ENGINEERING TRIPOS PART IIB 2012 DR J SAYIR

4F5 Advanced Wireless Communications, 2012
Crib

Question 1

(a) We first compute the square of the norms

$$\|x_1(.)\|^2 = \int_{-\infty}^{\infty} |x_1(t)|^2 dt = \int_{-1}^{1} t^2 dt = \frac{1}{3}t^3\Big|_{-1}^{1} = \frac{2}{3},$$
$$\|x_2(.)\|^2 = \int_{-1}^{1} t^4 dt = \frac{1}{5}t^5\Big|_{-1}^{1} = \frac{2}{5},$$
$$\|x_3(.)\|^2 = \int_{-1}^{1} t^6 dt = \frac{1}{7}t^7\Big|_{-1}^{1} = \frac{2}{7},$$

and take the square root to obtain the norms

$$\|x_1(.)\| = \frac{\sqrt{6}}{3}, \|x_2(.)\| = \frac{\sqrt{10}}{5}, \|x_3(.)\| = \frac{\sqrt{14}}{7}$$

or

$$\|x_1(.)\| = 0.816, \|x_2(.)\| = 0.632, \|x_3(.)\| = 0.535.$$

(b)

$$\langle x_1(.), x_2(.)\rangle = \int_{-\infty}^{\infty} x_1(t)x_2^*(t)dt = \int_{-1}^{1} t \cdot t^2 dt = \frac{1}{4}t^4\Big|_{-1}^{1} = 0$$
$$\langle x_1(.), x_3(.)\rangle = \int_{-1}^{1} t \cdot t^3 dt = \frac{1}{5}t^5\Big|_{-1}^{1} = \frac{2}{5}$$
$$\langle x_2(.), x_3(.)\rangle = \int_{-1}^{1} t^2 \cdot t^3 dt = \frac{1}{6}t^6\Big|_{-1}^{1} = 0$$

(c) We apply the formula

$$\tilde{f}_i(.) = x_i - \sum_{k=1}^{i-1}\langle x_i(.), f_k(.)\rangle f_k(.) \quad \text{and} \quad f_i(.) = \frac{\tilde{f}_i(.)}{\|\tilde{f}_i(.)\|}.$$

The first basis vector can be constructed by normalisation $f_1(.) = x_1(.)/\|x_1(.)\|$. Since $\langle x_1(.), x_2(.)\rangle = 0$, the second basis vector can also be constructed by normalisation $f_2(.) = x_2(.)/\|x_2(.)\|$. For the third basis vector,

$$\begin{aligned}\tilde{f}_3(.) &= x_3(.) - \langle x_3(.), f_1(.)\rangle f_1(.) - \langle x_3, f_2(.)\rangle f_2(.)\\ &= x_3(.) - \frac{\langle x_3(.), x_1(.)\rangle}{\|x_1(.)\|^2}x_1(.) - \frac{\langle x_3(.), x_2(.)\rangle}{\|x_2(.)\|^2}x_2(.)\\ &= x_3(.) - \frac{2/5}{2/3}x_1(.)\\ \tilde{f}_3(t) &= t^3 - \frac{3}{5}t \text{ for } t \in [-1,1], 0 \text{ otherwise,}\end{aligned}$$

then compute the squared norm

$$\begin{aligned}\|\tilde{f}_3(.)\|^2 &= \int_{-1}^{1}\left(t^3 - \frac{3}{5}t\right)^2 dt = \int_{-1}^{1}\left(t^6 - \frac{6}{5}t^4 + \frac{9}{25}t^2\right) dt\\ &= \frac{1}{7}t^7 - \frac{6}{25}t^5 + \frac{3}{25}t^3\Big|_{-1}^{1} = \frac{2}{7} - \frac{6}{25} = \frac{8}{175}\end{aligned}$$

and $f_3(.) = \tilde{f}_3(.)/\|\tilde{f}_3(.)\|$.

