Question 1

(i)

$$\mathrm{H} = -\sum_{\mathrm{i}=0}^{3} \mathrm{p}(\mathrm{x_i})\log_2 \mathrm{p}(\mathrm{x_i})$$

(1 mark)

$$p(x_0) = p(x_3) = p(1 < x < \infty) = \frac{1}{\sqrt{2\pi}} \int_1^{\infty} e^{-y^2/2}\ dy = 0.1611$$

$$p(x_1) = p(x_2) = p(0 < x < 1) = \frac{1}{\sqrt{2\pi}} \int_0^{1} e^{-y^2/2}\ dy = 0.5 - 0.1611 = 0.3389$$

(2 marks)

Therefore

H=1.9068 bits/symbol

(1 mark)

(ii) The capacity of this particular channel is given by

S/N=10^(40/10)=10000;

C = B*log2(1 + S/N) =39864bits/second.

(2 marks)

The entropy of the sampled signal is H=1.9068bits/symbol.

Given the channel capacity C=39864 bits/s, the maximum number of symbols transmitted without error is R=C/H=20906 symbols/s.

(2 marks)

Q1(b)

E_b/N_0 is related to the system signal to noise ratio through the following identities for binary systems (i.e. systems that send 1 bit per symbol):

$$\frac{\mathrm{E_b}}{\mathrm{N_0}} = \frac{\mathrm{ST}}{\mathrm{N_0}} = \frac{\mathrm{S}}{\mathrm{RN_0}} = \frac{\mathrm{S}}{\mathrm{N_oB}}\left(\frac{\mathrm{B}}{\mathrm{R}}\right)$$

Where B = bandwidth

S = average signal power

T = symbol (bit) time period

R = bit rate (data rate)

For the case where transmission bit rate is equal to the channel capacity, R=C, we have

$$\frac{\mathrm{E_b}}{\mathrm{N_0}} = \frac{\mathrm{S}}{\mathrm{N_oB}}\left(\frac{\mathrm{B}}{\mathrm{C}}\right)$$

Using

$$\frac{\mathrm{C}}{\mathrm{B}} = \log_2(1 + \frac{\mathrm{S}}{\mathrm{N_oB}})$$

i.e.

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{N_o}\left(\frac{C}{B}\right)\right)$$

let

$$x = \frac{E_b}{N_o}\left(\frac{C}{B}\right)$$

then

$$\frac{C}{B} = x \log_2 (1 + x)^{\frac{1}{x}}$$

i.e.

$$1 = \frac{E_b}{N_o} \log_2 (1 + x)^{\frac{1}{x}}$$

(3 marks)

In the limit, as $\frac{C}{B} \to 0$ and using

$$\lim_{x \to 0} (1 + x)^{1/x} = e$$

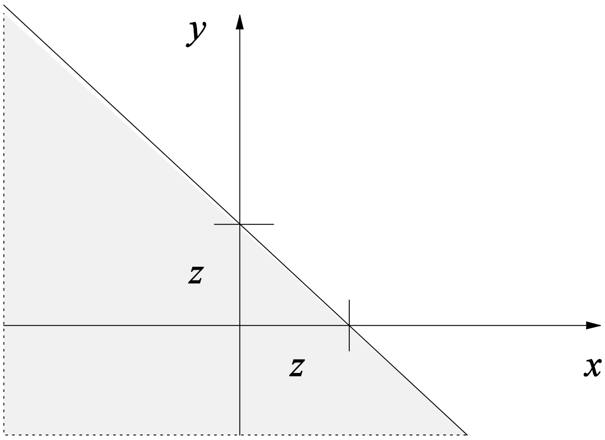
(2 marks)
we get

$$\frac{E_b}{N_o} = \frac{1}{\log_2 e} = 0.693 = -1.59 dB$$

(1 mark)

Q1(c)

When x and y are random variables the probability of their sum can be found from a consideration of the following figure:



(1 mark)

$$P(x+y \leq z) = P\{(x,y)\ in\ shaded\ area\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,dxdy$$
$$f_Z(z) = \frac{d}{dz}\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,dxdy$$
$$= \int_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx$$

(3 marks)

For statistically independent variables, we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)fY(z-x)dx$$
$$= \int_{-\infty}^{\infty} f_X(z-y)fY(y)dy$$

(2 marks)

Question 2

a.

