Paper Number(s): **E4.40**
SO20

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE
UNIVERSITY OF LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2002

MSc and EEE PART IV: M.Eng. and ACGI

INFORMATION THEORY

Monday, 22 April 10:00 am

There are SIX questions on this paper.

Answer FOUR questions.

Time allowed: 3:00 hours

Corrected Copy

Examiners responsible:

First Marker(s): Turner,L.F.

Second Marker(s): Barria,J.A.

Special instructions for invigilators: Grid to be provided.

Information for candidates: If you answer Question 3, please attach your grid securely to your main answer book.

1. Prove that for two probability sets

$$P_1, P_2, \ldots, P_M \qquad (P_1 + \ldots + P_M = 1)$$

and $$Q_1, Q_2, \ldots, Q_M \qquad (Q_1 + \ldots + P_M = 1)$$

$$\sum_{i=1}^{M} P_i \log \frac{1}{P_i} \leq \sum_{i=1}^{M} P_i \log \frac{1}{Q_i}$$

with equality if, and only if, $P_i = Q_i$ for all i.

Hence, or otherwise, prove that the entropy function H satisfies the condition

$H(P_1, \cdots, P_M) \leq \log M$
where $H(P_1, \cdots, P_M) = \sum_{i=1}^{M} P_i \log \frac{1}{P_i}$. [4]

If $H(x)$ denotes the entropy of a discrete random variable x, $H(y)$ is the entropy of a discrete random variable y, and $H(x/y)$ and $H(y/x)$ are the associated conditional entropies, prove that:

(i) $H(x,y) \leq H(x) + H(y)$
(ii) $H(xy) = H(x) + H(y/x) = H(y) + H(x/y)$
(iii) $H(x) - H(x/y) \geq 0$,
and explain the significance of the results. [9]

For the case in which the elements of the channel matrix are such that a) each of the rows of the matrix is a permutation of a basic set of numbers <u>and</u> b) each of the columns of the matrix is a permutation of a basic set of numbers, determine the relationship between the probabilities associated with the channel inputs if channel capacity is to be achieved.
[7]

2. Data is transmitted over a discrete memoryless noisy binary channel using pulses of amplitude $\pm V$ Volts. The channel is corrupted by zero-mean additive white Gaussian noise and attenuation can be neglected. If two decision thresholds at $\pm k.V$ are employed in the channel decision marking system and the received signal levels between the decision thresholds are considered to be 'ambiguous'; determine the capacity of the channel as a function of the various decision probabilities when

(i) $k < 1$

(ii) $k >> 1$. [14]

Explain the practical significance of the two approaches to decoding and discuss the associated advantages and disadvantages. [6]

If, in deriving your results, you employ any special arguments then these should be proved.

3. Figure 3.1 shows an image that is comprised of black and white pixels. The first-order statistics are fully representative of a class of images that are to be transmitted and it may be assumed that the pixels are statistically independent of each other.

The image is to be transmitted using a binary channel after it has been scanned on a line-by-line basis.

Determine the entropy of the image source and compare this with the image compression that can be achieved through the use of block encoding, with blocks of length 2. [10]

If the information was to be transmitted using your encoding scheme, and an error was to occur in the image in the seventh transmitted digit, what would be the output image obtained? Sketch the image on the grid provided. [7]

What general conclusion can you draw from your answer? [3]

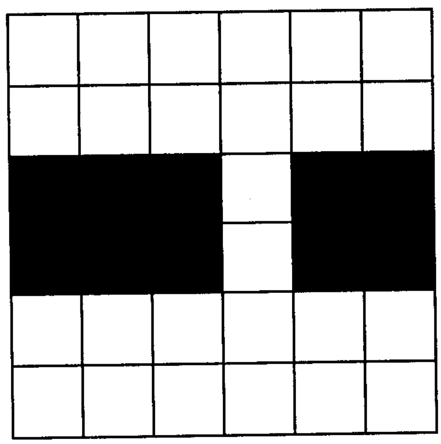


Figure 3.1

4. Suppose the pixels of an image can be represented by K binary digits. Suppose further that the image is to be transmitted over a binary communication channel using the well-known method of run-length coding.