In summary, we have, for $f_1(t)=f_2(t)=f_3(t)=0$ for $t \notin [-1,1]$, and for $t \in [-1,1]$,

$$f_1(t) = \frac{\sqrt{6}}{2}t, \; f_2(t) = \frac{\sqrt{10}}{2}t^2, \; \text{ and } \; f_3(t) = \frac{5\sqrt{14}}{4}\left(t^3 - \frac{3}{5}t\right),$$

or

$$f_1(t) = 1.225t, \; f_2(t) = 1.581t^2, \; \text{ and } \; f_3(t) = 4.677(t^3 - 0.6t).$$

(d) From the previous part, we know that

$$x_1(.) = \|x_1(.)\| f_1(.) = \frac{\sqrt{6}}{3}f_1(.)$$

and

$$x_2(.) = \|x_2(.)\| f_2(.) = \frac{\sqrt{10}}{5}f_2(.),$$

and

$$\begin{aligned}x_3(.) &= \tilde{f}_3(.) + \frac{3}{5}x_1(.) = \|\tilde{f}_3(.)\| f_3(.) + \frac{3}{5}\|x_1(.)\| f_1(.)\\ &= \sqrt{\frac{8}{175}}f_3(.) + \frac{3}{5}\frac{\sqrt{6}}{3}f_1(.).\end{aligned}$$

Thus, the vector representations of $x_1(.), x_2(.)$ and $x_3(.)$ are

$$\boldsymbol{x}_1 = (\frac{\sqrt{6}}{3}, 0, 0), \; \boldsymbol{x}_2 = (0, \frac{\sqrt{10}}{5}, 0), \; \boldsymbol{x}_3 = (\frac{\sqrt{6}}{5}, 0, \frac{2\sqrt{14}}{35}),$$

respectively, or

$$\boldsymbol{x}_1 = (0.816, 0, 0), \; \boldsymbol{x}_2 = (0, 0.632, 0), \; \boldsymbol{x}_3 = (0.145, 0, 0.214).$$

(e)

$$\|\boldsymbol{x}_1\|^2 = \left(\frac{\sqrt{6}}{3}\right)^2 = \frac{6}{9} = \|x_1(.)\|^2$$
$$\|\boldsymbol{x}_2\|^2 = \left(\frac{\sqrt{10}}{5}\right)^2 = \frac{2}{5} = \|x_2(.)\|^2$$
$$\|\boldsymbol{x}_3\|^2 = \left(\frac{\sqrt{6}}{5}\right)^2 + \left(\frac{2\sqrt{14}}{35}\right)^2 = \frac{2}{7} = \|x_3(.)\|^2.$$

These are equal to the squared norms computed in part (a) as expected.

The distance betweeen $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| = \left\|\left(\frac{\sqrt{6}}{3}, -\frac{\sqrt{10}}{5}, 0\right)\right\| = \sqrt{\frac{6}{9} + \frac{10}{25}} = \frac{4\sqrt{15}}{15},$$

or

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| = 1.033,$$

the distance between $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ is

$$\|\boldsymbol{x}_2 - \boldsymbol{x}_3\| = \left\|\left(-\frac{\sqrt{6}}{5}, \frac{\sqrt{10}}{5}, -\frac{2\sqrt{14}}{35}\right)\right\| = \sqrt{\frac{6}{25} + \frac{10}{25} + \frac{56}{35^2}} = \frac{2\sqrt{210}}{35},$$

or

$$\|\boldsymbol{x}_2 - \boldsymbol{x}_3\| = 0.828,$$

and the distance between $\boldsymbol{x}_3$ and $\boldsymbol{x}_1$ is

$$\|\boldsymbol{x}_3 - \boldsymbol{x}_1\| = \left\|\left(\frac{\sqrt{6}}{5} - \frac{\sqrt{6}}{3}, 0, \frac{2\sqrt{14}}{35}\right)\right\| = \frac{4\sqrt{105}}{105}$$

or

$$\|\boldsymbol{x}_3 - \boldsymbol{x}_1\| = 0.390.$$

(f) If M is the transmitted message, taking on values in the set $\{1, 2, 3\}$, and $\hat{M}$ is the reconstructed message at the receiver, the error probability for equiprobable messages is

