

These questions are based on the material in Section 1: Introduction to coding theory. You do not need to submit your answers to any of these questions.

1. The following ISBN was received with a smudge. What is the missing digit?

$$0 - 13 - 1x9139 - 9$$

2. Show that the following word is not a valid ISBN.

$$0 - 19 - 853803 - 2$$

Find a transposition error of adjacent digits that could have given this word.

3. Let  $C$  be a code of length 5, and suppose that we are transmitting codewords over a BSC with reliability  $p = 0.997$  and with randomly scattered noise.

- (a) For any codeword  $v \in C$ , what is the probability that  $v$  is received correctly?
- (b) Let  $v = 01100 \in C$  and  $x = 11100$ . What is  $\Phi_p(v, x)$ ?
- (c) Let  $v = 01100 \in C$  and  $w = 10101$ . What is  $\Phi_p(v, w)$ ?
- (d) For any codeword  $v \in C$ , if  $v$  is transmitted, what is the probability that a word is received which differs from  $v$  in one position (so one error has occurred)?

4. Explain why a channel with reliability  $p = 0$  is uninteresting.

5. What can be said about a channel with  $p = 0.5$ ?

6. What is the information rate of each of the following codes.

- (a)  $C = \{0000, 0101, 1010, 1111\}$
- (b)  $C = \{0000, 1110, 1111, 0101, 1010\}$
- (c)  $C = \{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$
- (d)  $C$  is a code of length 7 with 16 codewords.

7. Let  $C$  be the code whose codewords are all the words of length 3. Can  $C$  detect any errors? How many?

8. Let  $D$  be the code formed by adding a parity check digit to each codeword in the code  $C$  of Question 7, so that the number of ones in each codeword of  $D$  is even.

- (a) Write down the codewords in  $D$ .
- (b) Can  $D$  detect any errors? How many?
- (c) Suppose that the word 1101 is received. List the codeword(s) that were most likely to have been transmitted.

9. Let  $E$  be the 3-fold repetition code of length 9 formed by repeating each codeword in the code  $C$  of Question 7 three times.
  - (a) Can  $E$  detect any errors? How many?
  - (b) Suppose that the word 001000001 is received. List the codeword(s) that were most likely to have been transmitted.
  - (c) Suppose that the word 101100001 is received. List the codeword(s) that were most likely to have been transmitted.
10. Find the information rate for each of the codes  $C$ ,  $D$  and  $E$  from Questions 7–9.
11. Let  $C$  be a block code that can detect any single error.
  - (a) If  $C$  has length 4, what is the maximum number of codewords in  $C$ ?
  - (b) If  $C$  has length 5, what is the maximum number of codewords in  $C$ ?
  - (c) If  $C$  has length  $n$ , what is the maximum number of codewords in  $C$ ?
12. If  $C = \{010101, 110110, 101101, 100110, 011001\}$  and a word  $w = 101010$  is received, which codeword is most likely to have been sent?
13. Write down  $d(x, y)$  in each of the following cases:
  - (a)  $x = 1111, y = 0101$
  - (b)  $x = 0110, y = 1001$
  - (c)  $x = y = 1001$
  - (d)  $x = 11000, y = 10101$
  - (e)  $x = 1100101, y = 0111011$
14. From first principles, prove that Hamming distance is a metric (the properties of a metric are given on page 9 of the lecture notes).
15. Prove that  $wt(v + w) \leq wt(v) + wt(w)$ .
16. Consider the code  $C = \{001, 101\}$  and suppose that the codewords in  $C$  are sent across a BSC with reliability  $p = 0.9$ .
  - (a) Construct an IMLD table for  $C$ .
  - (b) For which received words will IMLD conclude that  $v = 001$  was transmitted?
  - (c) For which received words will IMLD conclude that  $v = 101$  was transmitted?
  - (d) If  $v = 001$  is transmitted, find the probability that IMLD will correctly conclude this after one transmission.
  - (e) If  $v = 101$  is transmitted, find the probability that IMLD will correctly conclude this after one transmission.

17. Consider the code  $C = \{000, 001, 110\}$  and suppose that the codewords in  $C$  are sent across a BSC with reliability  $p = 0.9$ .
- (a) For each word  $w \in K^3$  that could be received, find the word  $v \in C$  which IMLD will conclude was transmitted.
  - (b) For which received words will retransmission be requested?
  - (c) If  $v = 110$  is sent, find the probability that IMLD will correctly conclude this after one transmission.
  - (d) If  $v = 110$  is sent, find the probability that IMLD will conclude that 000 was sent.
18. Let  $C = \{001, 101, 110\}$ . Which of the error patterns 011 and 001 will  $C$  detect?
19. Let  $C = \{00000, 10101, 00111, 11100\}$ . Which of the error patterns 10101, 01010 and 11011 will  $C$  detect?
20. Let  $C = \{1101, 0110, 1100\}$ . Find all error patterns which can be detected by  $C$ .
21. Let  $C = \{1000, 0100, 0010, 0001\}$ . Find all error patterns which can be detected by  $C$ .
22. Find the distance of each of the following codes:
- (a)  $C = \{101, 111, 011\}$ .
  - (b)  $C = \{0000, 1001, 0110, 1111\}$ .
  - (c)  $C = \{00000, 11111\}$ .
  - (d)  $C = \{00000, 11100, 00111, 11011\}$ .
  - (e)  $C = \{00000, 11110, 01111, 10001\}$ .
  - (f)  $C = \{000000, 101010, 010101, 111111\}$ .
23. For each code in Question 22, if  $C$  is  $x$  error detecting, find  $x$ . In each case, find an error pattern of weight  $x + 1$  which  $C$  does not detect.
24. For each code in Question 22, if  $C$  is  $x$  error correcting, find  $x$ . In each case, find an error pattern of weight  $x + 1$  that  $C$  does not correct.

1. Let the missing digit be  $x$ . Then we require

$$1 \cdot 0 + 2 \cdot 1 + 3 \cdot 3 + 4 \cdot 1 + 5 \cdot x + 6 \cdot 9 + 7 \cdot 1 + 8 \cdot 3 + 9 \cdot 9 + 10 \cdot 9 \equiv 0 \pmod{11}$$

The left-hand side gives  $271 + 5x$ , so by inspection  $x = 3$ . (Note that  $271 \equiv 7 \pmod{11}$  so we require a value of  $x$  such that  $5x \equiv 4 \pmod{11}$ .)