If runs of lengths $l_1, l_2, l_3, \ldots, l_n$ can occur with associated probabilities $P_1, P_2, \ldots\ldots, P_n$, and successive runs are statistically independent, determine the data compression that can be achieved when the following encoding schemes are used:

a) At the start of each run a $(K + 1)$-binary digit word (a zero plus a K-bit word representing the amplitude of the pixels in the run) is used, and a binary 1 is used to represent each other pixel of the run. [6]

b) Optimum Run-Length encoding based on entropy considerations. [5]

c) Run-Length encoding in which a Shannon-Fano code is employed and a code-word of length L_i, is used to represent a run whose probability is P_i, $i=1,\ldots,n$. [5]

Compare the achievable compression ratios and draw conclusions as to the desirable properties of the run-lengths. [4]

Note: Compression Ratio, CR, is defined to be

$$CR = \frac{\text{Number of Data Digits Before Encoding}}{\text{Number of Data Digits After Encoding}}$$

5. Discuss the statement; "A restriction on the rate at which data pulses can be transmitted over a communication channel does not in itself place a fundamental restriction on the rate at which data can be transmitted over the channel". [2]

A binary symmetric channel has a cross-over probability of P. Determine **from first principles** the capacity of the channel. [6]

The channel is to be used to transmit information from an information source whose outputs are the integers 0,1 and 2, which occur with equal probabilities. Explain what has to be done in order to transmit data at, or close to, the channel capacity. [12]

6. Two modems whose constellation (signal-point) diagrams are shown in *Figure 6.1* are to be used to communicate data over a channel having a bandwidth of W Hz.

The noise which affects each dimension of the constellation identically and independently has an amplitude x whose probability density function $P(x)$ is

$$P(x) = \text{constant for } -q \leq x \leq q$$

$$= 0 \text{ elsewhere.}$$

Calculate and plot the capacity of the two modem systems as a function of q. [18]

What important practical conclusions can you draw from your plots? [2]

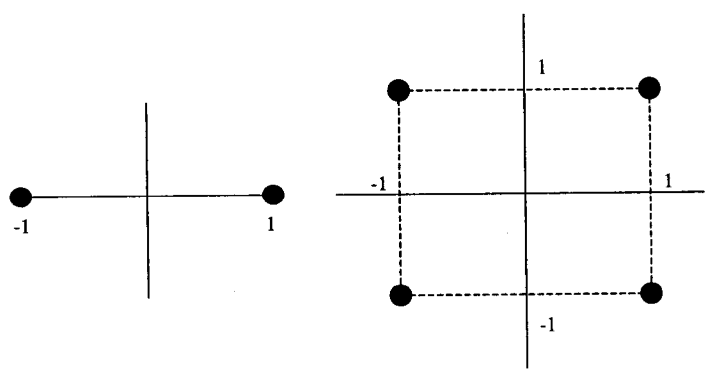
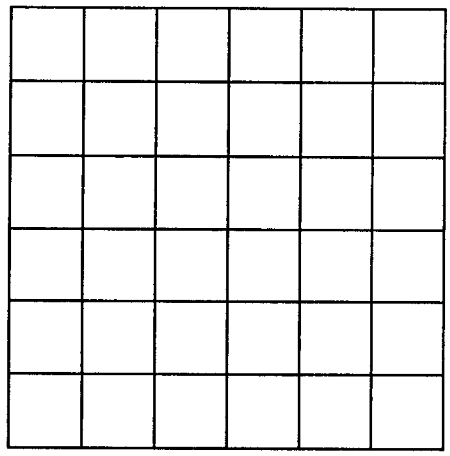


Figure 6.1



Grid

INFORMATION THEORY 2002

SOLUTIONS

Q1. Solution

Marks

Part 1 Now $\log x \le x - 1$, with equality at $x = 1$

$$\therefore \log \frac{Q_i}{P_i} \le \frac{Q_i}{P_i} - 1 \quad \text{, with equality iff } P_i = Q_i$$

and

$$P_i \log \frac{Q_i}{P_i} \le Q_i - P_i$$

$$\sum_{i=1}^{M} P_i \log \frac{Q_i}{P_i} \le \sum_{i=1}^{M} (Q_i - P_i) = 0$$

$$\therefore \sum_{i=1}^{M} P_i \log \frac{Q_i}{P_i} \le 0$$

with equality iff $P_i = Q_i$ for all i.