$$
\begin{aligned}
P_e &= \sum_{m=1}^{3} \frac{1}{3} \Pr(\hat{M} \neq M | M = m) \\
&= \frac{1}{3} \sum_{m=1}^{3} \Pr\left(\bigcup_{m' \neq m} \{\hat{M} = m'\} \middle| M = m \right) \\
&\leq \frac{1}{3} \sum_{m=1}^{3} \sum_{m' \neq m} \Pr(\hat{M} = m' | M = m) \\
&= \frac{1}{3} \sum_{m=1}^{3} \sum_{m' \neq m} Q\left(\frac{\|\boldsymbol{x}_{m'} - \boldsymbol{x}_m\|}{\sqrt{2N_0}} \right) \\
&= \frac{2}{3} \left[Q\left(\frac{4\sqrt{15}}{15} \right) + Q\left(\frac{2\sqrt{210}}{35} \right) + Q\left(\frac{4\sqrt{105}}{105} \right) \right]
\end{aligned}
$$

or

$$
P_e \leq \frac{2}{3} \left[Q(1.033) + Q(0.828) + Q(0.390) \right]
$$

Examiner's comment:

A popular straightforward question requiring tedious integrations. A surprising number of students were unable to solve definite integrals (the integral of x^2 from -1 to 1!)

Question 2

(a) We have

$$H(XZ|Y) = H(X|Y) + H(Z|XY) = H(Z|Y) + H(X|YZ),$$

where $H(Z|XY) = 0$ because Y determines Z (since $f(.)$ is a deterministic function), and $H(Z|Y) = 0$ for the same reason. Therefore,

$$H(X|Y) = H(X|YZ).$$

The *data processing theorem* states that, since the random variables X, Y and Z form a Markov chain, i.e., $H(X|Y) = H(X|YZ)$, the mutual information $I(X;Z)$ can not exceed the mutual information $I(X;Y)$ or $I(Y;Z)$.

(b) We can compute

$$P_Z(1) = \sum_x \sum_y P_{XYZ}(x, y, 1) = (1-p)\delta + p\delta = \delta$$

and thus $P_Z(0) = 1 - \delta$ doesn't depend on p. It is easy to see that

$$H(X|Y, Z = 0) = 0$$

because, given that Y is *not* an erasure ($Z = 0$), the output of the channel determines the input with no uncertainty. Similarly, we conclude that

$$H(X|Y, Z = 1) = H(X) = H_2(p),$$

because an erasure gives no information about the channel input. Thus the entropy of the channel input given an erasure is equal to its entropy with no observation.

(c) The mutual information is

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) - H(X|Y, Z) \\ &= H_2(p) - H(X|Y, Z = 0)P_Z(0) - H(X|Y, Z = 1)P_Z(1) \\ &= H_2(p) - H_2(p)\delta = H_2(p)(1 - \delta). \end{aligned}$$

$H_2(p)$ and hence mutual information is maximised for $p = 1/2$, for which $I(X;Y) = 1 - \delta$. This is also the capacity of the Binary Erasure Channel (BEC) since, by Shannon's coding theorem, this is equal to the maximum of the mutual information over all input distributions.

(d) The capacity of this channel is $C = 1 - \delta = 0.4$ bits per use. We are asked to communicate at a rate $R = 0.5 > C$ above capacity. Hence, it is not possible to achieve arbitrary reliability. The error probability can be lower bounded using the converse to Shannon's theorem

$$P_e \geq 1 - \frac{C}{R} - \frac{1}{nR} = \frac{1}{5} - \frac{2}{n}$$

where n is the codeword length.