There are 12 different months in each year and given the four persons, the total number of cases is 12X12X12X12=12^4.

(1 mark)

The number of cases where they are born at totally different months is given by

12X11X10X9

(1 mark)

So the probability of them born in totally different months is

12X11X10X9/12^4= 0.573

(1 mark)

So the probability of at least two of them born in the same month is

1-0.573=0.427

(1 mark)

b.

Let E_1 and E_2 be the events of an accident in the first and second years, respectively. Since the lawyers form 1/6 of the group and the miners 5/6

$$P(E_1) = \tfrac{1}{6}p_1 + \tfrac{5}{6}p_2$$

(2 marks)

The probability of a lawyer having a car accident in both years is p_1^2. For miners it is p_2^2. Hence

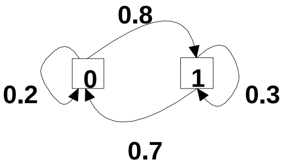
$$P(E_1 \cap E_2) = \tfrac{1}{6}{p_1}^2 + \tfrac{5}{6}{p_2}^2$$

(2 marks)
Therefore

$$P(E_2 \mid E_1) = \frac{p_1^{\,2} + 5p_2^{\,2}}{p_1 + 5p_2}$$

is the desired conditional probability.
(2 marks)

c.



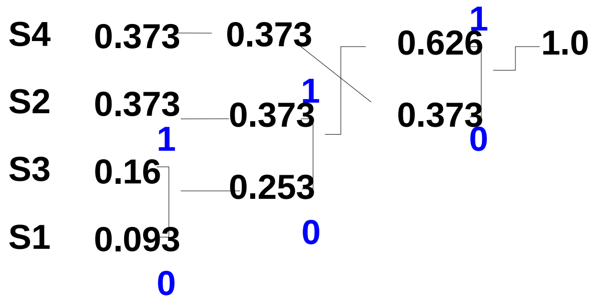
(i) Let the source be referred to as X then the Entropy is given by:
H(X) = P(0)*H(X|0) + P(1)*H(X|1).

We begin by solving for the *a priori* probabilities. This can be done using the fact that:
P(1) + P(0) = 1
Substituting this into
P(0) = P(0)P(0|0) + P(1)P(0|1)
and using the conditional probabilities indicated in the Markov model [P(1|1) = 0.4, etc] we get P(0) = 7/15 and therefore P(1) =8/15 .
(2 marks)

The conditional entropies H(X|0) and H(X|1) are then computed:
H(X|0) = P(0|0)*log2(1 / P(0|0)) + P(1|0)*log2(1 / P(1|0)) = 0.7219
H(X|1) = P(1|1)*log2(1 / P(1|1)) + P(0|1)*log2(1 / P(0|1)) = 0.8813
(2 marks)

So finally,
H(X) = P(0)*H(X|0) + P(1)*H(X|1) = 0.7219*7/15+0.8813*8/15 = 0.8069 bits / symbol
(1 mark)

(ii) To determine a unique, prefix-free coding scheme for this source close to its entropy, we can use extension code by using groups of two symbols as the basis and then apply Huffman coding to this new set of symbols. The new set of symbols are S1=00, S2=01, S3=11, and S4=10. Their associated probabilities are respectively:
p(00)=p(0)*p(0|0)=0.2*7/15=0.0933
p(01)=p(1)*p(0|1)=0.7*8/15=0.3733
p(11)=p(1)*p(1|1)=0.3*8/15=.1600
p(10)=p(0)*p(1|0)=0.8*7/15=0.3733
(2 marks)
Apply Huffman coding to these four symbols we have



So, the results are:
S1: 100
S2: 11
S3: 101
S4: 0
(2 marks)
[Note: variations are possible depending on the exact algorithm used. However, to be acceptable, the code must be uniquely decodable, prefix free and efficient]

(iii)
The average code length is given by:

$$\bar{n} = \sum_{i=1}^{4} n_i P_i$$

where n_i is the bit length of the codeword for symbol S_i, and P_i is the probability for S_i.
Average length $\bar{n}$ = 0.093*3+0.373*2+0.16*3+0.373*1=1.8780 bits / symbol
Back to the original symbol (binary symbol), it is 1.878/2=0.939bits/symbol. Then the coding efficiency is

$$\frac{0.8069}{\bar{n}} = 0.8593$$

(1 mark)

Question 3

a.