It thus follows directly that

$$\sum_{i=1}^{M} P_i \log \frac{1}{P_i} \le \sum_{i=1}^{M} P_i \log \frac{1}{Q_i} \quad (1)$$

with equality iff $P_i = Q_i$ for all i 4

Now let $Q_i = \frac{1}{M}$ for all i

$$\therefore \sum_{i=1}^{M} P_i \log 1/P_i \le \log M \ . \quad \text{QED}$$

Part 2

(i) $$H(XY) = -c \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i y_j) \log P(x_i y_j)$$

$$H(X) = -c \sum_{i=1}^{N} P(x_i) \log P(x_i)$$

$$H(Y) = -c \sum_{j=1}^{M} P(y_j) \log P(y_j)$$

$$\therefore H(X) + H(Y) = -c \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i y_j) \log P(x_i) P(y_j)$$

Let $P(x_i)P(y_j) = Q(x_i y_j)$ then it follows from (1) above that

$$-c \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i y_j) \log(P(x_i y_j)) \leq -c \sum_{i=1}^{N} \sum_{j=1}^{N} P(x_i y_j) \log Q(x_i y_j)$$

$$\therefore H(XY) \leq H(X) + H(Y) \quad (2)$$

with equality iff $P(x_i y_j) = P(x_i)P(y_j)$

Significance: The entropy of the joint distribution is equal to the sum of the entropies iff X & Y are statistically independent, ie on learning x we learn nothing of y, and vice versa.

4

(ii) $H(XY) \leq H(X) + H(Y/X) = H(Y) + H(X/Y)$

Now $H(Y/X) = -c \sum_{i=1}^{N} P(x_i) \sum_{j=1}^{M} p(y_j/x_i) \log P(y_j/x_i)$

$$= -c \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i y_j) \log P(y_j/x_i)$$

4

But $H(XY) = -c \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i y_j) \log P(x_i y_j)$

$$= -c \sum_{i=1}^{N} P(x_i) \log P(x_i) - c \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i y_j) \log P(y_j/x_i)$$

$$= H(X) + H(Y/X) \quad (3)$$

Similarly by using $P(x_i y_j) = P(y_j) P(x_i/y_j)$

we get $H(XY) = H(Y) + H(X/Y)$

Significance: uncertainty about X & Y is equal to uncertainty about Y plus the uncertainty about X, given that Y has been observed/learned.

(iii) $H(X/Y) \leq H(X)$.

This result follows directly from (2) and (3).

Significance: If X is dependent on Y, then having received Y we know something about X, ie our uncertainty is reduced as compared with our initial uncertainty, but if X & Y are statistically independent then $H(X/Y) = H(X)$, which is the most uncertain we can be about X.

Part 3

For the noisy channel

$$I = H(Y) - H(Y/X)$$

$$= H(Y) - \sum_{j=1}^{N} P(x_j) \sum_{i=1}^{K} P(y_i/x_j) \log \frac{1}{P(y_i/x_j)}$$

and $C = \underset{\max[P(X)]}{I}$

If the channel matrix is such that all rows are permutations of a set of numbers $a_1, \ldots, a_K$, then

$$\sum_{i=1}^{K} P(y_i/x_j) \log 1/P(y_i/x_j) = \sum_{i=1}^{K} a_i \log 1/a_i = \text{constant}$$

for all rows.

Thus $H(Y/X) = \sum_{j=1}^{N} P(x_j) \sum_{i=1}^{K} a_i \log(1/a_i) = \sum_{i=1}^{K} a_i \log 1/a_i$

Hence $C = I_{\max[P(X)]} = H(Y)_{\max P(X)} + \sum_{i=1}^{K} a_i \log a_i$

Now if all columns are permutations of a set of numbers $b_1, \dots, b_N$ (may be same set as set a_i's, or subset thereof)

