Paper Number(s): **E3.04**
SC3
ISE3.15

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE
UNIVERSITY OF LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2001

MSc and EEE/ISE PART III/IV: M.Eng., B.Eng. and ACGI

INFORMATION THEORY

Monday, 30 April 10:00 am

There are SIX questions on this paper.

Answer FOUR questions.

Corrected Copy
Q6 (i) (10:00)
Q2 (10:50)

Time allowed: 3:00 hours

Examiners: Turner,L.F. and Barria,J.A.

Special instructions for invigilators: None

Information for candidates: None

1 Prove the optimality of the Huffman encoding scheme and discuss its advantages and disadvantages when used as a practical method of source coding. Explain how you would implement the method in practice.

2 An analogue information signal is sampled and its sample values, x, have a probability density function, $F(x)$, given by

$$
\begin{aligned}
F(x) &= \tfrac{1}{4}\left(1 - \tfrac{x}{4}\right); && 0 \leq x \leq 4 \\
&= \tfrac{1}{4}\left(1 + \tfrac{x}{4}\right); && -4 \leq x \leq 0 \\
&= 0 \quad ; && \text{elsewhere}
\end{aligned}
$$

If the sampled process is quantized using a uniform 8-level quantizer, what is the entropy of the quantizer output?

After quantization, the signal is encoded using a 3-bit fixed-codeword-length encoder that uses its most significant digit as a sign digit and its other two digits as binary-coded-decimal digits with the digit pair 1, 1 being used for the largest magnitude sample. The output from the encoder is then transmitted over a binary communication channel that is corrupted by zero-mean additive white Gaussian noise of variance σ^2.

Derive an expression for the maximum rate in bits/binary digit, at which information can be transmitted error-free over the channel.

3

If, in connection with the quantizer, a simple 8-bit binary-coded decimal converter had been used, how would this have affected your results?

What is the maximum error-free rate of communication if the quantized samples are encoded using a Huffman encoder prior to transmission over the channel? Comment on the significance of your results.

3 An information source selects its outputs, one at a time, from a set of N symbols. The source has memory that extends over R successive outputs generated by the source.

If blocks, S_i, consisting of L successive source symbols are found to have associated probabilities $P(S_i)$, $i = 1, \ldots, N^L$, prove that provided $L > R$,

$$\widetilde{H} = -\sum_{i=1}^{N^L} P(S_i) \log P(S_i) = (L-R)\, H(x) + \delta,$$

where $H(x)$ is the entropy of the source, and δ is a positive constant that is independent of L.

Explain carefully what the expression $\widetilde{H}$ means and, further, explain the physical significance of the result you have proved.

An analogue signal is sampled and its sampled values, x, have a probability density function $P(x)$. If the sampled values are quantized using a uniform quantizer of step-size $\triangle$, derive an expression for the entropy of the quantizer output if successive samples are statistically independent. Examine what happens to the entropy as $\triangle \to 0$, and comment on the physical significance of the result.

Prove that the entropy function $H(P_1, \ldots, P_M) = -\sum_{i=1}^{M} P_i \log P_i$

satisfies the condition

$$H(P_1, \ldots, P_M) \leq \log M$$

Discuss and interpret the result.

4 Explain what you understand by the 'asymptotic equipartition theorem', and illustrate your answer with a simple example.

Explain what you understand by Shannon's random coding argument. Why is the concept so important in Information Theory?

Explain in detail why the random coding argument cannot be used to arrive at a practical channel coding scheme.

State Shannon's capacity theorem for noisy communication channels and explain through a carefully selected example how, in principle, block coding can be used to achieve the results promised by the theorem.

A 4-phase modulation system transmits information by selecting from the set of signal points shown in Figure 1. If the transmission system suffers from zero-mean additive white Gaussian noise, obtain the channel matrix in terms of notional transition probabilities, and determine the capacity as a function of these probabilities. What is the capacity at high signal-to-noise ratios?

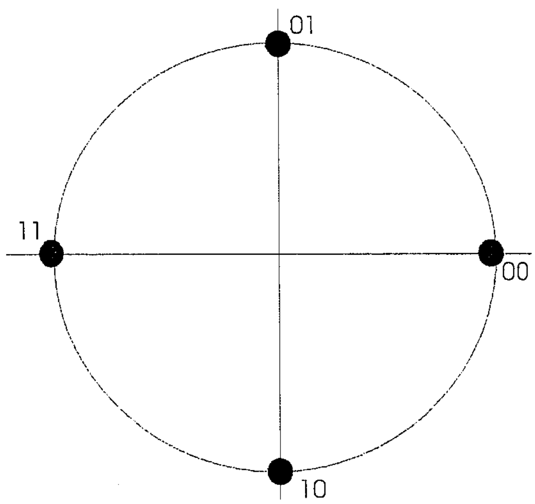


Figure 1

5 Binary information digits are to be transmitted over an associated binary communication channel. The communication system employs a simple code in which each information digit is sent three times and a majority logic decision maker is used at the receiver, If +v and -v volt pulses are used to transmit **ones** and **zeros** respectively and the channel is corrupted by zero mean additive white Gaussian noise of variance σ^2, derive from first principle an expression for the capacity of the overall binary communication system.