Additive White Gaussian Noise (AWGN) refers to a random signal with a constant power spectral density (white) and a Gaussian distributed amplitude. The term 'additive' indicates that this random noise signal is added to the original signal as it passes through the channel.
(2 marks)

AWGN is an appropriate model for noise because:

1) Noise in real systems results from the combination of interference from many different random sources, and a result from statistics (the central limit theorem) states that the sum of a large number of random variables (possibly each with different distributions) is a random variable with a distribution that approaches a Gaussian distribution.

(1 mark)

2) Many types of noise (notably thermal noise) have constant power spectrums over the operating bandwidths of communications systems and thus a white noise source makes a useful model.

(1 mark)

b.

Suppose there are M antennas at the transmitter and also M antennas at the receiver. Each of the transmit antennas sends a symbol at time n and together there are M symbols sent. We denote them by a vector X. At the receiver side, we also receive M symbols, denoted by a vector Y:

$$X = \begin{bmatrix} x_0 \\ \vdots \\ x_{M-1} \end{bmatrix} \qquad Y = \begin{bmatrix} y_0 \\ \vdots \\ y_{M-1} \end{bmatrix}$$

They have the following relationship: $Y = HX + N$

(2 marks)

Where H is the channel response matrix and N is the noise vector.

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{M-1} \end{bmatrix} = \begin{bmatrix} h_{0,0} & \cdots & h_{0,M-1} \\ \vdots & \ddots & \vdots \\ h_{M-1,0} & \cdots & h_{M-1,M-1} \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{M-1} \end{bmatrix} + N$$

(1 mark)

If we know the channel response matrix H (by some channel estimation methods) and it has full rank (which is usually true in strong multipath environments with antennas spaced sufficiently far away from each other), then given the received symbols Y, we can recover the original symbols X by matrix inversion or the maximum likelihood method, even if some of the values in H are zeros.

(2 marks)

Note for a SISO (single input and single output) system, if the channel response is zero, we will not be able to recover the original signals.

(1 mark)

c.

(i)

The error rate ε=0.02.

Let S_1 be the input with E1=0 and E2=1, and S_2 the output with F1=0 and F2=1. Then with equal probability at the input and also symmetric channel, we have

P(E1)=P(E2)=P(F1)=P(F2)=0.5.

(1 mark)

P(E2∩F2)=P(F2|E2)P(E2)=0.5(1-ε)

(1 mark)

I(E2,F2)=$\log_2$(P(F2,E2)/P(E2)/P(F2))= $\log_2$2(1-ε)=0.9709 bits.

(1 mark)

(ii)

$P(E2 \cap F1)=P(F1|E2)P(E2)=0.5\varepsilon$

(1 mark)

$I(E2,F1)=\log_2(P(F1,E2)/P(E2)/P(F1))= \log_2 2(\varepsilon)=-4.6439$ bits.

(1 mark)

(iii)

The mutual information between the two systems is

$H(S_2)=1$ bit

(1 mark)

$$H(S_2\ |S_1)=\frac{-1}{2}(1-\varepsilon)\log(1-\varepsilon)+\frac{-1}{2}\varepsilon\log\varepsilon+$$
$$\frac{-1}{2}\varepsilon\log\varepsilon+\frac{-1}{2}(1-\varepsilon)\log(1-\varepsilon)$$
$$=-(1-\varepsilon)\log(1-\varepsilon)-\varepsilon\log\varepsilon$$

(2 mark)

$I(S_1,S_2)=H(S_2)-H(S_2|S_1)=1-(-(1-\varepsilon)\log_2(1-\varepsilon)-\varepsilon\log_2\varepsilon)=0.8586$ bits.

(2 mark)

Question 4

a.

i) An outline of the SSL protocol is as follows:

Step 1 Authentication (using certificates and signatures)
Step 2 Exchange of secret key
Step 3 Use of secret key for secure communication
(2 marks)

Steps 1 and 2 are achieved using public key techniques, usually the RSA algorithm. The secure communication in step 3 then uses conventional (symmetric) encryption using the secret key.
(2 marks)

ii) The full communication process is as follows: (e = 3, d = 7, N = 33)

1. Alice publishes *N* and e as the public key for everyone to see, and keep d secret as the private key. (1 mark)

2. Before Bob sends the message *M* = 9 to Alice, he looks up *e* and *N* and calculates

   $C = M^e \bmod N = 3$. This is transmitted to Alice on a non-secure channel. Although other people can read this message, they would not be able to determine *M*, since a one-way function has been used to encrypt the message. (2 marks)