2. The given number is not a valid ISBN, since

$$1 \cdot 0 + 2 \cdot 1 + 3 \cdot 9 + 4 \cdot 8 + 5 \cdot 5 + 6 \cdot 3 + 7 \cdot 8 + 8 \cdot 0 + 9 \cdot 3 + 10 \cdot 2 = 207 \equiv 9 \pmod{11}.$$

To obtain a check sum that is congruent to  $0 \pmod{11}$ , we need to either add 2 or subtract 9. Suppose that the digit  $a$  occurs in position  $i$  and the digit  $b$  occurs in position  $(i + 1)$ . These two digits contribute a total of  $ia + (i + 1)b$  to the check sum. If the digits are transposed, then the new contribution is

$$ib + (i + 1)a = ia + i(b - a) + (i + 1)b + (i + 1)(a - b) = ia + (i + 1)b + (a - b).$$

Thus to obtain an additional contribution of 2 to the check sum, we look for adjacent digits  $ab$  where  $a = b + 2$ . Thus a transposition of digits in positions 5 and 6 (the digits 53) gives the valid ISBN 0 – 19 – 835803 – 2. We check this:

$$1 \cdot 0 + 2 \cdot 1 + 3 \cdot 9 + 4 \cdot 8 + 5 \cdot 3 + 6 \cdot 5 + 7 \cdot 8 + 8 \cdot 0 + 9 \cdot 3 + 10 \cdot 2 = 209 \equiv 0 \pmod{11}.$$

3. (a)  $p^5 = (0.997)^5 = 0.985$ .

(b)  $v$  and  $x$  differ in one position so  $\Phi_p(v, x) = p^4(1 - p)^1 = 0.00296$ .

(c)  $v$  and  $w$  differ in three positions so  $\Phi_p(v, w) = p^2(1 - p)^3 = 2.68 \times 10^{-8}$ .

(b) There are  $\binom{5}{1}$  words which differ in one position from  $v$ . Thus the probability that a word is received which differs from the codeword sent in one position is  $\binom{5}{1}p^4(1 - p)^1 = 0.0148$ .

4. A channel with reliability  $p = 0$  would guarantee that every bit of the received word is different from the transmitted word. Thus, by inverting each bit of the received word, we would in fact have a perfect channel, and perfect channels are boring.
5. A channel with  $p = 0.5$  is useless. Every word has equal probability of being received, regardless of which word is sent.

6. (a)  $\frac{1}{2}$     (b)  $\frac{1}{4} \log_2 5 = 0.58$     (c)  $\frac{3}{4}$     (d)  $\frac{4}{7}$

7.  $C$  cannot detect any errors since every word of length 3 is a codeword.

8. (a)  $D = \{0000, 1001, 0101, 0011, 1100, 1010, 0110, 1111\}$
- (b)  $D$  can detect any single error, since a single error would give a word with an odd number of ones. It cannot detect any double errors.
- (c) We list the codewords that are closest to 1101. They are 1001, 0101, 1100 and 1111.
9. (a)  $E$  can detect any two errors. It can detect some sets of three errors, but there are sets of three errors which it cannot detect.
- (b) The unique codeword that is closest to 001000001 is 001001001.
- (c) The unique codeword that is closest to 101100001 is 101101101.
10. The information rate  $C$  is 1, for  $D$  is  $3/4$  and for  $E$  is  $3/9 = 1/3$ .
11. If  $C$  can detect any single error, then it cannot have any codewords that only differ by one bit. Thus if it has the zero word, then it cannot have any words that have weight 1. Similarly, if  $C$  contains an even weight word, then there are several words that cannot appear in  $C$  and they all have odd weight. The best you can do is to have either all the words of even weight, or all the words of odd weight.
- (a) There are  $2^4 = 16$  words of length 4. The maximum number of codewords in  $C$  is 8.
- (b) There are  $2^5 = 32$  words of length 5. Thus the maximum number of codewords in  $C$  is 16.
- (c) There are  $2^n$  words of length  $n$ . Thus the maximum number of codewords in  $C$  is  $2^{n-1}$ .
12. By comparing  $w = 101010$  to each of the codewords in  $C$ , we see that 100110 is the most likely codeword to have been sent.
13. (a)  $d(x, y) = 2$     (b)  $d(x, y) = 4$     (c)  $d(x, y) = 0$     (d)  $d(x, y) = 3$     (e)  $d(x, y) = 5$
14. The four properties of a metric are given on page 9 of the lecture notes.
- (1) If  $x, y \in K^n$ , then by definition  $d(x, y)$  is non-negative.
- (2) If  $d(x, y) = 0$  then each corresponding pair of bits of  $x$  and  $y$  are the same, thus  $x = y$ . If  $x = y$  then clearly  $d(x, y) = 0$ .
- (3) Clearly the number of positions in which  $x$  differs from  $y$  is the same as the number of positions in which  $y$  differs from  $x$ .

14. (4) Let  $x, y, z \in K^n$  and let  $x = x_1x_2 \dots x_n$ ,  $y = y_1y_2 \dots y_n$  and  $z = z_1z_2 \dots z_n$ . We work componentwise and show that for each  $i$ ,

$$d(x_i, z_i) \leq d(x_i, y_i) + d(y_i, z_i).$$

If  $x_i = z_i$ , then  $d(x_i, z_i) = 0$ . We know that  $d(x_i, y_i) \geq 0$  and  $d(y_i, z_i) \geq 0$  so

$$0 = d(x_i, z_i) \leq d(x_i, y_i) + d(y_i, z_i).$$

If  $x_i \neq z_i$ , then  $d(x_i, z_i) = 1$ . Without loss of generality, let  $x_i = 1$  and  $z_i = 0$ . Then  $y_i$  must be equal to either  $x_i$  or  $z_i$ . Thus exactly one of  $d(x_i, y_i)$  and  $d(y_i, z_i)$  is equal to 1, and the other is equal to 0, so we have

$$1 = d(x_i, z_i) \leq d(x_i, y_i) + d(y_i, z_i) = 1.$$

Putting these two possibilities together completes the proof

15. The quantity  $wt(v) + wt(w)$  gives the number of ones in  $v$  plus the number of ones in  $w$ . The value  $wt(v + w)$  is the number of ones in  $v + w$  which equals the number of ones in  $v$  plus the number of ones in  $w$  minus 2 times the number of places in which  $v$  and  $w$  both have a one. Thus  $wt(v + w) \leq wt(v) + wt(w)$ .