If instead of using the simple code, each binary digit was transmitted only once, and information from the source has to be transmitted at the same rate in digits/second as when using the code, derive an expression for the channel capacity.

6 Text which is made up of a sequence of letters drawn from the twenty-six letter English Alphabet, together with the 'space letter' is to be transmitted over a noiseless communication channel and security is to be provided by encryption. Examine the following two proposed methods of encryption:

(i) the letters are represented respectively by the integer numbers 0 to 26 and, prior to transmission over a 27-level amplitude modulated channel, an integer drawn from the set 0 to 26 is added modulo-27 to the integer representing the letter. The added ~~letters~~ are equally integers probable and are statistically independent one from another.

(ii) the letters are first encoded using a binary source coder which removes all source redundancy and a binary digit is then added modulo-2 to each source digit. The added numbers are generated by an m-stage shift register system that generates sequences whose periodicity is $2^m - 1$ and is such that the number of **ones** in the sequence differs from the number of **zeros** by one.

Derive an expression for the capacity of a channel whose matrix is

$$\begin{array}{cc} & \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{array}{c} x_1 \\ \\ x_2 \end{array} & \left[\begin{array}{ccc} 1-p-q & q & p \\ \\ p & q & 1-p-q \end{array}\right] \end{array}$$

Explain the physical significance of the channel.

The channel is to be used for binary data transmission and a single-parity-check code is used in conjunction with the channel to provide error protection. The single-parity-check code takes blocks of k information digits; I_1, I_2, ..., I_k. It appends to these a check digit, C, which is obtained using the parity check equation

$$I_1 + I_2 + \ldots + I_k = C,$$

where + denotes modulo-2 addition.

Examine the error protection capabilities of the codes, given that the receiver re-computes the parity-check equation. Explain how the characteristics of the channel can be used to provide an increased protection against transmission errors.

E.3.14 INFORMATION THEORY

ISE3.15

Q1 SC3

SOLUTIONS 2001

Mark scheme missing

marks

An optimum instantaneous code must satisfy the following three conditions

(i) If $P_K \leq P_J$, then $l_K \geq l_j$

(ii) The longest two codewords have the same length

(iii) The longest two codewords differ only in the last bit and correspond to the two least likely symbols.

Proof Consider the optimum code C_m and form the

(i) code C_m' by interchanging l_K and l_j

If we do this then

$$L(C_m') - L(C_m) = \sum_{\substack{i=1 \\ i \neq K \\ i \neq j}}^{m} P_i l_i + P_j l_K + P_K l_j$$

$$- \sum_{\substack{i=1 \\ i \neq k \\ i \neq j}}^{m} P_i l_i + P_j l_j + P_K l_K$$

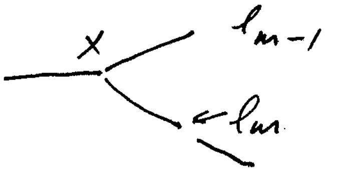
Square Box = marks

$\boxed{15}$

$$= (P_j - P_K)(l_K - l_j)$$

and hence since $P_j \geq P_K$ it follows that $l_K \geq l_j$ since $L(C_m)$ is optimum ie min. length

(ii) Consider now the two longest codewords



Now to point x they have common stem and we can clearly remove final digit in l_m and still maintain prefix condition

2/23

hence $l_m = l_{m-1}$

iii) Since they have common stems and are of equal length they differ in the final digit.

This proves the proposition that code satisfies conditions (i), (ii) and (iii)

Now consider the formation of a merged code with the merged symbol having the stems of the two-lowest-probability codewords and let it be assigned probability $P_{m-1} + P_m$.

The two codes are thus as follows

	C_{m-1}	C_m
P_1	$\underbrace{w_1}_{\text{word}} \; \underbrace{l_1}_{\text{length}}$	$w_1 \; l_1$
$\vdots$		$\vdots$
P_{m-2}	$w_{m-2} \; l_{m-2}$	$w_{m-2} \; l_{m-2}$
$P_{m-1} + P_m$	$w_{m-1} \; l'_{m-1}$	$w_{m-1}.0 \quad l'_{m-1}+1 \; (= l_{m-1})$
		$w_{m-1}.1 \quad l'_{m-1}+1 \; (= l_m)$

Thus if we denote $L(C_{m-1})$ as the average length of C_{m-1} and $L(C_m)$ as the average length of C_{m-1}

we see immediately that

$$L(C_m) = \sum_{i=1}^{m} P_i l_i = \sum_{i=1}^{m-2} P_i l_i + P_{m-1}(l'_{m-1}+1) + P_m(l'_{m-1}+1)$$

$$= L(C_{m-1}) + P_{m-1} + P_m$$

Hence the average length of C_m differs from that of C_{m-1} by a fixed amount that is independent of C_{m-1}

Hence minimizing the length of C_{m-1} minimizes length of C_m.

We now ensure that the code C_{m-1} satisfies the conditions (i), (ii), (iii).

Then form code C_{m-2} from C_{m-1} by the same merging process and continue to repeat until arrive at final two symbols which are then encoded using 0 and 1 respectively.

