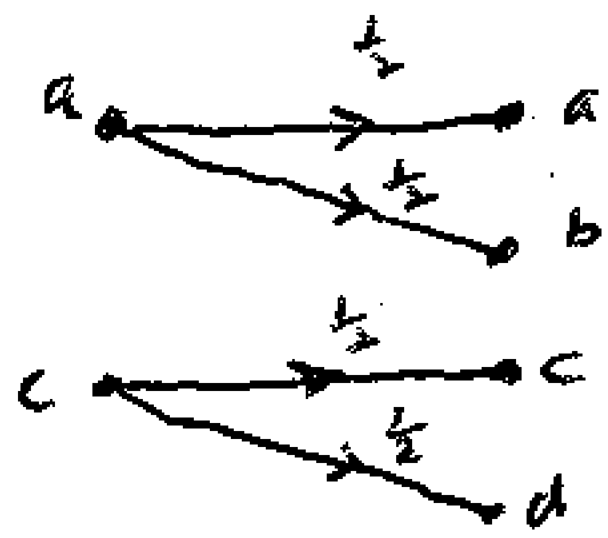
PREVIOUSLY THIS WAS THOUGHT TO BE IMPOSSIBLE!

C IS A <u>FUNDAMENTAL</u> <u>LIMIT</u> ON TRANSMISSION SPEED FOR $P_e \rightarrow 0$.

<u>KEY IDEA</u>: FOR ERRORLESS TRANSMISSION POSSIBLE OUTPUTS OF CHANNEL MUST BE <u>DISTINGUISHABLE</u>.