then $P(y_i) = \sum_{j=1}^{N} P(x_j) P(y_i/x_j)$

$$= P(x_1) b_1 + \dots + P(x_N) b_N$$

7

If we put $P(x_1) = P(x_2) = \dots = P(x_N) = 1/N$ then

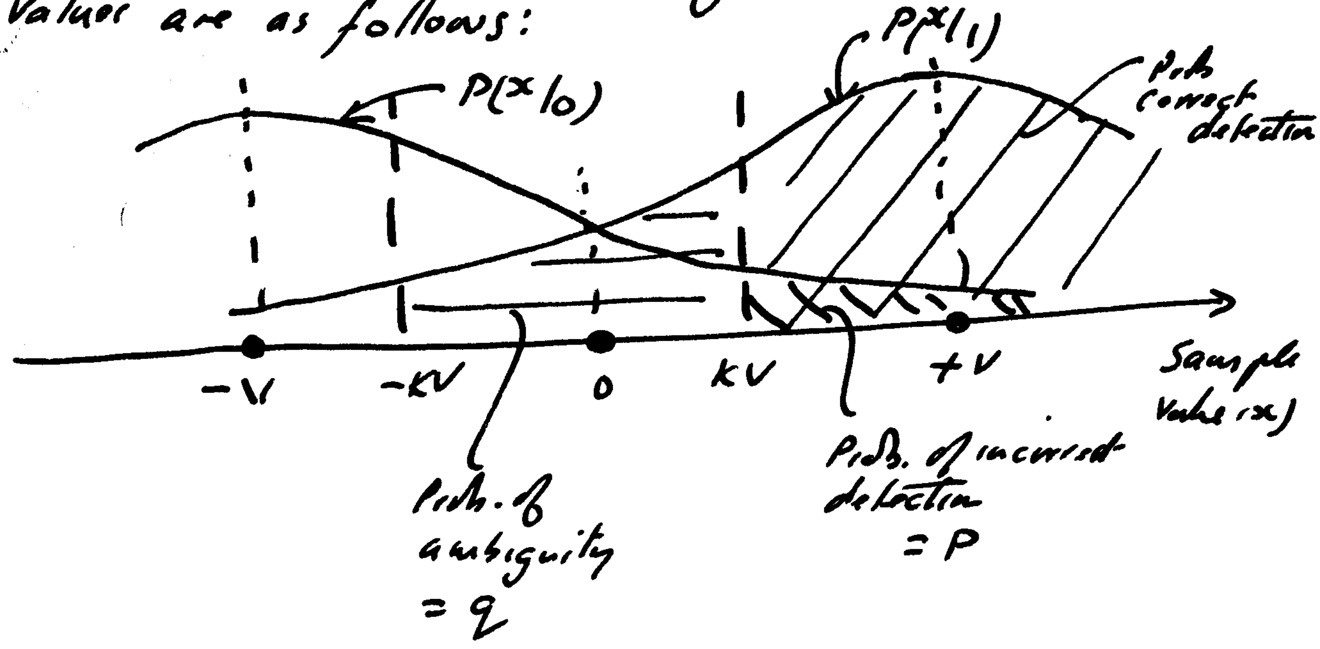
$P(y_i) = 1/K$ for all i

and hence $\cancel{P(y_i)}$ $H(Y)$ is maximized at $\log K$ and hence capacity is achieved.

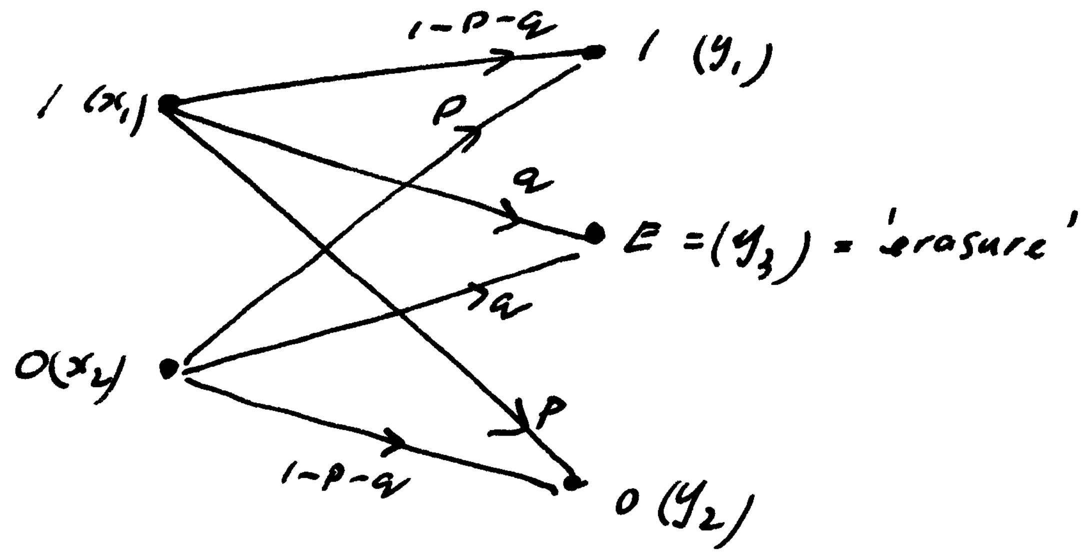
The relationship is that for ~~equiprob~~ equally probable input symbols we get equally probable outputs — hence capacity is achieved.