This completes the proof.

Part 2

Although the Huffman code is optimum it suffers, as do all lossless encoding techniques from a number of problems. The ~~two~~ most important are.

(i) If the source statistics change then the code is no longer optimum and can result in an increase (data expansion) rather than compression

(ii) If the number of symbols is large (say scan speech, or text) then the codebook size grows exponentially with N the size of the no. of symbols and L the block length encoded

i.e. no of symbols in extended codebook $\Rightarrow N^L$

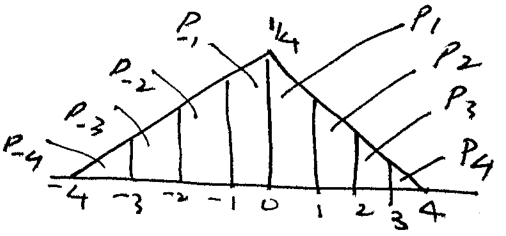
$\boxed{5}$

Part 3 4/23

Easy to implement using simple ROM look-up table

Q2 Part 1 5/23

The pdf is



and the quantization levels are as indicated.
The probabilities associated with the quantization levels are

$$P_1 = P_{-1} = 7/32$$
$$P_2 = P_{-2} = 5/32$$
$$P_3 = P_{-3} = 3/32$$
$$P_4 = P_{-4} = 1/32$$

∴ The entropy of the quantized source is

$$H(X) = -2\left\{ \tfrac{7}{32}\log_2 \tfrac{7}{32} + \tfrac{5}{32}\log_2 \tfrac{5}{32} + \tfrac{3}{32}\log_2 \tfrac{3}{32} + \tfrac{1}{32}\log_2 \tfrac{1}{32} \right\}$$

The encoder as as follows:

7

	Codeword
L_4 ($P_4 = 1/32$)	100
L_3 ($P_3 = 3/32$)	101
L_2 ($P_2 = 5/32$)	110
L_1 ($P_1 = 7/32$)	111
L_{-1} ($P_{-1} = 7/32$)	011
L_{-2} ($P_{-2} = 5/32$)	010
L_{-3} ($P_{-3} = 3/32$)	001
L_{-4} ($P_4 = 1/32$)	000

MSB LSB

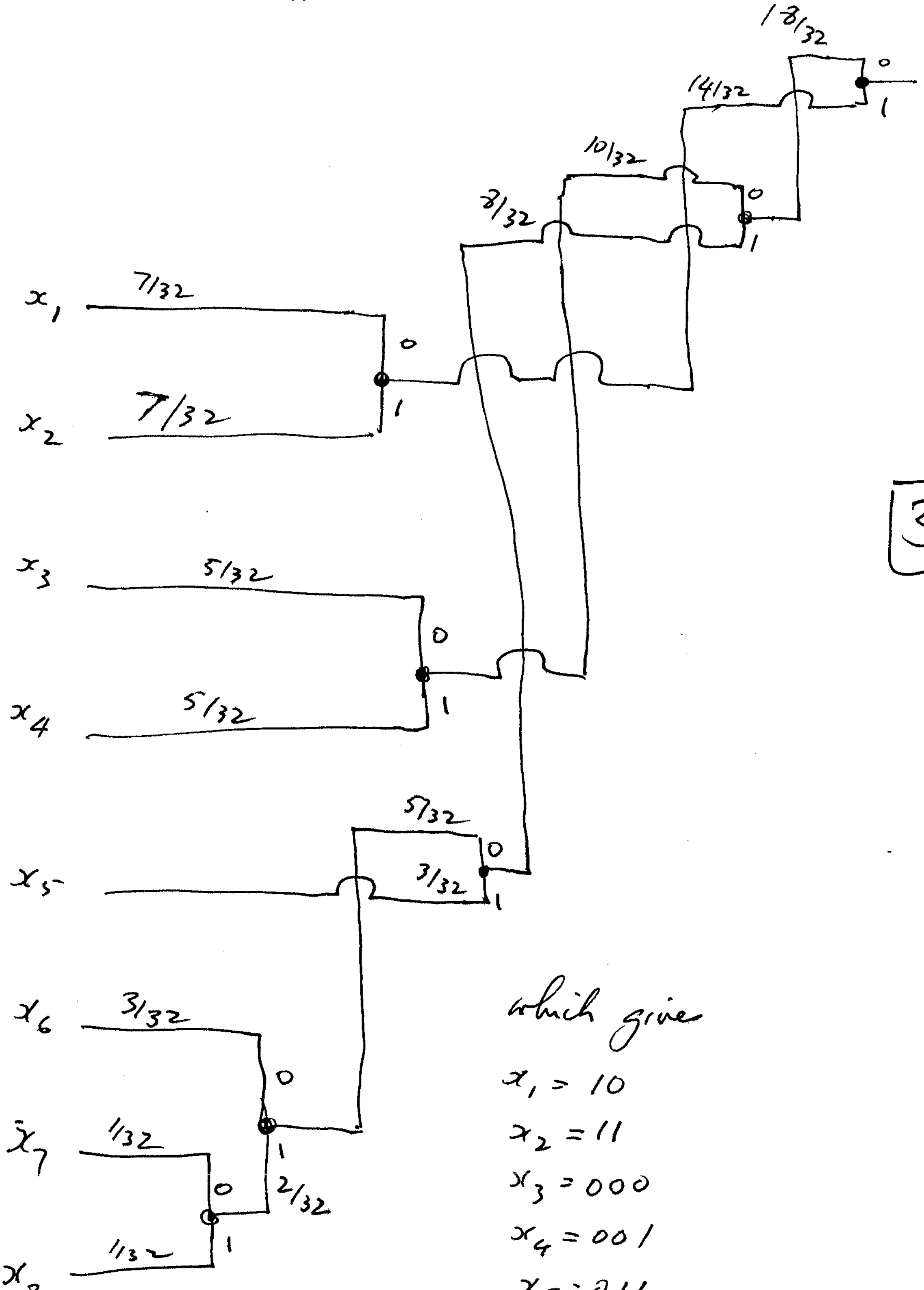
The number of 0's used on average/sample