CAN'T USE b, d IF OUTPUTS ARE TO BE DISTINGUISHABLE

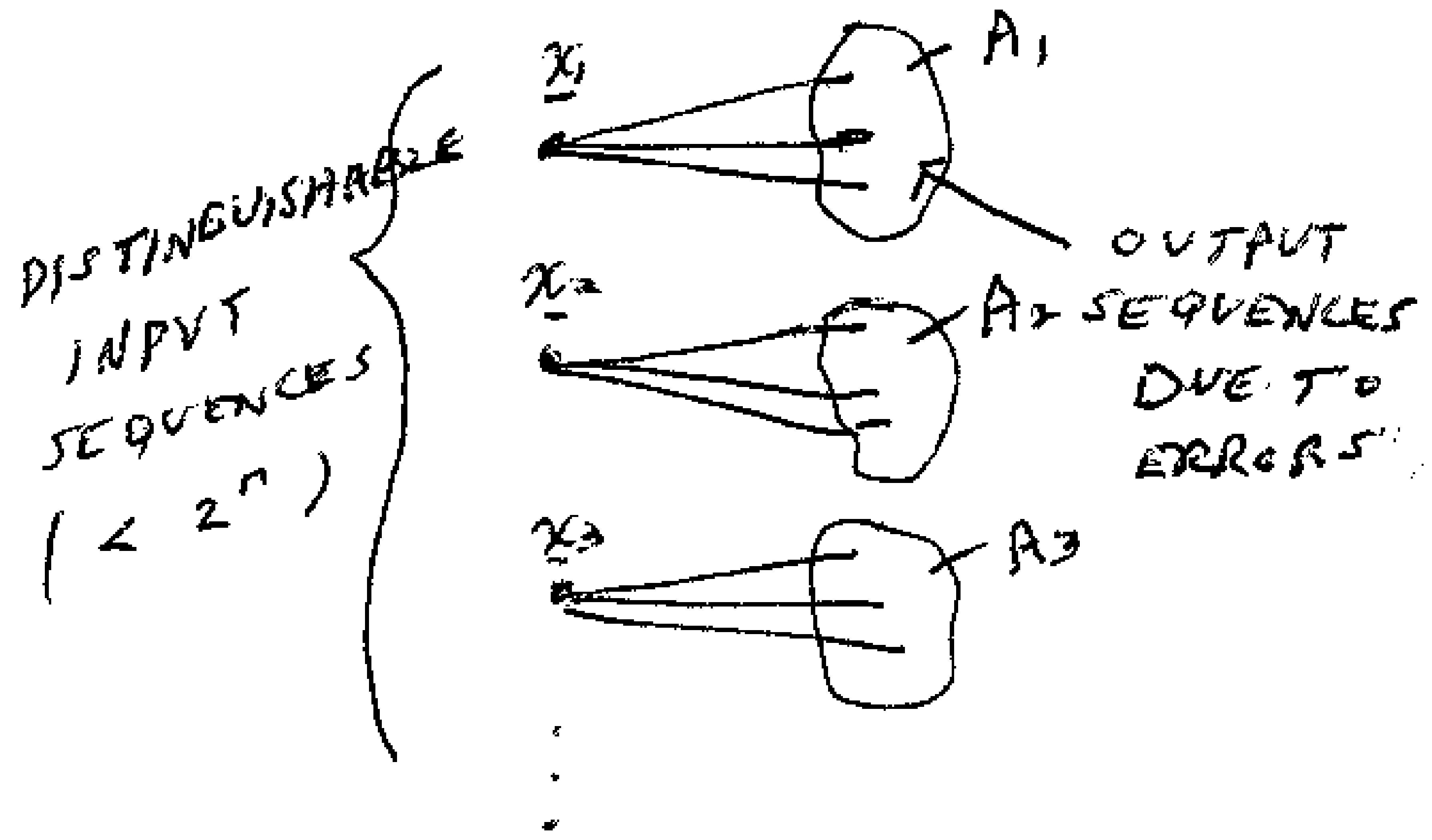


GIVEN OUTPUT WE CAN FIND INPUT WITHOUT ERROR

Figure 12.10 An example of a discrete channel.

USE ONLY INPUTS THAT RESULT IN DISJOINT OUTPUTS. FOR A BINARY DMC WE CAN ONLY DISTINGUISH A SUBSET OF THE 2^n OUTPUT SEQUENCES THAT ARE POSSIBLE FOR AN INPUT SEQUENCE OF n BITS.

HOW MANY SEQUENCES ARE IN EACH <u>DISJOINT</u> SUBSET A_i AS SHOWN?



$\underline{x}_i = (a_{1i}, a_{2i}, \ldots, a_{ni})$

$\underline{x}_i$ = GIVEN SEQUENCE OF n BITS

IF THIS PICTURE HOLDS, WE CAN DECODE WITH NO ERROR. (IF $\underline{x} \in A_i \Rightarrow \underline{x}_i$ TRANSMITTED)

FOR EACH $\underline{x}_i$ TRANSMITTED WE WILL HAVE $n\epsilon$ ERRORS (WITH PROB. 1 AS $n \to \infty$)

$\Rightarrow$ $n\epsilon$ BITS WILL BE DIFFERENT THAN $\underline{x}_i$

$\binom{n}{n\epsilon}$ POSSIBLE OUTPUTS FOR EACH $\underline{x}_i$

<u>EXAMPLE</u>: $n=4$, $\epsilon = 1/4 \Rightarrow n\epsilon = 1$ ERROR

TRANSMIT 0110 BUT RECEIVE

1110, 0010, 0100, 0111 = $\binom{4}{1}$ SEQUENCES

$\Rightarrow$ SIZE OF $A_i = \binom{n}{n\epsilon}$

SIZE OF $A_1 \cup A_2 \cup A_3 \cup \ldots$ = TOTAL POSSIBLE OUTPUT SEQUENCES

$$= 2^{nH(Y)} \qquad \text{(TYPICAL ONES)}$$

FINALLY NUMBER OF <u>DISTINGUISHABLE INPUT</u> SEQUENCES IS

$$M = \frac{2^{nH(Y)}}{\binom{n}{n\epsilon}} = \frac{\#\text{ TYPICAL OUTPUT SEQUENCES}}{\#\text{ SEQUENCES IN EACH } A_i \text{ SET}}$$

NOW $\binom{n}{n\epsilon} = \dfrac{n!}{(n-n\epsilon)!\,(n\epsilon)!}$

$$n! \approx \sqrt{2\pi n}\; n^n e^{-n} \qquad \text{(STIRLING APPROX.)}$$

$$\binom{n}{n\epsilon} \approx \frac{\sqrt{2\pi n}\; n^n e^{-n}}{\sqrt{2\pi n(1-\epsilon)}\,[n(1-\epsilon)]^{n(1-\epsilon)} e^{-n(1-\epsilon)}} \cdot \frac{1}{\sqrt{2\pi n\epsilon}\,[n\epsilon]^{n\epsilon} e^{-n\epsilon}}$$

$$= \frac{n^n}{[n(1-\epsilon)]^{n(1-\epsilon)} (n\epsilon)^{n\epsilon}} \quad \frac{1}{\sqrt{1-\epsilon}\,\sqrt{2\pi n\epsilon}}$$

$$\log_2 \binom{n}{n\epsilon} \approx n\log_2 n - n(1-\epsilon)\log_2 n - n(1-\epsilon)\log_2(1-\epsilon)$$
$$- n\epsilon \log_2 n\epsilon - \tfrac{1}{2}\log_2(1-\epsilon) - \tfrac{1}{2}\log 2\pi$$
$$- \tfrac{1}{2}\log_2 n\epsilon$$