Q2 Solution

Assume that attenuation in cable is neglected. In this case the conditional pdf's relating to the received sample values are as follows:



Thus the channel transition diagram is as follows:



and the associated matrix is

$$\begin{array}{c|ccc} & y_1 & y_2 & y_3 \text{ — outputs} \\ \hline x_1 & 1-p-q & p & q \\ x_2 & p & 1-p-q & q \end{array}$$

input

Thus the channel matrix is 'uniform' from the input

$\therefore \; I = H(Y) + (1-q-p)\log(1-q-p) + q\log q + p\log p$

← expect proof of this

Now although the channel appears to be symmetrical it is NOT 'uniform' from the output

But it is easy to show that in this case $\underset{P(X)}{I_{\max}} = C$ 14

is obtained when $P(x_1) = P(x_2) = \frac{1}{2}$

← expect this to be proved.

This condition makes $H(Y)$ maximum.

With these probabilities

$$P(y_1) = P(x_1)(1-p-q) + P(x_2)p = (1-q)\tfrac{1}{2}$$

$$P(y_2) = q\,P(x_1) + q\,P(x_2) = q$$

$$P(y_3) = P(x_1)p + P(x_3)(1-p-q) = (1-q)^{1/2}$$

and hence the capacity is

$$C = \overbrace{-q\log q - \tfrac{1}{2}(1-q)\log\tfrac{1}{2}(1-q) - \tfrac{1}{2}(1-q)\log\tfrac{1}{2}(1-q)}^{H(Y)}$$

$$+ (1-p-q)\log(1-p-q) + q\log q + p\log p$$

$$= (1-q)\left[1-\log_2(1-q)\right] + (1-p-q)\left[\log_2(1-p-q)\right]$$

$$+ p\log_2 p \quad \text{bits/input} \qquad \text{Eqn (1)}$$

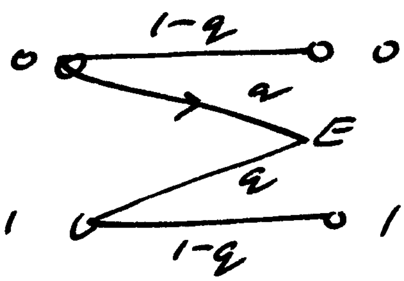
Note: (i) If $q=0$ then this reduces to

$1 + p\log p + (1-p)\log(1-p)$ which is the capacity of a B.S.C

(ii) the capacity in (1) is changed from that of BSC since we are saying if signals are somewhat uncertain (are in ambiguous region) then this will be indicated (soft decision) 3

If K becomes large, then $p \to 0$ ie we are attempting to rule out errors, at the expense of an increasing number of 'ambiguous'/erasure outputs. 3

The transition diagram becomes



Q3 Solution

Part 1

$$P_{white} = P_1 = \frac{26}{36} = 0.72$$

$$P_{black} = P_2 = \frac{10}{36} = 0.28$$

these are to be taken as representative probabilities.

The entropy of the source is

$$H = -\left\{0.72\log_2 0.72 + 0.28\log_2 0.28\right\}$$
$$= 0.855 \text{ bits/pixel}$$

Now if we encode using blocks of length 2, we have

X_1 (WW)	$P_1P_1 = 0.52$	0	
X_2 (WB)	$P_1P_2 = 0.20$	1 0	
X_3 (BW)	$P_2P_1 = 0.20$	1 1 0	
X_4 (BB)	$P_2P_2 = 0.08$	1 1 1	10

Therefore: the average codeword length

$$= 0.52 \times 1 + 2 \times 0.20 + 3 \times 0.20 + 3 \times .08$$
$$= 0.88 \text{ binary digits}$$

Using the block encoding the image is represented by the sequence

$$X_1 X_1 X_1 X_1 X_1 X_1 X_4 X_3 X_4 X_4 X_3 X_4 X_1 X_1 X_1 X_1 X_1 X_1$$

(sequence 1)

and the corresponding sequence of digits for transmission is 000000011111011111111011000000

= 30 digits

The compression ratio is thus $\frac{36}{30} = 1.2$.

Note that $\frac{30}{36} = 0.833$ bits/pixel. This is less than the entropy and I explain them to explain why.

(The image is not large enough for all black/white, white/black possibilities to occur — note that $X_2 = WB$ does not occur.