$$= \frac{5}{32}\times 1 + \frac{3}{32}\times 1 + \frac{1}{32}\times 2 + \frac{7}{32}\times 1 + \frac{5}{32}\times 2 + \frac{3}{22}\times 2 + \frac{1}{32}\times 3 = \frac{36}{32}$$

No. of 1's $= \frac{60}{32}$

$$\therefore P_{zero} = 0.375$$
$$P_{one} = 0.625$$

With a Huffman encoder we obtain the codebook using the following construction



$\boxed{3}$

which gives

$x_1 = 10$

$x_2 = 11$

$x_3 = 000$

$x_4 = 001$

$x_5 = 011$

$x_6 = 0100$

$x_7 = 01010$

$x_8 = 01011$

Hence average no on 1's $= 40/32$

average no of 0's $= 49/32$

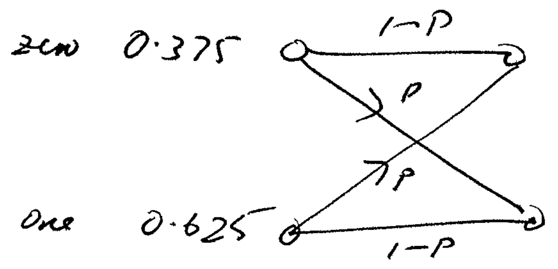
Hence $P_{zero} = 49/89 = 0.55$

$P_{one} = 40/89 = 0.45$ Note closer to 50:50 than in uncoded case.

The maximum amount of information that can be sent error free per transmitted digit is

$$I = H(X) - H(X/Y) \qquad (1)$$

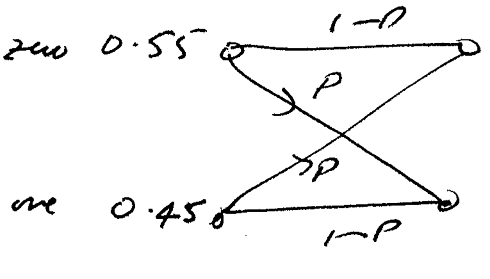
With the 3-bit encoder we thus have



Case I

6

and with the Huffman encoder we have



Case II

Note P is same in both cases since determined by noise alone

Hence in case I we obtain

$$H(X) = -\left\{0.375 \log_2 0.375 + 0.625 \log_2 0.625\right\} \qquad (2)$$

$$H(X/Y) = +\sum_i H(X/y_i)\, P(y_i)$$

and $$H(X/y_i) = -\left\{ P(x_1/y_i) \log_2 P(x_1/y_i) + P(x_2/y_i) \log_2 P(x_2/y_i) \right\}$$

$$\ldots = -\left\{ P\left(\tfrac{x_1}{y_1}\right) \log_2 P\left(\tfrac{x_1}{y_1}\right) + P\left(\tfrac{x_2}{y_1}\right) \log_2 P\left(\tfrac{x_2}{y_1}\right) \right\}$$

$$H\left(\tfrac{x}{y_2}\right) = -\left\{ P\left(\tfrac{x_1}{y_2}\right) \log_2 P\left(\tfrac{x_1}{y_2}\right) + P\left(\tfrac{x_2}{y_2}\right) \log P\left(\tfrac{x_2}{y_2}\right) \right\}$$

Thus
$$H\left(\tfrac{x}{y}\right) = -\left\{ P(y_1) P\left(\tfrac{x_1}{y_1}\right) \log_2 P\left(\tfrac{x_1}{y_1}\right) + P(y_1) P\left(\tfrac{x_2}{y_1}\right) \log_2 P\left(\tfrac{x_2}{y_1}\right) + P(y_2) P\left(\tfrac{x_1}{y_1}\right) \log_2 P\left(\tfrac{x_1}{y_1}\right) + P(y_2) P\left(\tfrac{x_2}{y_2}\right) \log_2 P\left(\tfrac{x_2}{y_2}\right) \right\} \quad -(3)$$