16. (a) The IMLD table is as follows.

received word $w$	error pattern		decode as
	001 + $w$	101 + $w$	
000	001	101	001
100	101	001	101
010	011	111	001
001	000	100	001
110	111	011	101
101	100	000	101
011	010	110	001
111	110	010	101

- (b) IMLD will conclude that  $v = 001$  was transmitted for the received words 000, 010, 001 and 011.  
 (c) IMLD will conclude that  $v = 101$  was transmitted for the received words 100, 110, 101 and 111.  
 (d)  $L(001) = \{000, 010, 001, 011\}$  so

$$\begin{aligned} \Theta_p(001, C) &= \Phi_p(001, 000) + \Phi_p(001, 010) + \Phi_p(001, 001) + \Phi_p(001, 011) \\ &= p^2(1-p) + p(1-p)^2 + p^3 + p^2(1-p) \\ &= (0.9)^3 + 2(0.9)^2(0.1) + (0.9)(0.1)^2 \\ &= 0.9 \end{aligned}$$

16. (e)  $L(101) = \{100, 110, 101, 111\}$  so

$$\begin{aligned} \Theta_p(101, C) &= \Phi_p(101, 100) + \Phi_p(101, 110) + \Phi_p(101, 101) + \Phi_p(101, 111) \\ &= p^2(1-p) + p(1-p)^2 + p^3 + p^2(1-p) \\ &= (0.9)^3 + 2(0.9)^2(0.1) + (0.9)(0.1)^2 \\ &= 0.9 \end{aligned}$$

17. (a) The IMLD table is as follows.

received word $w$	error pattern			decode as
	$000 + w$	$001 + w$	$110 + w$	
000	000	001	110	000
100	100	101	010	–
010	010	011	100	–
001	001	000	111	001
110	110	111	000	110
101	101	100	011	001
011	011	010	101	001
111	111	110	001	110

(b) Retransmission will be requested for the received words 100 and 010.

(c)  $L(110) = \{110, 111\}$  so

$$\begin{aligned} \Theta_p(110, C) &= \Phi_p(110, 110) + \Phi_p(110, 111) \\ &= p^3 + p^2(1-p) \\ &= (0.9)^3 + (0.9)^2(0.1) \\ &= 0.81 \end{aligned}$$

(d) IMLD only concludes that 000 was sent if 000 is received. Therefore we have to calculate the probability that 000 is received when  $v = 110$  is sent. So

$$\Phi_p(110, 000) = p(1-p)^2 = (0.9)(0.1)^2 = 0.009.$$

18. For each error pattern  $u$ , we need to check if  $v + u$  is a codeword for any  $v \in C$ . If there is a  $v \in C$  for which  $v + u \in C$ , then  $C$  does not detect  $u$ .

$$001 + 011 = 010 \notin C \quad 101 + 011 = 110 \in C$$

So  $C$  does not detect the error pattern 001.

$$001 + 001 = 000 \notin C \quad 101 + 001 = 100 \notin C \quad 110 + 001 = 111 \notin C$$

So  $C$  does detect the error pattern 001.

19. Since  $00000 + 10101 = 10101 \in C$ , the code  $C$  does not detect 10101.

$$00000 + 01010 = 01010 \notin C \quad 10101 + 01010 = 11111 \notin C$$

$$00111 + 01010 = 01101 \notin C \quad 11100 + 01010 = 10110 \notin C$$

So  $C$  does detect the error pattern 01010.

Since  $00111 + 11011 = 11100 \in C$ , the code  $C$  does not detect 11011.

20. From lectures we know that  $C$  cannot detect any error patterns which are the sum of a pair of codewords. Thus  $C$  cannot detect:

$$1101 + 1101 = 0000 \quad 1101 + 0110 = 1011 \quad 1101 + 1100 = 0001 \quad 0110 + 1100 = 1010$$

So  $C$  can detect the error patterns  $K^4 - \{0000, 1011, 0001, 1010\}$ .

21. From lectures we know that  $C$  cannot detect any error patterns which are the sum of a pair of codewords. Thus  $C$  cannot detect:

$$1000 + 1000 = 0000 \quad 1000 + 0100 = 1100 \quad 1000 + 0010 = 1010 \quad 1000 + 0001 = 1001$$

$$0100 + 0010 = 0110 \quad 0100 + 0001 = 0101 \quad 0010 + 0001 = 0011$$

So  $C$  can detect the error patterns  $K^4 - \{0000, 1100, 1010, 1001, 0110, 0101, 0011\}$ .

22. (a)  $\delta = 1$    (b)  $\delta = 2$    (c)  $\delta = 5$    (d)  $\delta = 3$    (e)  $\delta = 2$    (f)  $\delta = 3$

23. (a)  $C$  is 0 error-detecting. It cannot detect the error pattern 010 of weight 1.

(b)  $C$  is 1 error-detecting. It cannot detect the error pattern 1001 of weight 2.

(c)  $C$  is 4 error-detecting. It cannot detect the error pattern 11111 of weight 5.

(d)  $C$  is 2 error-detecting. It cannot detect the error pattern 11100 of weight 3.

(e)  $C$  is 1 error-detecting. It cannot detect the error pattern 10001 of weight 2.

(f)  $C$  is 2 error-detecting. It cannot detect the error pattern 101010 of weight 3.

24. (a)  $C$  is 0 error-correcting. It cannot correct the error pattern 010 of weight 1.

(b)  $C$  is 0 error-correcting. It cannot correct the error pattern 1000 of weight 1.

(c)  $C$  is 2 error-correcting. It cannot correct the error pattern 11100 of weight 3.

(d)  $C$  is 1 error-correcting. It cannot correct the error pattern 11000 of weight 2.

(e)  $C$  is 0 error-correcting. It cannot correct the error pattern 10000 of weight 1.

(f)  $C$  is 1 error-correcting. It cannot correct the error pattern 101000 of weight 2.

These questions are based on the material in Section 2: Linear Codes I. You do not need to submit your answers to any of these questions.

1. Determine whether the following code is linear.

$$C = \{0000, 1001, 1010, 0011, 1111\}$$

2. For each of the following sets  $S$ , list the elements of the linear code  $\langle S \rangle$ .
  - (a)  $S = \{010, 011, 111\}$
  - (b)  $S = \{0101, 1010, 1100\}$
  - (c)  $S = \{11000, 01111, 11110, 01010\}$
  - (d)  $S = \{10101, 01010, 11111, 00011, 10110\}$
  - (e)  $S = \{0001111, 0110101, 1010011, 1011100, 1100110\}$
3. For each of the linear codes  $\langle S \rangle$  in Question 2, state the distance of the linear code, and how many errors it can detect and correct.
4. Let  $S = \{0101, 1010, 1100\}$ . From first principles, find the dual code  $C^\perp = S^\perp$ .
5. For each set  $S$  in Question 2, use Algorithm 2.13 to find a basis  $B$  for the code  $C = \langle S \rangle$  and to find a basis  $B^\perp$  for the code  $C^\perp = S^\perp$ .
6. For each set  $S$  in Question 2, find the dimension of each code  $C = \langle S \rangle$  and each dual code  $C^\perp$ .
7. Verify that the following matrix is a generating matrix for the linear code  $C = \{0000, 1110, 1011, 0101\}$ .