Part 2

Suppose an error occurs in the 7th digit as indicated in the question. Then the received binary sequence is

e

00000000111101111111101110000000

and this will be decoded as

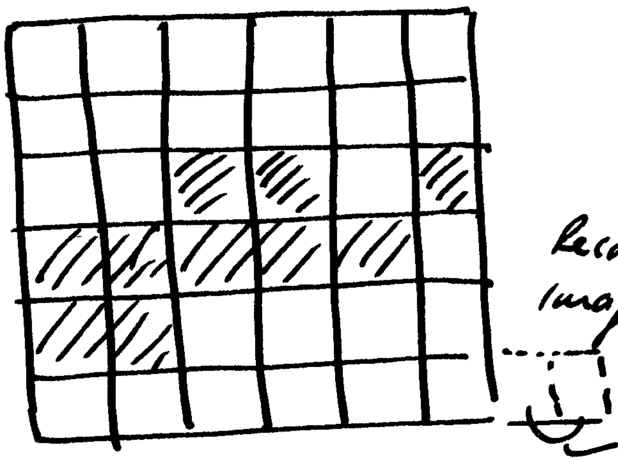
$X_1 X_1 X_1 X_1 X_1 X_1 X_1 X_4 X_2 X_4 X_4 X_3 X_4 X_1 X_1 X_1 X_1 X_1 X_1$

note symbol recovery from this point on (Sequence 2)

I expect them to note that after the error the sequence 'recovers' ie code is self synchronous in respect of letters/symbols.

But!

10/21



I expect the following points to be made:

(1) the major point to note is that although symbol synchronisation is recovered, spatial positioning is lost and the image would be of little value.

(2) This type of effect is not unusual since we know that removal of redundancy makes messages more sensitive to errors. 3

(3) We have to have low error rates for image transmission and 'end-of-line' signals are usually employed.

Q4 Solution

11/21

Suppose the source generates a long sequence of N runs, containing runs of the various lengths $l_1, l_2, \dots, l_n$.

Thus the number of runs of length l_i will be NP_i, $i = 1, \dots, n$

Thus the number of symbols generated will be

$$S = \sum_{i=1}^{N} NP_i l_i = N.\,l_{ave}, \text{ where } l_{ave} = \text{average run length}$$

Thus the number of bits generated is Q, where

$$Q = KN.\sum_{i=1}^{n} P_i l_i$$ — this is data generated in <u>uncompressed</u> form.

With coding scheme A

The number of binary digits generated by the encoding with respect to the N runs is

$$R = S + KN = N.\sum_{i=1}^{n} P_i l_i + KN$$

6

and the compression ratio is

$$CR_1 = \frac{KN\sum_{i=1}^{n} P_i l_i}{KN + \sum_{i=1}^{n} P_i l_i} = \frac{K\sum_{i=1}^{N} P_i l_i}{K + \sum_{i=1}^{N} P_i l_i} \quad (1)$$

With coding scheme B

With optimum run-length encoding the average number of binary digits use to encode a run-length is

$$H = -\sum_{i=}^{n} P_i \log_2 P_i \quad \text{bits (binary digits)}$$

Thus the number of digits used to represent the N runs is

$$\underbrace{KN}_{\text{to indicate 'amplitude'}} + N\sum_{i=1}^{n} P_i \log_2 1/P_i$$

and the compression ratio in this case is

$$CR_2 = \frac{KN\sum_{i=1}^{n} P_i l_i}{KN + N\sum_{i=1}^{N} P_i \log_2 1/P_i} = \frac{K\sum_{i=1}^{n} P_i l_i}{K + \sum P_i \log_2 1/P_i} \quad (2)$$

5

With coding scheme C

The average code-word length is $\sum_{i=1}^{n} P_i L_i$ (per run)

and hence the average number of digits used is

$$KN + N\sum_{i=1}^{n} P_i L_i$$

and the compression ratio is

$$CR_3 = \frac{KN \sum_{i=1}^{n} p_i l_i}{KN + N \sum_{i=1}^{n} p_i L_i}$$

$$= \frac{K \sum_{i=1}^{n} p_i l_i}{K + \sum_{i=1}^{n} p_i L_i} \quad (3)$$

5

I want the following points to be made in respect of discussion and comparision

(i) We note from (1) that

$$CR_1 = \frac{K}{1 + K / \sum_{i=1}^{n} p_i L_i}$$

So the longer the average run-length the better (ie higher the compression ratio)

(ii) Limiting value of $CR_1 = K$

(iii) The same general conclusion can be drawn with respect to (2) since $\sum p_i \log 1/p_i$ is fixed and is in a sense independent of the run length, so we would like long runs so that

$$\frac{\sum p_i \log 1/p_i}{\sum p_i l_i} \to 0 \quad \text{and } CR \to \text{large.}$$

4

ie the entropy / average run length becomes small.

iv, The Shannon coding technique ~~w~~ is reasonably efficient and hence

$$\sum_i p_i L_i \simeq \sum p_i \log_2 1/p_i .$$

and hence we would expect (3) to be ~~[illegible]~~ slightly worse than (2), but not too different