But
$$P(y_1) = P(x_1) P\left(\tfrac{y_1}{x_1}\right) + P(x_2) P\left(\tfrac{y_1}{x_2}\right)$$
$$= 0.375(1-p) + 0.625p$$
$$= 0.375 + 0.25p$$
$$P(y_2) = .375p + .625(1-p)$$
$$= 0.625 - 0.25$$

and
$$P\left(\tfrac{x_1}{y_1}\right) = \frac{P(x_1, y_1)}{P(y_1)} = \frac{P(x_1) P\left(\tfrac{y_1}{x_1}\right)}{P(y_1)} = \frac{0.375(1-p)}{0.375 + 0.25p} \quad (4)$$

$$P\left(\tfrac{x_1}{y_2}\right) = \frac{P(x_1, y_2)}{P(y_2)} = \frac{0.375p}{.625 - 0.25p} \quad (5)$$

$$P\left(\tfrac{x_2}{y_1}\right) = \frac{P(x_2) P\left(\tfrac{y_1}{x_2}\right)}{P(y_1)} = \frac{0.625p}{0.375 + 0.25p} \quad (6)$$

$$P\left(\tfrac{x_2}{y_2}\right) = \frac{P(x_2) P\left(\tfrac{y_2}{x_2}\right)}{P(y_2)} = \frac{0.625(1-p)}{0.625 - 0.25p} \quad (7)$$

...bstituting Eqn's (4) → (6) into Equation 3 and then ...mbining with Eq (2) into (1) gives the required ...xpression.

9/23

I am looking for an understanding and statement

(i) that I is determined by $P(x_1)$, $P(x_2)$ for a given fixed P.

(ii) that through equations (4) to (6) $H(x/y)$ can be determined in terms of the measurable values P, $P(x_1)$ & $P(x_2)$.

Part 2

Case II The result is identical to case I except that $P(x_1)$ and $P(x_2)$ are now 0·55 and 0·45 respectively. As these are much closer to 0·5 ie the inputs being equi-probable I will be closer to the maximum ~~theoretical~~ capacity of

$$1 + P\log_2 P + (1-P)\log_2(1-P)$$

4

Part 3

I am also looking for a statement that the source coding by Huffman is more efficient so that fewer digits are needed to send same output – ie double gain

Part 4

If a BCD encoder had been used then

$P(1) = P(0) = 0·5$

Hence $I = 1 + P\log_2 P + (1-P)\log_2(1-P)$ — hence higher capacity

But Source coder still better and

aim should be to use extended length Huffman to get $P(0) = P(1)$ together with source coder gain.

Q3.

Part 1 The sequence $S_i = x_{i_1}, x_{i_2}, \ldots, x_{i_L}$; where each x_{i_j} is chosen from the set of symbols $x_1, \ldots, x_N$

Thus $$\tilde{H} = -\sum_{i=1}^{N^L} P(S_i) \log P(S_i)$$

$$= -\sum_{i_1=1}^{N} \cdots \sum_{i_L=1}^{N} P(x_{i_1}, \ldots, x_{i_L}) \log P(x_{i_1}, \ldots, x_{i_L}) \quad \text{--- (1)}$$

But $$P(x_{i_1}, \ldots, x_{i_L}) = P(x_{i_1}, \ldots, x_{i_R}) P\left(\frac{x_{i_{R+1}}}{x_{i_1}, \ldots, x_{i_R}}\right) \cdot P\left(\frac{x_{i_{R+2}}}{x_{i_2}, \ldots, x_{i_{R+2}}}\right) \cdot \cdots \cdot P\left(\frac{x_{i_L}}{x_{i_{L-R}}, \ldots, x_{i_{L-1}}}\right)$$

Hence on substituting this into (1) we obtain

$$\tilde{H} = -\sum_{i_1=1}^{N} \cdots \sum_{i_L=1}^{N} P(x_{i_1}, \ldots, x_{i_L}) \left[\log P(x_{i_1}, \ldots, x_{i_R}) + \log P\left(\frac{x_{i_{R+1}}}{x_{i_1}, \ldots, x_{i_R}}\right) + \cdots + \log P\left(\frac{x_{i_L}}{x_{i_{L-R}}, \ldots, x_{i_{L-1}}}\right) \right]$$

But $$-\sum_{i=1}^{N} \cdots \sum_{i_L=1}^{N} P(x_{i_1}, \ldots, x_{i_L}) \log P\left(\frac{x_{i_j}}{x_{i_{j-R}}, \ldots, x_{i_{j-1}}}\right) = H(X)$$

$j = R+1, \ldots, L$

The true entropy of the Markov source

Thus we have

$$\tilde{H} = (L-R)H(X) - \sum_{i_1=1}^{N} \cdots \sum_{i_L=1}^{N} P(x_{i_1},\ldots,x_{i_L}) \log P(x_{i_1},\ldots,x_{i_R})$$

$$= (L-R)H(X) - \sum_{i=1}^{N} \cdots \sum_{i_R=1}^{N} P(x_{i_1},\ldots,x_{i_R}) \log P(x_{i_1},\ldots,x_{i_R})$$

But since $-\sum P_i \log P_i \geq 0$ we immediately have

$$\tilde{H} = (L-R)H(X) + \delta$$

where $\delta \geq 0$, and from it is clear that this is independent of L, provided $L > R$.