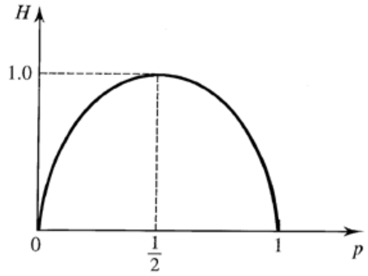
3. To decipher the message Alice uses the value only she knows, i.e. d=7. To get the real message Alice calculates $M = C^d \bmod N = 3^7 \bmod 33 = 9$. (1 mark)

b.

(i)
The highest entropy occurs when the events (or symbols) have equal probabilities.
(1 mark)
As an example, consider a two symbol (binary) source where *p* is the probability of transmitting a '1' we have:

$$H = p\log_2\frac{1}{p} + (1-p)\log_2\frac{1}{1-p}$$

(2 marks)



The maximum entropy value 1.0 is achieved when we have p=0.5.
(1 mark)

(ii) To prove this condition, we need to use the following theorem:
For x>0, lnx≤x−1, with equality only when x=1.
(2 marks)

Assume firstly that no p_k is zero. Then,

$$\sum_{k=1}^{n} p_k \ln\frac{1}{np_k} \leq \sum_{k=1}^{n} p_k\left(\frac{1}{np_k}-1\right)$$

$$=\sum_{k=1}^{n}\left(\frac{1}{n}-p_k\right) = 1-1 = 0$$

$$-\sum_{k=1}^{n} p_k \ln p_k \leq \sum_{k=1}^{n} p_k \ln n = \ln n$$

Hence,
ln can be converted to log by multiplying by a suitable positive constant. Since multiplication by a positive constant does not invalidate an inequality, we have

$$-\sum_{k=1}^{n} p_k \log p_k \leq \log n \quad \text{or} \quad H(S) \leq \log n$$

Since every p_k is positive, equality will arise only if it does for every term in the First Result. So we have $1/(np_k)=1$ for all k. The theorem is therefore proved for non-zero probabilities.
(4 marks)

If a probability, say p_j , is zero, $p_j \ln p_j=0$, and

$$p_j \ln\frac{1}{np_j} < \frac{1}{n} - p_j$$

Thus the first result continues to hold and the upper bound on H(S) then follows. The above inequality prevents equality in the theorem. The proof is finished.
(2 marks)

Question 5

a. i) (2 marks)

The minimum Hamming distance between the codes is $D_{\min} = 2$.

(1 mark)

The number of bit errors that can be detected per code word is given by $D_{\min} - 1 = 1$.

(1 mark)

a. ii) (2 marks)

Forward error correction is impossible for 1-bit errors.

(1 mark)

This is because a 1-bit error will result in a codeword that has the same distance to the original codeword as that to any other valid codeword, i.e., $D_{\min}/2 - 1 = 0$.

(1 mark)

a. iii) (2 marks)

To send eight messages, we need 3 bits per message.

(1 mark)

The linear block code in question uses 4 bits per codeword. So the code rate is ¾ = 75%.

(1 mark)

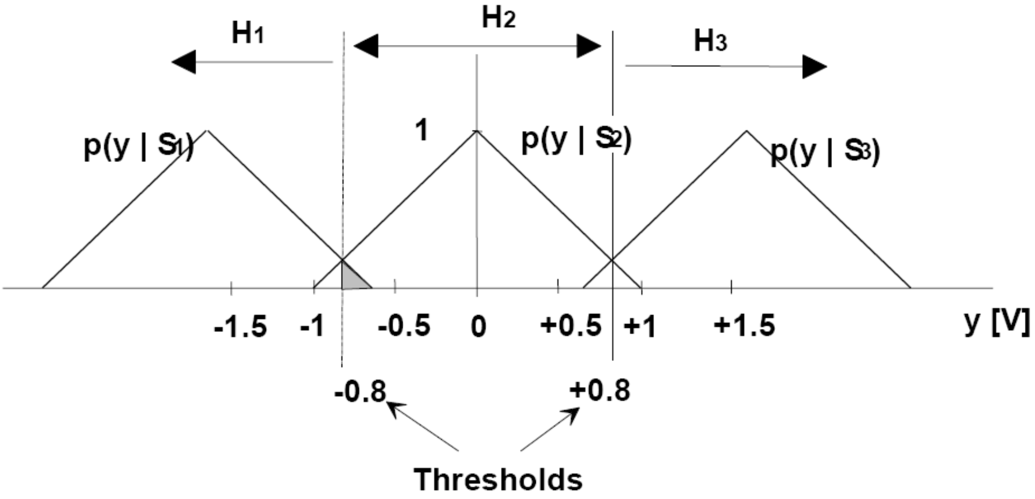
b. i) (4 marks)