(e) We define a new function $f(y_1, y_2) = 1$ for $y_1 = y_2 = \Delta$ and $f(y_1, y_2) = 0$ for all other values of y_1 and y_2. Observe that

$$H(XZ|Y_1, Y_2) = H(X|Y_1, Y_2) + H(Z|X, Y_1, Y_2) = H(Z|Y_1, Y_2) + H(X|Z, Y_1, Y_2)$$

where $H(Z|Y_1, Y_2) = H(Z|X, Y_1, Y_2) = 0$, leading us to conclude that

$$H(X|Y_1, Y_2) = H(X|Y_1, Y_2, Z).$$

Thus, the mutual information of interest can be stated as

$$\begin{aligned} I(X; Y_1, Y_2) &= H(X) - H(X|Y_1, Y_2) = H(X) - H(X|Y_1, Y_2, Z) \\ &= H(X) - H(X|Y_1, Y_2, Z = 0)P_Z(0) - H(X|Y_1, Y_2, Z = 1)P_Z(1) \\ &= H_2(p) - H_2(p)P_Z(1), \end{aligned}$$

following from the fact that, if any or both of the channel outputs are non-erased ($Z = 0$), our uncertainty about the channel input is zero.

Since the two BECs are independent, the probability of them both being erased is $\delta_1\delta_2$, leading to the expression

$$I(X; Y_1, Y_2) = H_2(p)(1 - \delta_1\delta_2)$$

which is maximised for $p = 1/2$, for which we get the capacity of the channel with input X and output (Y_1, Y_2),

$$C = \max_{P_X} I(X; Y_1, Y_2) = 1 - \delta_1\delta_2.$$

(f) We start by analysing channel 2. If either Y_1 or Y_2 or both are erased, our uncertainty about X_2 is not reduced by observation of the channel outputs. Only when both are non-erased can we deduce X_2 with certainty. Let $Z = f(Y_1, Y_2)$ be an indicator function that takes the value 0 if both channel outputs are non-erased and 1 in any other case. We have $P_Z(0) = (1-\delta)^2$ and $P_Z(1) = 1-(1-\delta)^2$. As before, it is easy to show that $H(X_2|Y_1, Y_2) = H(X_2|Y_1, Y_2, Z)$. Thus,

$$\begin{aligned} I(X_2; Y_1, Y_2) &= H(X_2) - H(X_2|Y_1, Y_2, Z = 0)P_Z(0) - H(X_2|Y_1, Y_2, Z = 1)P_Z(1) \\ &= 1 - \left(1 - (1 - \delta)^2\right) \\ &= (1 - \delta)^2. \end{aligned}$$

As for channel 1, given X_2, the channel between X_1 and (Y_1, Y_2) is equivalent to the parallel BEC channel setup of the previous part with $\delta_1 = \delta_2 = \delta$ and uniform input, and thus

$$I(X_1; X_2, Y_1, Y_2) = 1 - \delta^2.$$

Examiner's comment:

An apparently simple but in fact difficult question that tested the fundamental understanding of concepts such as entropy, equivocation and mutual information.

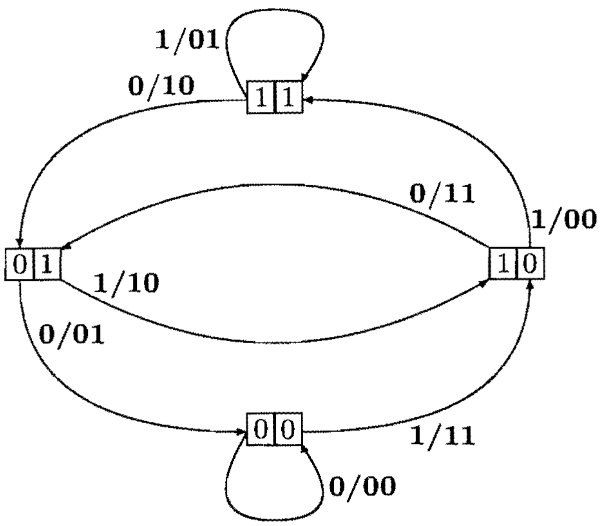
Question 3

(a) The convolutional encoder outputs two binary code digits for every source binary digit. Thus its rate is $R = 1/2$.