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Give two other generating matrices for the code  $C$ .

8. Verify that the following matrix is a parity check matrix for the linear code  $C = \{0000, 1110, 1011, 0101\}$ .

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

9. Let  $S = \{001110, 100101, 110110, 101011\}$ .

- (a) Find a generating matrix in RREF for the linear code  $C = \langle S \rangle$ .  
 (b) Assign letters to the words in  $K^3$  as follows:

000	100	010	001	110	101	011	111
space	D	E	H	L	N	P	S

Using your generating matrix from part (a), encode the message SEND HELP.

- (c) State the distance of  $C$  and hence determine how many errors it can detect and correct.  
 (d) Find a parity check matrix for  $C$ .  
 (e) The three words 111101, 010010 and 101011 are received. Use the parity check matrix from part (d) to determine whether it is likely that errors have occurred in the transmission of these words.  
 (f) Assume that at most one error has occurred in each of the received words in part (e). By comparing the received words in part (e) to the words used in the encoding in part (b), determine the most likely intended 3-letter message.

10. Let  $C$  be the linear code

$$C = \{00000, 101100, 111010, 010110, 011001, 110101, 100011, 001111\}.$$

- (a) Find a generating matrix  $\mathbf{G}_C$  in RREF for  $C$ .  
 (b) Find a parity check matrix  $\mathbf{H}_C$  for  $C$ . Verify that  $v\mathbf{H}_C = \mathbf{0}$  for the codeword  $v = 100011$ .  
 (c) Find a generating matrix  $\mathbf{G}_{C^\perp}$  for  $C^\perp$ .  
 (d) Find a parity check matrix  $\mathbf{H}_{C^\perp}$  for  $C^\perp$ . Verify that  $\mathbf{G}_{C^\perp}\mathbf{H}_{C^\perp} = \mathbf{0}$ .  
 (e) State the dimensions of  $C$  and  $C^\perp$ .

11. Consider the code

$$C = \{00000, 10010, 10100, 00110\}.$$

- (a) Show that  $C$  is not a systematic code.  
 (b) Find a systematic code  $C'$  which is equivalent to  $C$ .

12. This question involves recovering messages encoded using a generating matrix in non-standard form.

(a) Let  $C_1$  be the code with generating matrix  $\mathbf{G}_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$ .

Let  $u$  be a message word of length 3, encoded using  $\mathbf{G}_1$ , and corresponding to the codeword  $v = u\mathbf{G}_1$ . If you received the codeword  $v$ , how would you recover the message word  $u$ ?

(b) Let  $C_2$  be the code with generating matrix  $\mathbf{G}_2 = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$ .

Let  $u$  be a message word of length 3, encoded using  $\mathbf{G}_2$ , and corresponding to the codeword  $v = u\mathbf{G}_2$ . If you received the codeword  $v$ , how would you recover the message word  $u$ ?

13. The following matrix  $\mathbf{H}$  is a parity check matrix for a linear code  $C$ .

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- (a) Determine the distance of  $C$ .  
 (b) Find a nonzero codeword of minimum weight in  $C$ .

14. Let  $C$  be the linear code with generating matrix  $\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$ .

- (a) Find a parity check matrix for  $C$ .  
 (b) List the cosets of  $C$ .  
 (c) Use parts (a) and (b) to construct an SDA for  $C$ .  
 (d) The matrix  $\mathbf{G}$  was used to encode an English message according to the following key:

000	100	010	001	110	101	011	111
space	A	E	I	C	D	H	M

The following message is received (over a noisy channel)

0111011001000100011010000.

Assuming that at most one error occurred in each word, determine the most likely intended message.

1. The code  $C$  is not linear since  $0011 \in C$  and  $1111 \in C$  but  $0011 + 1111 = 1100 \notin C$ .
2. (a)  $\langle S \rangle = \{000, 010, 011, 111, 100, 001, 101, 110\} = K^3$   
 (b)  $\langle S \rangle = \{0000, 0101, 1010, 1100, 1111, 1001, 0110, 0011\}$   
 (c)  $\langle S \rangle = \{00000, 11000, 01111, 11110, 01010, 10111, 00110, 10010, 10001, 00101, 10100, 01001, 11101, 01100, 11011, 00011\}$   
 (d)  $\langle S \rangle = \{00000, 10101, 01010, 11111, 00011, 10110, 01001, 11100\}$   
 (e)  $\langle S \rangle = \{0000000, 0001111, 0110101, 1010011, 1011100, 1100110, 0111010, 1101001\}$
3. (a)  $\delta = 1$ , no errors detected or corrected.  
 (b)  $\delta = 2$ , 1-error detecting, 0-error correcting.  
 (c)  $\delta = 2$ , 1-error detecting, 0-error correcting.  
 (d)  $\delta = 2$ , 1-error detecting, 0-error correcting.  
 (e)  $\delta = 4$ , 3-error detecting, 1-error correcting.
4. Let  $x_1x_2x_3x_4 \in C^\perp$ . Then by the definition of  $C^\perp$  we have

$$\begin{aligned} (x_1x_2x_3x_4) \cdot (0101) &= 0 && \text{so } x_2 + x_4 = 0 \\ (x_1x_2x_3x_4) \cdot (1010) &= 0 && \text{so } x_1 + x_3 = 0 \\ (x_1x_2x_3x_4) \cdot (1100) &= 0 && \text{so } x_1 + x_2 = 0 \end{aligned}$$

Thus we have  $x_1 = x_2 = x_3 = x_4$ . Thus

$$C^\perp = \{0000, 1111\}.$$

5. (a) We apply Algorithm 2.13.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

which is in RREF. So a basis for  $C$  is  $\{100, 010, 001\}$ . Note that this agrees with the fact that  $C = K^3$  as calculated in Question 2.

Since  $\mathbf{G} = \mathbf{I}_3$ , there is no matrix  $\mathbf{X}$  from which to create a parity check matrix. Note that  $k = 3$ ,  $n = 3$  and the number of columns of a parity check matrix is  $n - k = 0$ . Thus the dual code is  $C^\perp = \{000\}$  and so a basis for  $C^\perp$  is  $B^\perp = \emptyset$ .

5. (b) We apply Algorithm 2.13.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

which is in RREF. So a basis for  $C$  is  $\{1001, 0101, 0011\}$ .