8

The physical significance

for those x_{i_j}, $j > R$ it follows that the same generates $H(x)$ for each such symbol since the full memory constrains is involved, and it is for this ~~the~~ condition that $H(x)$ is defined.

For x_{i_1} there is no memory involved hence the info generated is $\sum_{i=1}^{N} -P(x_{i_1}) \log P(x_{i_1}) \geq 0$

3

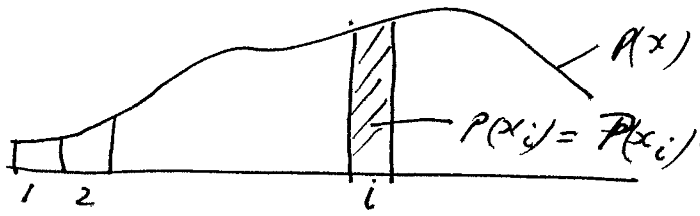
For x_{i_2}, the memory due to first symbol is involved, hence less info than for x_{i_1} ≥ 0

For each successive x_{i_j} $j = 3, \ldots, R-1$ the memory constrains increases so that less info generated ≥ 0

Part 2 $\qquad\qquad\qquad\qquad$ 12/23

The sampled signal has pdf $p(x)$

hence let the probability of sample being in range i be as indicated below



where $P(x_i)$ is average value of $p(x)$ in i^{th} interval

Hence $$H(X) = -\sum_{i=1}^{M} P(x_i)\Delta \log P(x_i)\Delta \quad ; \quad M = \frac{X}{\Delta} \leftarrow \text{range}$$

$$= -\left\{\sum_{i=1}^{M} P(x_i)\Delta \log P(x_i) + \sum P(x_i)\Delta \log \Delta\right\}$$

4

and as $\Delta \to 0$ we obtain

$$H(X) = -\int_0^X p(x)\log p(x)\,dx$$

$$- \log\Delta \int_0^X p(x)\,dx$$

tends to $-\infty$

hence $H(X) \to \infty$ as we would expect since quantizing to infinite accuracy.

Part 3

Let $P_1, P_2, \dots, P_M$ $\quad P_1 + \dots + P_M = 1$

and $Q_1, Q_2, \dots, Q_M$ $\quad Q_1 + \dots + Q_M = 1$

be two different probability sets

Now $\log x \leq x - 1$ with equality at $x = 1$

$\therefore \log\left(\frac{Q_i}{P_i}\right) \leq \frac{Q_i}{P_i} - 1$ with equality if $P_i = Q_i$ for all i

Thus $P_i \log\left(\frac{Q_i}{P_i}\right) \leq Q_i - P_i$

and $\sum_{i=1}^{M} P_i \log\left(\frac{Q_i}{P_i}\right) \leq 0$

$\boxed{5}$

Hence it follows that

$$\sum P_i \log(1/P_i) \leq \sum P_i \log 1/Q_i$$

with equality if $P_i = Q_i$ for all i

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

$\therefore \sum_{i=1}^{M} P_i \log \frac{1}{P_i} \leq \log M$ with equality if and only if $P_i = \frac{1}{M}$ for all i

Q 4 Part 1

The AEPT states that the source will produce a typical sequence (ie one of a typical set, containing correct number of each symbol type, with the correct intersymbol conditional probabilities) and it will do so with probability that approaches unity. The probability of producing a typical sequence approaches zero as the sequence length becomes large.

3

If a source has an entropy of H bits/output symbol then the entropy of an N symbol set is NH bits/sequence of N symbols. Accordingly from the noiseless coding theorem we know that as $N \to \infty$ we need use only NH binary digits to encode the source sequence without loss. Thus the number of sequences 2^{NH} with high probability

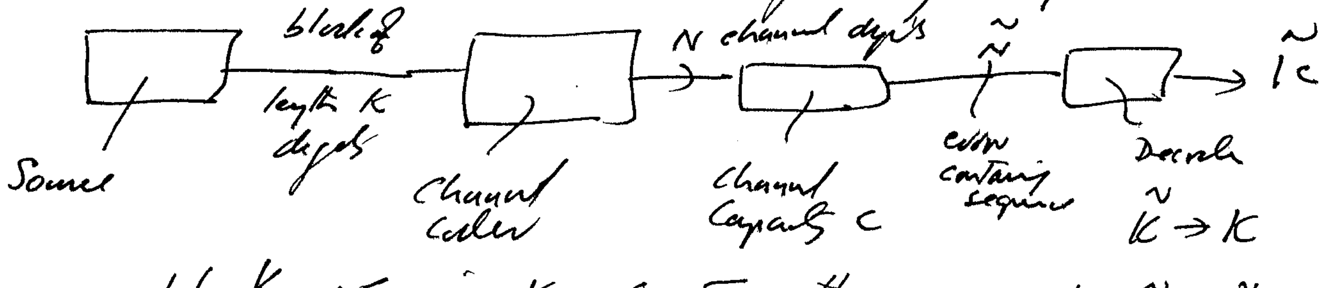
Part 2

Shannons random coding argument assumes that each codeword of length N binary digit (say) is selected at random from the complete set of 2^N possible codewords. This construction procedure allows all codewords to be the same (very bad = useless), but when we average the codeword error probability over all possible codes (good, bad and indifferent) the average error probability goes to zero as $N \to \infty$. From this it can be concluded that a code(s) with error probability $= 0$ must exist