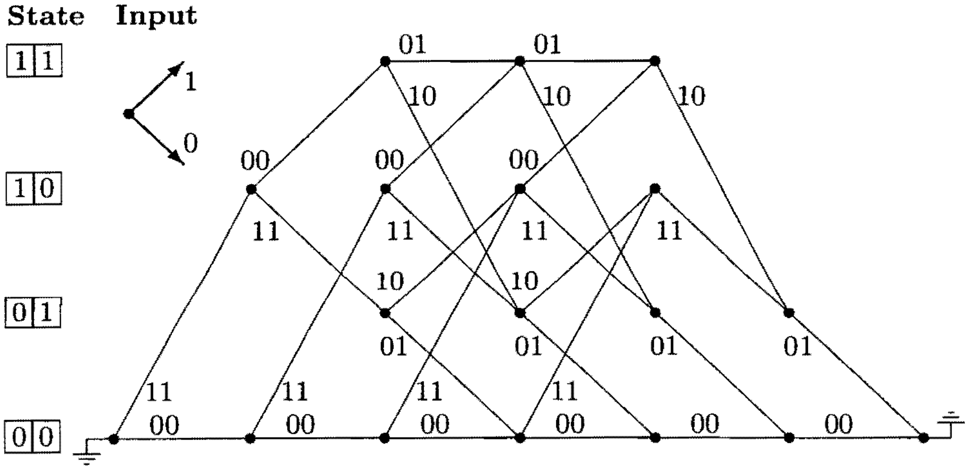
The binary connection polynomials for this encoder are $(1,1,0)$ and $(1,1,1)$. The generators in octal form for this encoder are thus $(6,7)_8$.

Since there is no universally accepted convention on whether to express connection polynomials from right to left or from left to right, an answer of $(3,7)_8$ will also be accepted if properly motivated.

(b) The encoder has two binary memory elements and therefore has $2^2 = 4$ states. The state diagram is drawn below.



(c) The trellis is drawn in the picture below. Two zero digits are required to return it to its initial state.



(d) If $\boldsymbol{X} = X_1, \ldots, X_n$ is the codeword and $\boldsymbol{Y} = Y_1, \ldots, Y_n$ is the sequence of channel outputs, then the maximum likelihood rule for picking the estimate $\hat{\boldsymbol{x}}$ of $\boldsymbol{X}$ for the observation $\boldsymbol{Y} = \boldsymbol{y} = (y_1, \ldots, y_n)$ is

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}).$$

For a memoryless channel, this can be rewritten as follows

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \arg\max_{(x_1,\ldots,x_n)} \prod_{i=1}^{n} P_{Y_i|X_i}(y_i|x_i) \\
&= \arg\max_{(x_1,\ldots,x_n)} \log\left(\prod_{i=1}^{n} P_{Y_i|X_i}(y_i|x_i)\right) \\
&= \arg\max_{(x_1,\ldots,x_n)} \sum_{i=1}^{n} \log\left(P_{Y_i|X_i}(y_i|x_i)\right) \\
&= \arg\max_{(x_1,\ldots,x_n)} \sum_{i=1}^{n} \log\left(\kappa + P_{Y_i|X_i}(y_i|x_i)\right),
\end{aligned}$$

where the last step follows because adding a constant κ to every term in the sum will shift the maximum by $n\kappa$, but will not affect in any way the value $\hat{\boldsymbol{x}}$ achieving that maximum.

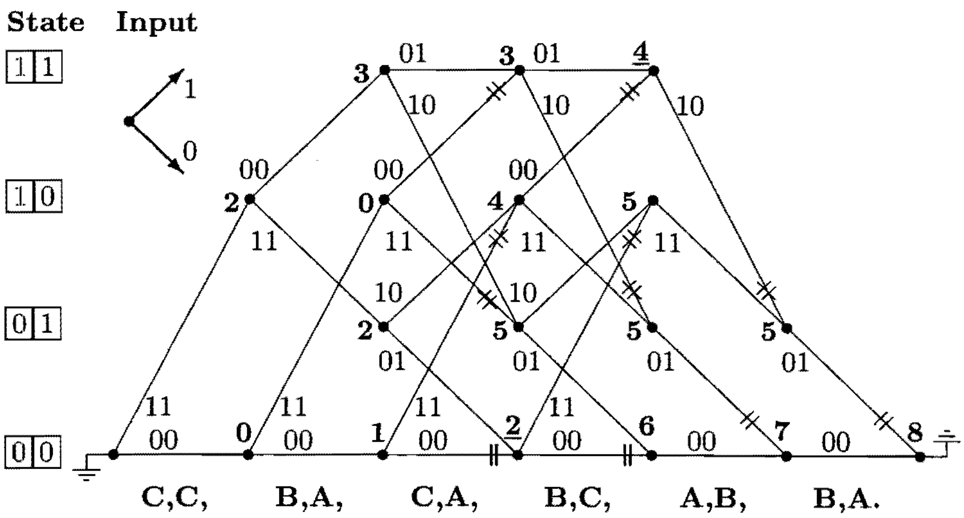
We write an initial logarithmic metric $\log P_{Y_i|X_i}(y_i|x_i)$ table without added constant for the channel as:

x_i \ y_i	A	B	C
0	-1	-2	-2
1	-2	-2	-1

Adding 2 to every metric in the table gives us a non-negative metric that is easier to work with for decoding:

x_i \ y_i	A	B	C
0	1	0	0
1	0	0	1

(e) We apply the Viterbi decoder to the trellis for this sequence as illustrated in the picture below:



Retracing our steps back through the trellis, we obtain the sequence

11 00 10 01 00 00

resulting in the maximum likelihood information sequence

1 1 0 0 (0 0)

where the last two digits are zero-stuffing and thus not part of the information sequence. Note that the two metric values underlined in the trellis correspond to ties between paths of equal metric, where the winning path was selected arbitrarily among the two incoming paths. Since none of these ties are to be found along the overall winning path, the maximum likelihood solution above is unique.

Examiner's comment:

An easy question that required the students to repeat operations that were treated extensively in examples during the lecture.

Question 4

(a) The product $B_c T_c$ of the coherence bandwidth B_c with the coherence time T_c determines whether the channel is underspread or overspread.

The coherence bandwidth B_c is simply the inverse of the delay spread T_d, i.e.,

$$B_c = \frac{1}{T_d} = \frac{1}{100 \times 10{-9}} = 10MHz.$$

The coherence time T_c is the inverse of the Doppler spread B_d, which, in Jakes' model, is equal to $2f_m$.[1]

We have

$$f_m = \frac{v f_c}{c}$$

where $v = 10$ km/h $= 10/3.6 = 2.78$ m/s is the velocity, c is the speed of light, and f_c is the carrier frequency, i.e.,

$$T_c = \frac{1}{B_d} = \frac{1}{2f_m} = \frac{c}{2vf_c} = \frac{3 \times 10^8}{2 \times 2.78 \times 5 \times 10^9} = 10.8 \text{ ms}.$$

Thus,

$$B_c T_c = 10^7 \times 1.08 \times 10^{-2} = 1.08 \times 10^5 >> 1$$

and therefore the channel is *underspread.*

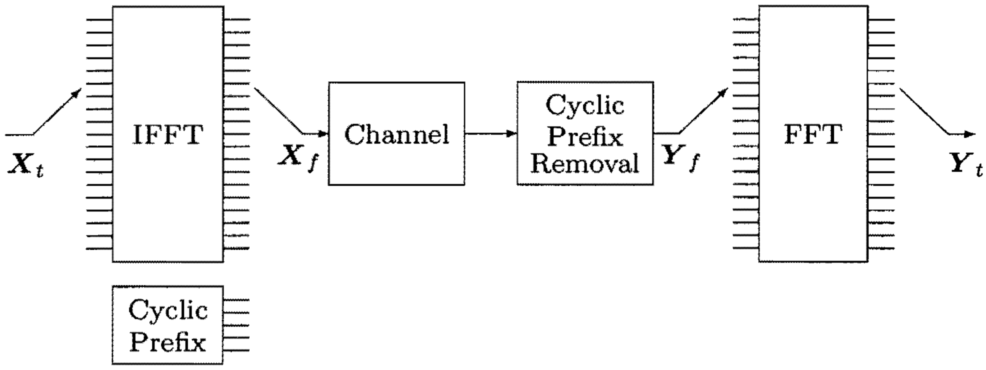
(b) The codeword duration of 3.5 μs is far smaller than the coherence time of $T_c = 10.8$ ms, thus the channel is *flat in time* (not selective.)