Now  $\mathbf{G} = (\mathbf{I}_3 \ \mathbf{X})$  so  $\mathbf{H} = \begin{pmatrix} \mathbf{X} \\ \mathbf{I}_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ .

Thus a basis for  $C^\perp$  is  $\{1111\}$ . Note that this agrees with the code  $C^\perp$  whose codewords were calculated in Question 4.

(c) We apply Algorithm 2.13.

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \end{aligned}$$

which is in RREF. So a basis for  $C$  is  $\{10001, 01001, 00101, 00011\}$ .

Now  $\mathbf{G} = (\mathbf{I}_4 \ \mathbf{X})$  so  $\mathbf{H} = \begin{pmatrix} \mathbf{X} \\ \mathbf{I}_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ .

Thus a basis for  $C^\perp$  is  $\{11111\}$ .

(d) We apply Algorithm 2.13

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which is in RREF. So a basis for  $C$  is  $\{10101, 01001, 00011\}$ .

5. (d) Now  $\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ , so  $\mathbf{G}' = \left( \begin{array}{ccc|cc} 1 & 2 & 4 & 3 & 5 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{array} \right) = (\mathbf{I}_3 \ \mathbf{X})$ .

Thus  $\mathbf{H}' = \begin{pmatrix} \mathbf{X} \\ \mathbf{I}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix}$ , and so  $\mathbf{H} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$ .

Thus a basis for  $C^\perp$  is  $\{10100, 11011\}$ .

(e) We apply Algorithm 2.13.

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which is in RREF. So a basis for  $C$  is  $\{1010011, 0110101, 0001111\}$ .

Now  $\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$ , so  $\mathbf{G}' = \left( \begin{array}{cccc|cccc} 1 & 2 & 4 & 3 & 5 & 6 & 7 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{array} \right) = (\mathbf{I}_3 \ \mathbf{X})$ .

Thus  $\mathbf{H}' = \begin{pmatrix} \mathbf{X} \\ \mathbf{I}_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ 4 \\ 3 \\ 5 \\ 6 \\ 7 \end{matrix}$ , and so  $\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ .

Thus a basis for  $C^\perp$  is  $\{1110000, 0101100, 1001010, 1101001\}$ .

6. In each case, we have  $\dim C + \dim C^\perp = n$ .

	(a)	(b)	(c)	(d)	(e)
$\dim C$	3	3	4	3	3
$\dim C^\perp$	0	1	1	2	4

7. Each of the four codewords occurs as a linear combination of the rows of the matrix  $\mathbf{G}$ .

$$\begin{aligned} 0000 &= 0(1110) + 0(1011) & 1110 &= 1(1110) + 0(1011) \\ 1011 &= 0(1110) + 1(1011) & 0101 &= 1(1110) + 1(1011) \end{aligned}$$

Thus  $\mathbf{G}$  is a generating matrix for  $C$ .

Two other generating matrices for  $C$  are  $\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$  and  $\begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$ .

8. Let  $v = x_1x_2x_3x_4 \in C$ . If  $\mathbf{H}$  is a parity check matrix for  $C$  then  $v\mathbf{H} = \mathbf{0}$ . If

$$(x_1 \ x_2 \ x_3 \ x_4) \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = (0 \ 0),$$

then

$$x_1 + x_2 + x_4 = 0 \quad \text{and} \quad x_2 + x_3 + x_4 = 0.$$

These equations imply that  $x_1 = x_3$ . If  $x_1 = x_3 = 0$  then  $x_2 = x_4$  and these could be 0 or 1. If  $x_1 = x_3 = 1$  then exactly one of  $x_2$  or  $x_4$  is 1. Thus the words which satisfy  $v\mathbf{H} = \mathbf{0}$  are precisely the words of  $C$

$$\{0000, 0101, 1110, 1011\}.$$

Thus  $\mathbf{H}$  is a parity check matrix of  $C$ .

9. (a) We put the vectors in  $S$  into a matrix and reduce.

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus a generating matrix in RREF is

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

(b) Let  $u \in K^3$ . Then  $u$  is encoded as  $u\mathbf{G}$ .

S	E	N	D		H	E	L	P
111	010	101	100	000	001	010	110	011
111000	010011	101011	100101	000000	001110	010011	110110	011101

So the encoded message would be the bits (one after another and in order) that appear in the bottom row of the table.

9. (c) The distance of  $C$  is  $\delta = 3$ . Thus  $C$  is 2-error detecting and 1-error correcting.  
 (d) The matrix  $\mathbf{G}$  from part (a) is used in Algorithm 2.13. Thus

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is a parity check matrix for  $C$ .

- (e) Given a received word  $w$ , if  $w\mathbf{H} = \mathbf{0}$  then  $w$  is a codeword. Otherwise, an error has occurred.

$$(1 \ 1 \ 1 \ 1 \ 0 \ 1) \mathbf{H} = (1 \ 0 \ 1)$$

$$(0 \ 1 \ 0 \ 0 \ 1 \ 0) \mathbf{H} = (0 \ 0 \ 1)$$

$$(1 \ 0 \ 1 \ 0 \ 1 \ 1) \mathbf{H} = (0 \ 0 \ 0)$$

Thus errors have occurred in the received words 111101 and 010010 but 101011 is a codeword, so it is likely that no errors occurred in its transmission.

- (f) The received word 111101 is closest to the codeword 011101, corresponding to the letter P.

The received word 010010 is closest to the codeword 010011, corresponding to the letter E.

The received word 101011 is the codeword corresponding to the letter N.

Thus it is likely that the intended message was PEN.

10. (a) The dimension of  $C$  is  $k = 3$ , so a generating matrix will have three rows. Choose any three linearly independent codewords, place them in the rows of a matrix and put this matrix into RREF.

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Thus  $\mathbf{G}_C = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$  is a generating matrix for  $C$  in RREF.

- (b) We use Algorithm 2.13 to find a parity check matrix for  $C$ . We have

$$\mathbf{G}_C = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} = (\mathbf{I}_3 \ \mathbf{X})$$

10. (b) Thus we have  $k = 3$ ,  $n = 6$  and a parity check matrix for  $C$  is

$$\mathbf{H}_C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Let  $v = 100011$  and check that  $v\mathbf{H}_C = \mathbf{0}$ .

$$v\mathbf{H}_C = (1 \ 0 \ 0 \ 0 \ 1 \ 1) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (0 \ 0 \ 0)$$

(c) A generating matrix for  $C^\perp$  is the transpose of a parity check matrix for  $C$ . Thus

$$\mathbf{G}_{C^\perp} = \mathbf{H}_C^T = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

(d) A parity check matrix for  $C^\perp$  can be found either by applying Algorithm 2.13 to the generating matrix  $\mathbf{G}_{C^\perp}$  from part (c), or by taking the transpose of the generating matrix  $\mathbf{G}_C$  from part (a). Here we'll take the transpose of  $\mathbf{G}_C$ .

$$\mathbf{H}_{C^\perp} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Now we verify that  $\mathbf{G}_{C^\perp}\mathbf{H}_{C^\perp} = \mathbf{0}$ .

$$\mathbf{G}_{C^\perp}\mathbf{H}_{C^\perp} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(e) The dimensions of both  $C$  and  $C^\perp$  are 3.