15/21

Q5 Solution

Part 1

The statement in itself does not constitute a restriction since we can send as many data bits/pulse as we like, by increasing the size of the symbol set. But if we are power constrained then the pulses get 'closer together' and hence if noise is present then the system error probability increases. 2

Part 2

$$C = I_{\max P[X]} = \max_{P[X]} \left\{ H(X) - H(X/Y) \right\} \text{ or } \max_{P(X)} \left\{ H(Y) - H(Y/X) \right\}$$

I will accept a solution in which I is expressed as a function of $P(x_1)$, $P(x_2)$ (the input probabilities) and P (error probability), and the function is then maximised by obtaining $\frac{\partial I}{\partial P(x_1)} = 0$; $\frac{\partial I}{\partial P(x_2)} = 0$ with a solution being obtained for $P(x_1)$, $P(x_2)$.

The result is $P(x_1) = P(x_2) = 1/2$

This is a difficult approach and I will accept the following alternative :

16/21

$$I = H(Y) - H(Y/X)$$

and the channel matrix is

$$\begin{array}{c|cc} & y_1 & y_2 \\ \hline x_1 & (1-p) & p \\ x_2 & p & (1-p) \end{array}$$

Now $H(Y/X) = \sum_{j=1}^{2} P(x_j)\left[p \log p + (1-p)\log(1-p)\right]$

$$= p \log p + (1-p)\log(1-p)$$

and $H(Y) = -\left\{ P(y_1)\log P(y_1) + P(y_2)\log P(y_2) \right\}$

But $P(y_1) = P(x_1)\,P(y_1/x_1) + P(x_2)\,P(y_1/x_2)$

$$= P(x_1)(1-p) + p.P(x_2)$$

$P(y_2) = P(x_1)\,P(y_2/x_1) + P(x_2)\,P(y_2/x_2)$ 6

$$= P(x_1).p + (1-p)P(x_2)$$

Now if $P(y_1) = P(y_2)$ then $H(Y)$ is maximised

hence if we set $P(x_1) = P(x_2) = 1/2$

then $P(y_1) = P(y_2)$ and hence

$$C = 1 + p\log_2 p + (1-p)\log_2(1-p)$$

Part 3

In order to transmit at capacity it is necessary that the input binary digits to the channel be used with equal probability.

Now the source is

		$P(x_i)$
X_1	0	1/3
X_2	1	1/3
X_3	2	1/3.

Consider the following coding in which X's are mapped into binary digits.

Case I

X_1 00

X_2 10 $\quad \therefore P(0) = P(1) = 1/2$

X_3 11 $\quad \therefore$ looks to be perfect

But the entropy of the source $= \log_2 3 = 1.59$ bits/symbol

We are using 2 bits/symbol $\therefore$ wasteful

Case II

Using a Shannon/Fano code we obtain

X_1 0 $\quad P(0) = 0.4$

X_2 10 $\quad P(1) = 0.6$

X_3 11 $\quad \therefore$ Poor match to required $P(0) = P(1) = 0.5$

With this code average codeword length $= 5/3 = 1.67$ bits/symbol.

Case III Apply SF code to extended symbol set

Prob 1/9	$x_1 x_1$	0 0 0
	$x_1 x_2$	0 0 1
	$x_1 x_3$	0 1 0
	$x_2 x_1$	0 1 1
	$x_2 x_2$	1 0 0
	$x_2 x_3$	1 0 1
	$x_3 x_1$	1 1 0
	$x_3 x_2$	1 1 1 0
1/9	$x_3 x_3$	1 1 1 1

$\therefore P(0) = 0.45$, $P(1) = 0.55$ } quite close to 0.5

Average codeword length

$= \frac{29}{9}$ / pair of symbols

$= \frac{29}{18} = 1.61$ bits/symbol.

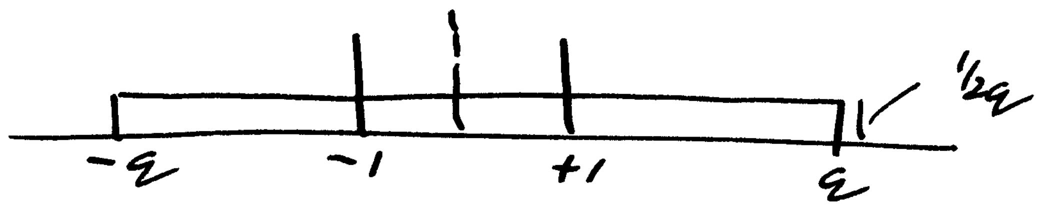
If we repeat by coding longer blocks of source symbols then $P(0) \rightarrow 0.5$, $P(1) \rightarrow 0.5$ and $\bar{n} \rightarrow H(X)$

The final requirement is for a proper channel coder which is inserted between the source coder and the channel to achieve reliable communication — care needed to ensure that $P(0) = P(1) = 0.5$ requirement is retained at coder output.