For a maximum likelihood detector, we choose the hypothesis that has the highest likelihood value for the observed signal when compared with all possible hypotheses:

$$\hat{H}_{ML} = \arg\max_i \{p(y \mid S_i)\}$$

(1 mark)

Considering a graphical representation of the situation:



(1 mark)

It is clear that the decision thresholds are at –0.8V and +0.8V. The full decision rule is as follows:
Choose H1 if y < -0.8,
Choose H2 if –0.8< y < 0.8, and
Choose H3 if y > 0.8

(2 marks)

b. ii) (6 marks)

The probability of error is calculated by considering all the different possible ways of making an error:
Pe = P(H1|S2)P(S2) + P(H1|S3)P(S3) + P(H2|S1)P(S1) + P(H2|S3)P(S3) + P(H3|S1)P(S1) + P(H3|S2)P(S2)

(2 marks)

From the diagram, it is clear that P(H1|S3) and P(H3|S1) are both zero.
P(H2|S1) is the shaded area under the graph of p(y|S1) so

$$P(H_2 \mid S_1) = \int_{-0.8}^{0.8} p(y \mid S_1)dy = 0.5 \times (\text{triangle base}) \times (\text{triangle height})$$
$$= 0.5 \times 0.2 \times 0.2$$
$$= 0.02$$

(2 marks)

$P(S_1)$ = 1/3.
Also, because of symmetry, P(H1|S2) = P(H2|S3) = P(H3|S2) = P(H2|S1) so Pe = 4* P(H2|S1) * P(S1) = 0.02667

(2 marks)

b. iii) (4 marks)

If P(S2) → 1, then the decision thresholds would move outwards towards –1 and 1.

(2 marks)

In the limit the MAP error probability would tend to zero (If P(S2) = 1 then the receiver could always just choose H2 and would never make an error.)

(2 marks)

Question 6

a. i) (3 marks)

There are three properties:

(a) Balance property. We want roughly the same number of 0s as we have 1s in the sequence.

(b) Run property. A run sequence is a sequence of consecutive bits with the same value. We want

a. 1/2 of run sequences to be of length 1

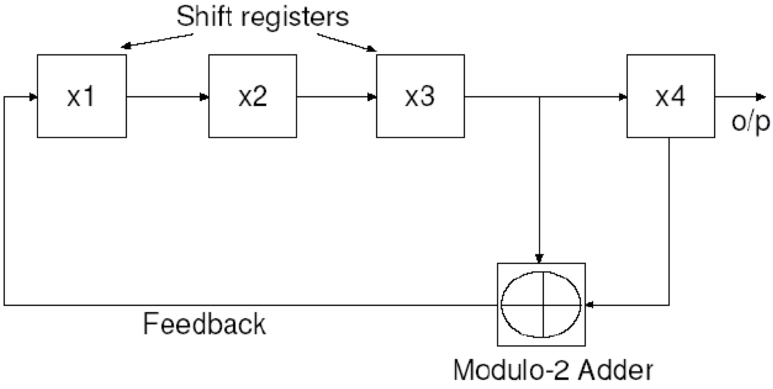
b. 1/4 of run sequences to be of length 2

c. 1/8 of run sequences to be of length 3…

(c) Correlation property. We would like the PN sequence to have very low correlation with shifted copies of itself (as with white noise). This also helps the receiver synchronise correctly.

a. ii) (3 marks)

PN sequences for spread spectrum systems can be produced using circuits consisting of shift registers with feedback. An example is given below.



b. (8 marks)

Message delay is defined as: $D = w + \tau$, where
w = the average waiting time before transmission of the message begins;
τ = transmission time.