11. (a) The code  $C$  is not systematic. The RREF of a generating matrix for a linear code is unique and the RREF generating matrix for  $C$  is

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

This matrix does not contain the  $2 \times 2$  identity matrix in its first two columns, and hence  $C$  is not systematic.

- (b) If we swap columns 2 and 3 in the generating matrix above we obtain a generating matrix

$$\mathbf{G}' = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

This gives the equivalent code  $C' = \{00000, 10010, 01010, 11000\}$ , which is systematic.

12. (a) Since  $\mathbf{G}_1$  is in RREF, the bits of the message word  $u$  will appear in the positions corresponding to the leading columns in the codeword  $v$ . Let  $u = x_1x_2x_3$  be a message word. Then

$$u\mathbf{G}_1 = (x_1 \ x_2 \ x_3) \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} = (x_1 \ x_2 \ x_1 + x_2 \ x_3 \ x_1 + x_3 \ x_1 + x_2).$$

Thus to recover  $u$  from  $v$ , take bits 1, 2 and 4 (in order) of  $v$ .

- (b) Since  $\mathbf{G}_2$  is not in RREF, we cannot recover the bits of  $u$  directly from the codeword  $v$ . However if we let  $u = x_1x_2x_3$  then we have

$$\begin{aligned} u\mathbf{G}_2 &= (x_1 \ x_2 \ x_3) \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \\ &= (x_1 \ x_1 + x_2 \ x_2 + x_3 \ x_1 + x_3 \ x_1 + x_2 \ x_1 + x_3) \end{aligned}.$$

Thus, if the codeword is  $v = v_1v_2v_3v_4v_5v_6$ , then  $x_1 = v_1$ ,  $x_2 = v_2 + x_1$  and  $x_3 = v_3 + x_2$ .

13. (a) In the parity check matrix  $\mathbf{H}$ , no single row is a linearly dependent set (no row of all zeros), but the set of rows 3 and 5 is a linearly dependent set (since row 3 plus row 5 gives zero). Thus by Theorem 2.42, the distance of  $C$  is 2.

- (b) Since rows 3 and 5 give a linearly dependent set, the word of length 6 with ones in positions 3 and 5 must be a codeword. To check this, we see that

$$(0 \ 0 \ 1 \ 0 \ 1 \ 0)\mathbf{H} = \mathbf{0}.$$

Thus 001010 is a word of minimum weight in  $C$ .

14. (a) Since we have  $\mathbf{G} = (\mathbf{I}_3 \ \mathbf{X})$ , Algorithm 2.13 gives a parity check matrix  $\mathbf{H}$  for  $C$  where

$$\mathbf{H} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

- (b) Since we have  $n = 5$  and  $k = 3$ , there are  $2^{5-3} = 4$  cosets of  $C$ , each containing 8 words. The cosets of  $C$  are:

$$\begin{aligned} C &= \{00000, 10001, 01011, 00101, 11010, 10100, 01110, 11111\} \\ 10000 + C &= \{10000, 00001, 11011, 10101, 01010, 00100, 11110, 01111\} \\ 01000 + C &= \{01000, 11001, 00011, 01101, 10010, 11100, 00110, 10111\} \\ 00010 + C &= \{00010, 10011, 01001, 00111, 11000, 10110, 01100, 11101\} \end{aligned}$$

- (c) An SDA for  $C$  is:

coset leader(s)	syndrome
00000	00
10000 or 00100 or 00001	01
00010	10
01000	11

- (d) First the received message is partitioned into received words of length 5. Then the syndrome of each received word is calculated, which determines the most likely error pattern(s). From this we can calculate the most likely corresponding codeword (and hence the English letter).

received word $w$	syndrome $w\mathbf{H}$	most likely error pattern(s)	corresponding codeword	English letter
01110	00	00000	01110	H
11001	11	01000	10001	A
00010	10	00010	00000	space
00110	11	01000	01110	H
10000	01	10000	00000	space
		00100	10100	D
		00001	10001	A

Thus the most likely message is HA H?, where the ? is one of space, D or A. Using the redundancy of the English language, we could guess that the intended message is HA HA.

These questions are based on the material in Section 3: Linear Codes II. You do not need to submit your answers to any of these questions.

1. Use Theorem 3.5 to show that there exists a linear code of length  $n = 5$ , dimension  $k = 2$  and distance  $\delta = 3$ . Use the method outlined in the proof of Theorem 3.5 to construct such a code.
2. Determine a lower and upper bound on the maximum number of codewords in a linear code with length  $n = 8$  and distance  $\delta = 3$ .
3. Prove Theorem 3.14: If  $C$  is a perfect code of length  $n$  and distance  $\delta = 2t + 1$  then  $C$  will correct all error patterns of weight less than or equal to  $t$ , and no other error patterns.
4. Let  $C$  be the linear code with generating and parity check matrices given by:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- (a) Find a generating matrix  $\mathbf{G}^*$  and a parity check matrix  $\mathbf{H}^*$  for the extended code  $C^*$  formed by adding a parity check bit to the start of each codeword of  $C$ , and verify that  $\mathbf{G}^*\mathbf{H}^* = \mathbf{0}$ .
- (b) What are the length, dimension and distance of  $C^*$ ?
- (c) Let  $B$  be the code with codewords

$$\{00000, 11000, 00111, 11111\}.$$

Create a new code  $D$  using the  $(a \mid a + b)$  construction, where  $a \in C$  and  $b \in B$ .

- (i) List the codewords of  $D$ .
  - (ii) What are the length, dimension and distance of  $D$ ?
  - (iii) Use the generating and parity check matrices for codes  $C$  and  $B$  to find a generating matrix  $\mathbf{G}_D$  and a parity check matrix  $\mathbf{H}_D$  for the code  $D$ .
  - (iv) Verify that the distance of  $D$  and the matrix  $\mathbf{H}_D$  satisfy Theorem 2.42.
5. Compare the length, number of codewords, distance and rate of the following codes:
    - the Hamming code of length 31;
    - the extended Hamming code of length 32;
    - the simplex code of length 31;
    - the first order Reed-Muller code of length 32;
    - the extended Golay code,  $C_{24}$ .