12

Q6 Solution

Consider the given noise pdf



From this we see that the probability of the noise taking the signal level across a decision boundary in each dimension is $\frac{1}{2q}(q-1) = P$.

Thus for the two systems we have channel matrices as follows:

Binary

$$\begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{bmatrix} (1-P) & P \\ P & (1-P) \end{bmatrix} \end{array}$$

4-phase

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} (1-P)^2 & P(1-P) & P^2 & P(1-P) \\ P(1-P) & (1-P)^2 & P(1-P) & P^2 \\ P^2 & P(1-P) & (1-P)^2 & P(1-P) \\ P(1-P) & P^2 & P(1-P) & (1-P)^2 \end{bmatrix} \end{array}$$

where $P = \frac{1}{2q}(q-1)$

6

Both channels are doubly symmetric and hence the respective capacities are

20/21

Binary $C = 2W\left[1 + p\log_2 p + (1-p)\log_2(1-p)\right]$ (1)

bits/sec.

4-phase

$$C = 2W\left\{2 + p^2\log_2 p^2 + 2(1-p)p\log_2(1-p) + (1-p)^2\log_2(1-p)^2\right]$$ (2)

bits/sec.

4

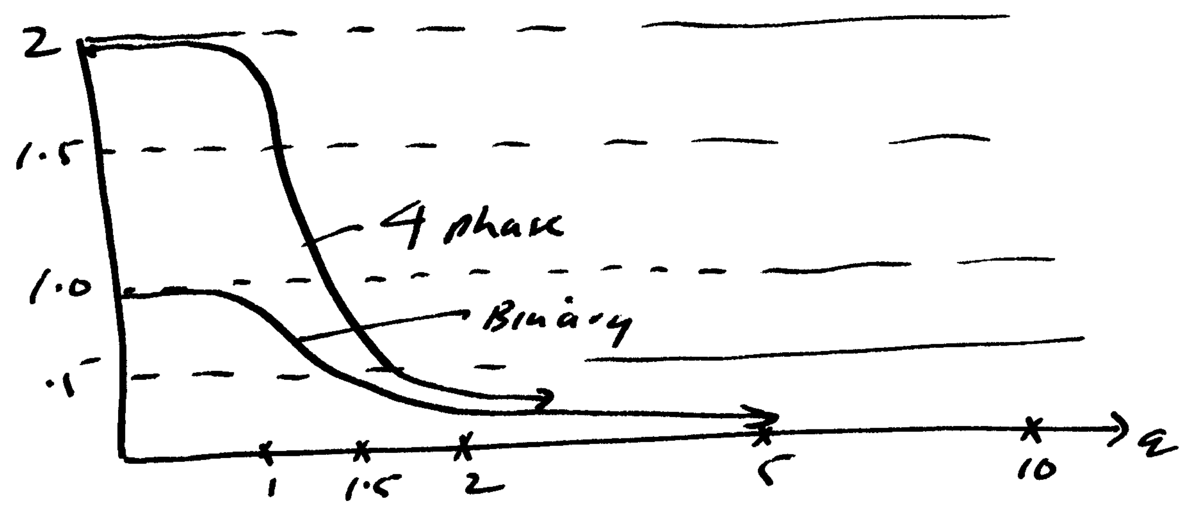
Now we can evaluate the capacity expressions as function of q.

bits/symbol

q	p	p^2	$(1-p)$	$p(1-p)$	$(1-p)^2$	C_{binary}	C_{4phase}
≤ 1	0	0	1	0	·1	1	2
1·5	·167	·028	·83	·139	·694	·35	·69
2	·25	·063	·75	·188	·563	·19	·38
5	·4	·16	·60	·240	·360	·03	·07
10	·45	·20	·55	·248	·303	·01	·02
100	·495	·25	·5	·25	·25	0	0

$p = \frac{1}{2q}(q-1)$

and on plotting we obtain



The main points are

(i) at high S/N we get main from using more symbol points

(ii) at low SNR although $C_{4P} > C_B$ the difference becomes small and gets less

(iii) at very low S/N $C \to 0$

10