Let us assume that each user wishes to transmit messages (packets) of b bits every T seconds. We further assume that b is chosen so that the channel is fully utilised, i.e. $Mb/T = R$. Also the time slots used for the TDMA system are T/M seconds in length.

(2 marks)

FDMA:
$w_{fdma} = 0$ (continuous transmission)
$\tau_{fdma} = T$
so $D_{fdma} = T$

(2 marks)

TDMA:
Each packet is sent within slots of T/M seconds in length, i.e. $\tau_{tdma} = T/M$.

(1 mark)

For the waiting time:
Packet 1 is transmitted immediately.
Packet 2 is transmitted T/M seconds later.
…
Packet m is transmitted $(m - 1)T/M$ seconds later.
The average waiting time is

$$w_{tdma} = \frac{1}{M}\sum_{m=1}^{M}(m-1)\frac{T}{M}$$

$$= \frac{T}{M^2}\sum_{m=1}^{M}(m-1)$$

$$= \frac{T}{M^2}\frac{(M-1)M}{2} = \frac{T}{2}(1-\frac{1}{M})$$

(2 marks)

So the message delay for TDMA

$$D_{tdma} = w_{tdma} + \tau_{tdma} = \frac{T}{2}(1-\frac{1}{M}) + \frac{T}{M}$$

(1 mark)

c. (6 marks)

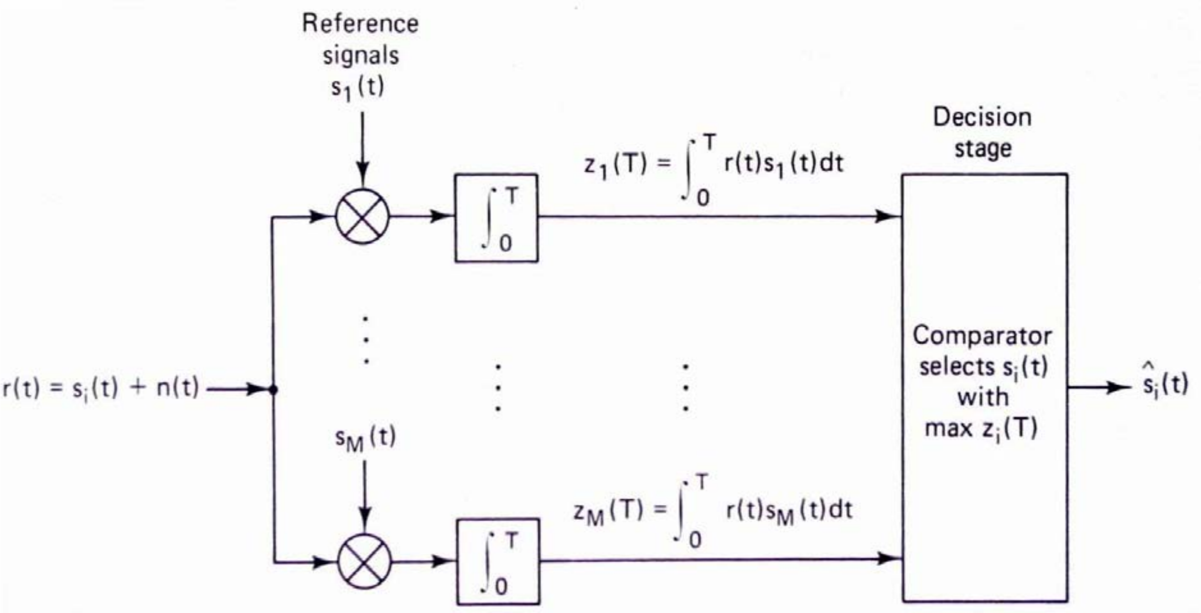
A correlation receiver works by comparing the incoming signal with a number of different reference signals and makes a decision about which symbol has been received based on which reference signal is most strongly correlated with the incoming signal.

(1 mark)

The correlation is computed by multiplying the incoming and reference signals and integrating over one symbol period.

(1 mark)

The structure of the receiver is illustrated below:


Reference signals $s_1(t)$
$z_1(T) = \int_0^T r(t)s_1(t)dt$
Decision stage
Comparator selects $s_i(t)$ with max $z_i(T)$
$r(t) = s_i(t) + n(t)$
$\hat{s}_i(t)$
$s_M(t)$
$z_M(T) = \int_0^T r(t)s_M(t)dt$
$\int_0^T$

(4 marks)