6. Let  $C$  be the code constructed from codes  $A$  and  $B$  using the  $(a \mid a + b)$  construction.
- (a) How many errors can be corrected by  $C$  if  $A$  is the Golay code  $C_{23}$  and  $B$  is the code of length 23 and dimension 1 containing the all zeros word and the all ones word?
  - (b) How many errors can be corrected by  $C$  if  $B$  is the Golay code  $C_{23}$  and  $A$  is the code of length 23 and dimension 1 containing the all zeros word and the all ones word?

7. Let  $C$  be the Hamming code of length 3, and  $C^\perp$  be the simplex code of length 3. List the codewords of  $C$  and  $C^\perp$ .
8. Use mathematical induction to prove property 4 of Theorem 3.37: For all  $r \geq 1$ ,  $RM(r - 1, m)$  is contained in  $RM(r, m)$ .
9. (a) Let  $Rep(n)$  be the  $n$ -fold repetition code with two codewords, the zero word and all ones word. Prove that

$$\mathbf{H}_{Rep(n)} = \begin{pmatrix} 1 \dots 1 \\ \mathbf{I}_{n-1} \end{pmatrix}$$

is a parity check matrix for  $Rep(n)$ .

- (b) Let  $\mathbf{H}_{r,m}$  be a parity check matrix for  $RM(r, m)$ . Prove that

$$\mathbf{H}_{1,m+1} = \begin{pmatrix} \mathbf{H}_{1,m} & \mathbf{H}_{Rep(2^m)} \\ \mathbf{0} & \mathbf{H}_{Rep(2^m)} \end{pmatrix}.$$

10. Suppose that a message was encoded using the first order Reed-Muller code  $RM(1, 3)$  with generating matrix  $\mathbf{G}_{1,3}$  and the word  $w = 11001000$  was received. Apply Algorithm 3.43 to determine the most-likely intended message word.
11. Suppose that a message was encoded using the first order Reed-Muller code  $RM(1, 2)$  with generating matrix  $\mathbf{G}_{1,2}$  and the word  $w = 1101$  was received. What does Algorithm 3.43 tell you about the most likely transmitted codeword?
12. Consider the extended Golay code  $C_{24}$  with generating matrix  $\mathbf{G} = \begin{pmatrix} \mathbf{I}_{12} & \mathbf{B} \end{pmatrix}$  and parity check matrix  $\mathbf{H} = \begin{pmatrix} \mathbf{I}_{12} \\ \mathbf{B} \end{pmatrix}$  as given in the lecture notes.
- (a) Decode the received word  $w = 111\ 000\ 100\ 100\ 000\ 101\ 001\ 001$ .
  - (b) Decode the received word  $w = 101\ 110\ 101\ 010\ 001\ 001\ 011\ 111$ .
13. Consider the Golay code  $C_{23}$  with generating matrix  $\mathbf{G} = \begin{pmatrix} \mathbf{I}_{12} & \hat{\mathbf{B}} \end{pmatrix}$  as given in the lecture notes.
- (a) Decode the received word  $w = 010\ 000\ 100\ 001\ 010\ 100\ 011\ 01$ .
  - (b) Decode the received word  $w = 001\ 000\ 110\ 010\ 001\ 101\ 010\ 10$ .

1. By Theorem 3.5 there exists a linear code of length  $n$ , dimension  $k$  and distance at least  $\delta$  if

$$\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\delta-2} < 2^{n-k}.$$

For the parameters given we have

$$\binom{4}{0} + \binom{4}{1} = 5 \quad \text{and} \quad 2^{5-2} = 8.$$

Since  $5 < 8$ , such a linear code does exist.

In the proof of Theorem 3.5, we construct an  $n \times (n - k)$  matrix which can then be used as a parity check matrix for our linear code. The goal is to construct a  $5 \times 3$  matrix in which no set of  $\delta - 1 = 2$  rows is linearly dependent, and in which there is a linearly dependent set of three rows. We can choose any word of length 3 for each row of the matrix, provided that we haven't chosen that word already. We must also ensure that the columns of the matrix are linearly independent. To do this, choose the first three rows of the matrix to be the rows of  $\mathbf{I}_3$ . For example, we could have the matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

No pair of rows form a linearly dependent set but rows 1, 2 and 4 (for example) do form a linearly dependent set. To construct a code with the required parameters, we apply Algorithm 2.13 in reverse to obtain the generating matrix

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Thus an example of a code with length 5, dimension 2 and distance 3 is:

$$C = \{00000, 11010, 01101, 10111\}.$$

(Note that there are many different choices for  $\mathbf{H}$  and hence many different codes with these parameters.)

2. By Theorem 3.2, we know that

$$|C| \leq \frac{2^8}{\binom{8}{0} + \binom{8}{1}} = 28.4.$$

Thus, since  $C$  is a linear code, we have  $|C| \leq 16$ .

2. By Corollary 3.6, we know that

$$|C| \geq \frac{2^7}{\binom{7}{0} + \binom{7}{1}} = 16.$$

Thus the upper and lower bounds are equal and so the maximum number of codewords in a linear code with length 8 and distance 3 is 16.

3. Suppose that  $C$  is a perfect code of length  $n$  and distance  $2t + 1$ . Since  $C$  is perfect we have

$$|C| = \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{t}}.$$

Since there are  $2^n$  words of length  $n$  and the cosets of  $C$  partition these words into sets of size  $|C|$ , there are precisely  $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{t}$  cosets of  $C$ .

Also since  $C$  has distance  $2t+1$ , we know that  $C$  is  $t$ -error correcting, that is, every error pattern of weight less than or equal to  $t$  must be correctable. This is only possible if every word of weight less than or equal to  $t$  is a unique coset leader. There are precisely  $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{t}$  such words.

So each word of weight less than or equal to  $t$  must be a coset leader and there are no other coset leaders. Thus,  $C$  will correct all error patterns of weight less than or equal to  $t$  and no other error patterns.

4. (a)

$$\mathbf{G}^* = ( \mathbf{b} \quad \mathbf{G} ) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

$$\mathbf{H}^* = \begin{pmatrix} 1 & 0 \dots 0 \\ j & \mathbf{H} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{H}^* = \begin{pmatrix} 0 \dots 0 & 1 \\ \mathbf{H} & j \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

In each case,  $\mathbf{G}^* \mathbf{H}^* = \mathbf{0}$ .

(b) Length  $n = 6$ , dimension  $k = 2$ , distance  $\delta = 4$ .