The importance of the argument is that it forms the basis of proof of the channel coding (capacity) theorem. Without it, the coding theorem could not be proved since codes satisfying the theorem have yet to be found. Although the random coding argument is very likely to result in the generation of a very good code (one satisfying the theorem) it could not be tested easily, and more important, ~~it~~ no practical decoding scheme exists – the decoder would have no algorithmic structure and decoding would have to be as a maximum likelihood basis, and this would infinite complexity since 2^{NH} grows without limit as $N \to \infty$ $\boxed{A}$

Part 3

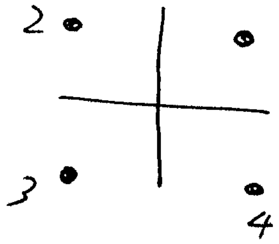
The coding theorem states that if a source has an entropy H, and the channel has capacity C, then the output from the source can be transmitted/communicated error free, provided $H = C - \delta$, δ as small as we like. If $H > C$ then error free transmission/comm^n is not possible. Block encoding can be used in the following way 3



If $\frac{K}{N} < C$ ie $\frac{K}{N} = C - \delta$ then can go from $\tilde{N}$ to $\tilde{K} = K$ with probability 1

Part 4

The channel matrix is obtained by considering the errors that can occur



let the input be denoted
$x_1 (=1)$
$x_2 (=2)$
$x_3 (=3)$
$x_4 (=4)$

and the output be y_1, y_2, y_3, y_4

The channel matrix is

$$\begin{array}{c|c}
 & \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\
\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} & \begin{bmatrix} P_1 & P_2 & P_3 & P_2 \\ P_2 & P_1 & P_2 & P_3 \\ P_3 & P_2 & P_1 & P_2 \\ P_2 & P_3 & P_2 & P_1 \end{bmatrix}
\end{array}$$

P_1 = prob. of correct detection, same for each transmitted signal point

$P_1 \neq P_2 \neq P_3$

Now the channel is doubly uniform so it has a capacity

$$C = \log K + \sum_{i=1}^{K} a_i \log a_i$$

where K is the number of signal points, and a_i is a transition probability.

From the matrix we have

$$C = \log_2 4 + P_1 \log P_1 + P_3 \log_2 P_3 + 2 P_2 \log_2 P_2$$

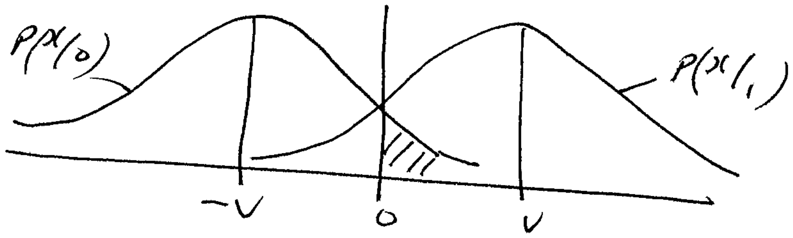
bits/signal

Hence if S/N is large $P_1 \Rightarrow 1$; $P_2 \Rightarrow P_3 \Rightarrow 0$

$\therefore$ Capacity = 2 bits/transmitted pulse

10

Q5 - Consider the transmission of a +V or -V volt pulse
Then we have



The probability of pulse being in error (P_E) is

$$P = 1 - \int_0^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-V)^2}{2\sigma^2}}\,dx$$

$$= 1 - \frac{1}{2} - \int_0^{V} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-V)^2}{2\sigma^2}}\,dx$$

Change variable so that $z = \frac{x-V}{\sqrt{2}\,\sigma}$

Then we obtain
$$P = \frac{1}{2} - \int_{z=-\frac{V}{\sqrt{2}\sigma}}^{z=0} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-z^2}\,dz \cdot \sqrt{2}\,\sigma$$

$$= \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_{-\frac{V}{\sqrt{2}\sigma}}^{z=0} e^{-z^2}\,dz$$

$$= \frac{1}{2}\left\{1 - \frac{2}{\sqrt{\pi}} \int_0^{V/\sqrt{2}\sigma} e^{-z^2}\,dz\right\}$$

$$= \frac{1}{2}\left\{1 - \operatorname{erf}\left(\frac{V}{\sqrt{2}\sigma}\right)\right\}$$ 1

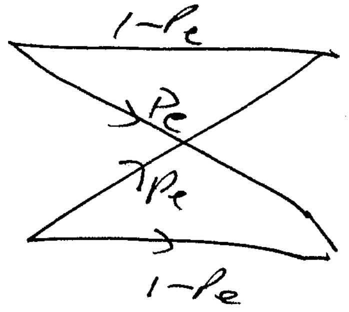
$\boxed{8}$

Now since a simple repetition code is to be used, the probability of error is the probability that two, or three digits will be in error ie

$$P_e = \sum_{i=2}^{3}\binom{3}{i}p^i(1-p)^{3-i} = 3p^2(1-p)+p^3$$

where p is given by (1)