The signal bandwidth of 20 MHz is larger than the coherence bandwidth of $B_c = 10$ MHz, thus the channel is *selective in frequency.*

(c) (i) In a multipath wireless channel model, the signal emitted from the transmit antenna travels along several paths to the destination antenna. Along each of these paths, the signal is reflected by one or more scatterers. The scatterers give rise to paths of different lengths and hence taps in the discrete channel model of different gains and phases. Correlated scatterers result in paths of correlated lengths and hence in taps of correlated gains and phases. In order for the channel taps to be independent, the scatterers need to be uncorrelated.

(ii) The block diagram of OFDM is drawn below,

[1] Due to an inconsistency on this point in a previous exam crib, a Doppler spread of $B_d = f_m$ will also be accepted as a correct assumption and it does not change the solution.



where $\boldsymbol{X}_t$ is the signal vector in the "time domain", $\boldsymbol{X}_f$ is the signal vector transformed into the "frequency domain" via application of an inverse fast Fourier transform (IFFT). The last elements of the transformed block are repeated and appended to the beginning of the block as a "cyclic prefix" to ensure that the convolution applied by the channel is equivalent to a cyclic convolution in the frequency domain, equivalent to a multiplication in the time domain. The signal is then serialised and forwarded to a frequency-selective channel, which is modelled as a combination of a linear filter followed by an Additive White Gaussian Noise (AWGN) phase. At the receiver, the cyclic prefix is removed, and an FFT is performed to recover the signal in the time domain.

The result of these operations is to convert a frequency-selective channel into parallel flat channels, thereby replacing a channel with memory by parallel channels without memory.

(iii) We have, for $k \neq i$ and $k - i < n$,

$$\begin{aligned}
\mathrm{E}[H_i H_k^\star] &= \sum_{\ell=0}^{N_p-1} \sum_{m=0}^{N_p-1} \mathrm{E}[h_\ell h_m^\star] e^{-j2\pi\frac{i\ell}{n}} e^{+j2\pi\frac{km}{n}} \\
&= \sum_{\ell=0}^{N_p-1} \mathrm{E}[h_\ell h_\ell^\star] e^{j2\pi\frac{\ell(k-i)}{n}} \\
&= \sigma^2 \sum_{\ell=0}^{N_p-1} e^{j2\pi\frac{\ell(k-i)}{n}} \\
&= \sigma^2 \frac{1 - e^{j2\pi\frac{N_p}{n}(k-i)}}{1 - e^{j2\pi\frac{k-i}{n}}},
\end{aligned}$$

where the first step follows from the independence of the channel taps and from the fact that they have zero mean and the second step follows from the fact that they all have equal variance σ^2. In the last step, the expression in the denominator is non-zero because $0 < 2\pi\frac{k-1}{n} < 2\pi$ and the expression in the numerator is zero for all $k \neq i$ if and only if N_p is an integer multiple of n.

(iv) The linear code stated has block length 7 and rate $R = 4/7$. If OFDM is used with DFT length n equal to the code block length, then, referring to the previous part, we are now in the case where $N_p = n$ and thus N_p is an integer multiple of n and the DFT coefficients $H_1, \ldots, H_n$ are uncorrelated. In this case, the diversity achieved is simply the diversity of the code, which is equal to its minimum distance, or minimum Hamming weight.

The minimum distance of the code, and hence the diversity order achieved, is $d_{\min} = 3$. This can be shown through the following argument: the generator matrix is systematic, so any combination of 3 or more of its rows will have weight at least 3 because the systematic part will have weight at least 3; any combination of 2 rows in the generator matrix has weight at least 3 as well, because its systematic part has weight 2 and the parity-check part of two rows would have to be identical to achieve overall weight 2, which is not the case for any 2 rows; finally, there are 3 rows of weight 3 and thus the code has at least 3 codewords of weight 3.

(v) For $N_p = 2$, the DFT coefficients are correlated and this will reduce the diversity order of the system.