TERMS ARE SMALL
OTHERS ARE $O(n)$

$$= +n\epsilon\log_2 n - n(1-\epsilon)\log_2(1-\epsilon) - n\epsilon\log_2 n\epsilon$$

$$= -n(1-\epsilon)\log_2(1-\epsilon) - n\epsilon\log_2\epsilon = nH_b(\epsilon)$$

WHERE $H_b(\epsilon) = -\epsilon \log_2 \epsilon - (1-\epsilon)\log_2(1-\epsilon)$

$$\Rightarrow \quad M = \frac{2^{nH(Y)}}{2^{nH_b(\epsilon)}} = 2^{n(H(Y)-H_b(\epsilon))}$$

THUS, WE CAN ONLY TRANSMIT AT MOST M MESSAGES, IE, $x_1, x_2, \ldots, x_M$ USING n BITS.

OR

$\log_2 M$ INFORMATION BITS PER n CHANNEL USES OR

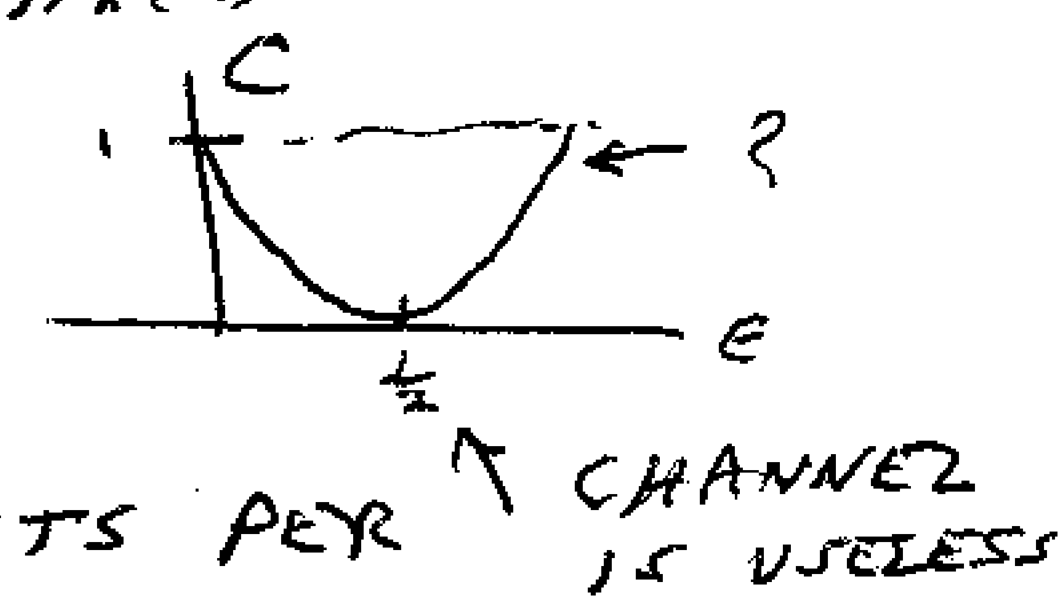
$$\frac{\log_2 M}{n} = R = H(Y) - H_b(\epsilon)$$

R ↑ TRANSMISSION RATE

FINALLY, SINCE WE MAY BE ABLE TO INCREASE R BY ADJUSTING $H(Y)$ (HAVE TO MODIFY $P(0)$, $P(1)$ AT INPUT, AND $H(Y) \le 1$ (BINARY CASE)

$$R_{max} = C = 1 - H_b(\epsilon) = \max_{P(0),P(1)} H(Y) - H_b(\epsilon)$$

C = CHANNEL CAPACITY



MEASURED IN INFORMATION BITS PER CHANNEL USE.