The channel is thus



I expect an argument (not just a statement) that

$$C = 1 + P_e\log_2 P_e + (1-P_e)\log_2(1-P_e) \quad (2)$$

— will accept argument that channel is doubly symmetric ie $C = K + \sum a_i \log a_i$

hence (2)

8

Now, if instead of using three pulses, a single pulse is used, then only 1/3 of the bandwidth would be necessary and hence the noise power is reduced from σ^2 to $\frac{\sigma^2}{3} = \sigma_i^2$ where $\sigma_i = \frac{\sigma}{\sqrt{3}}$

Hence the expression for P is simply obtained by replacing σ by $\frac{\sigma}{\sqrt{3}}$ in equation 1

thus $$P = \frac{1}{2}\left\{1 - erf\left(\frac{V\sqrt{3}}{\sqrt{2}\sigma}\right)\right\}$$
$$= \frac{1}{2}\left\{1 - erf\left(\frac{V}{\sigma}\sqrt{3/2}\right)\right.$$

and the capacity is

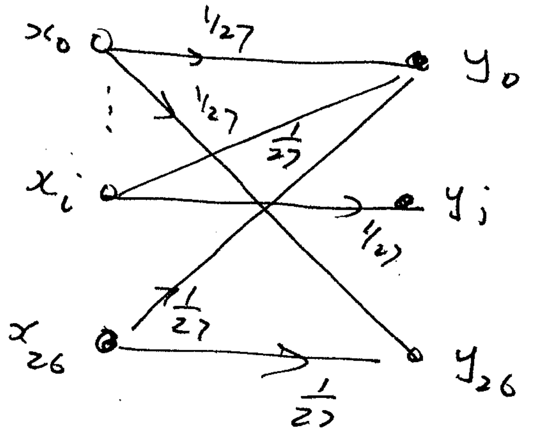
$$C = 1 + P\log_2 P + (1-P)\log_2(1-P)$$

Q6. 20/23

Although the wording is long, the question is relatively simple.

Part 1

The encryption channel diagram is



and its matrix is

$$\begin{array}{c} \\ x_0 \\ \\ x_{26} \end{array}\begin{array}{c} \begin{array}{ccc} y_0 & & y_{26} \end{array} \\ \begin{bmatrix} \frac{1}{27} & \cdots & \frac{1}{27} \\ & & \\ \frac{1}{27} & & \frac{1}{27} \end{bmatrix} \end{array}$$

The channel is doubly-uniform and has a capacity

$$C = \log 27 - \sum_{i=0}^{27} \frac{1}{27} \log \frac{1}{27}$$

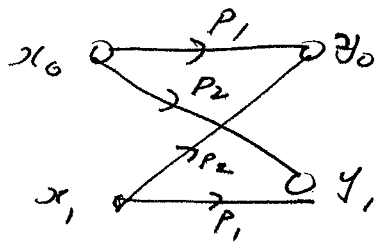
$$= 0$$

$\boxed{8}$

The proposed scheme thus communicates no information with respect to the encryption and hence it is totally secure. It is the so-called ONE-TIME PAD. The pad must not be used more than once if security is to be maintained. Re-use ~~pad~~ provide the cryp-analyst with information that he can use.

Part 2 2/23

In this case the channel is



but $p_1 = p_2$ in this case although they are very close.

They are so close that the capacity of the channel is very nearly zero, but not quite. Thus the system appears to be secure. It is not, however, for the following reasons 6

(i) a small amount of encryption information is actually transmitted

and (ii) more importantly the pseudo-random sequence is from a limited set each member of which can easily be re-generated. Addition of the regenerated sequence will remove the encryption, and all that is left is the 'encryption' due to source coding, which is easily broken.

Part 3

The channel is uniform from the input and hence its capacity is

$$C = \max_{P(X)} I \quad \text{where} \quad I = H(Y) + (1-p-q)\log(1-p-q) + q\log q + p\log p$$

Now to maximise I we have to maximise $H(Y)$.

But
$$P(y_1) = P(x_1)(1-p-q) + P(x_2)p$$
$$P(y_2) = q\,P(x_1) + q\,P(x_2)$$
$$P(y_3) = p.\,P(x_1) + P(x_2)(1-p-q)$$

and it is easy to show that $P(x_1) = P(x_2) = 1/2$ maximises $H(Y) = \sum_{i=1}^{3} P(y_i)\log 1/P(y_i)$

we get
$$P(y_1) = (1-q)\tfrac{1}{2}$$
$$P(y_2) = q$$
$$P(y_3) = (1-q)\,1/2$$

Note not equally-likely $\therefore H(Y) \neq \log 3$.

on substituting we get

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

4

Part 4

The single-parity check code can detect any pattern of errors that contains an odd number of errors — but cannot correct.

With the channel we will get a sequence as follows

101110010E111--0 etc

But change E to 0 or 1 and then re-working the parity-check equation we have a real chance of correcting the most likely source of error. The same argument applies to patterns of more than one error. Clearly we cannot be certain that the correction is satisfying and any even number of errors that are not indicated as erasures will defeat the system.

2