(c) (i) The codewords of  $D$  are:

000000000	000000111	0000011000	0000011111
1010110101	1010110010	1010101101	1010101010
0101101011	0101101100	0101110011	0101110100
1111011110	1111011001	1111000110	1111000001

- 4.(c) (ii)  $C$  is a  $(5, 2, 3)$  code and  $B$  is a  $(5, 2, 2)$  code, so  $D$  has length  $n = 10$ , dimension  $k = 4$  and distance  $\delta = \min\{6, 2\} = 2$ .

(iii)

$$\mathbf{G}_D = \begin{pmatrix} \mathbf{G}_C & \mathbf{G}_C \\ \mathbf{0} & \mathbf{G}_B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

$$\mathbf{H}_D = \begin{pmatrix} \mathbf{H}_C & \mathbf{H}_B \\ \mathbf{0} & \mathbf{H}_B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(iv)  $D$  has distance  $\delta = 2$  and rows 6 and 7 of  $\mathbf{H}_D$  form a linearly dependent set, while no single row is a linearly dependent set. Thus, Theorem 2.42 is satisfied.

5. The following table gives the length  $n$ , number of codewords  $|C|$ , distance  $\delta$ , and rate for each of the five codes.

	$n$	$ C $	$\delta$	rate
Hamming length 31	31	$2^{26}$	3	$26/31$
Ext Hamming length 32	32	$2^{26}$	4	$26/32$
Simplex length 31	31	$2^5$	16	$5/31$
Reed-Muller $RM(1, 5)$	32	$2^6$	16	$6/32$
Ext Golay $C_{24}$	24	$2^{12}$	8	$12/24$

The simplex and Reed-Muller codes give high error correction but not many codewords and a low information rate. The Hamming and extended Hamming codes give low error correction but many codewords and a high information rate. The extended Golay code is in the middle of the two extremes.

6. (a) The code  $C_{23}$  has  $\delta = 7$  and the  $(23, 1)$  repetition code  $Rep(23)$  has  $\delta = 23$ , so if  $a \in C_{23}$  and  $b \in Rep(23)$ , then the distance of  $C$  will be  $\min\{14, 23\} = 14$ . Thus the code  $C$  is 6-error correcting.
- (b) If  $a \in Rep(23)$  and  $b \in C_{23}$ , then the distance of  $C$  will be  $\min\{46, 7\} = 7$ . Thus the code  $C$  is 3-error correcting.

7. A parity check matrix for the Hamming code of length  $3 = 2^r - 1$ , so  $r = 2$  is

$$\mathbf{H} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus a generating matrix for the Hamming code of length 3 is  $\mathbf{G} = (1 \ 1 \ 1)$ . The simplex code of length 3 is the dual of the Hamming code of length 3, so the codewords of  $C^\perp$  are generated by the columns of  $\mathbf{H}$ . The codewords of  $C$  and  $C^\perp$  are thus

$$C = \{000, 111\} \quad \text{and} \quad C^\perp = \{000, 110, 101, 011\}.$$

8. By looking at the codes listed on page 51 of the lecture note, the statement is clearly true for  $m \leq 2$  and all  $1 \leq r \leq m$ .

Consider  $r = 1$ . We want to show that  $RM(0, m) \subseteq RM(1, m)$ . Since

$$\mathbf{G}_{1,m} = \begin{pmatrix} \mathbf{G}_{1,m-1} & \mathbf{G}_{1,m-1} \\ \mathbf{0} & \mathbf{G}_{0,m-1} \end{pmatrix}$$

and the top row of  $\mathbf{G}_{1,m-1}$  is the all ones vector, we see that  $RM(1, m)$  contains the all ones word (of length  $2^m$ ). Since  $RM(1, m)$  is a linear code, it must also contain the zero word (of length  $2^m$ ). But the code  $RM(0, m)$  contains only the zero word and the all ones word (each of length  $2^n$ ), so  $RM(0, m) \subseteq RM(1, m)$ .

Now suppose that for  $m' < m$  and any  $r'$  where  $1 \leq r' \leq m'$  we have that  $RM(r' - 1, m') \subseteq RM(r', m')$ . Also suppose that for  $r' < r$  we have  $RM(r' - 1, m) \subseteq RM(r', m)$ . We need to prove that  $RM(r - 1, m) \subseteq RM(r, m)$ . Since  $RM(r - 1, m - 1) \subseteq RM(r, m - 1)$ , we know that  $\mathbf{G}_{r-1,m-1}$  is a submatrix of  $\mathbf{G}_{r,m-1}$  (having the same number of columns). Also since  $RM(r - 2, m - 1) \subseteq RM(r - 1, m - 1)$ , we know that  $\mathbf{G}_{r-2,m-1}$  is a submatrix of  $\mathbf{G}_{r-1,m-1}$  (having the same number of columns). Thus

$$\mathbf{G}_{r-1,m} = \begin{pmatrix} \mathbf{G}_{r-1,m-1} & \mathbf{G}_{r-1,m-1} \\ \mathbf{0} & \mathbf{G}_{r-2,m-1} \end{pmatrix}$$

is a submatrix (with the same number of columns) of

$$\mathbf{G}_{r,m} = \begin{pmatrix} \mathbf{G}_{r,m-1} & \mathbf{G}_{r,m-1} \\ \mathbf{0} & \mathbf{G}_{r-1,m-1} \end{pmatrix}$$

and hence  $RM(r - 1, m) \subseteq RM(r, m)$ .

9. (a) The code  $Rep(n)$  has length  $n$  and dimension 1. Thus a parity check matrix for  $Rep(n)$  must be an  $n \times (n-1)$  matrix and must have linearly independent columns. The given matrix satisfies these conditions. Also, a generating matrix  $\mathbf{G}_{Rep(n)}$  for  $Rep(n)$  is the  $1 \times n$  matrix consisting of a single row of ones. Clearly

$$\mathbf{G}_{Rep(n)}\mathbf{H}_{Rep(n)} = (1 \ 1 \ \dots \ 1)\mathbf{H}_{Rep(n)} = \mathbf{0}.$$

Thus  $\mathbf{H}_{Rep(n)}$  is a parity check matrix for  $Rep(n)$ .

- (b) Since  $RM(1, m+1)$  is built from  $RM(1, m)$  and  $RM(0, m) = Rep(2^m)$  using the  $(a \mid a+b)$  construction, this result follows from Theorem 3.25 and part (b) above. We can verify that

$$\begin{aligned} \mathbf{G}_{1,m+1}\mathbf{H}_{1,m+1} &= \begin{pmatrix} \mathbf{G}_{1,m} & \mathbf{G}_{1,m} \\ 0 \dots 0 & 1 \dots 1 \end{pmatrix} \begin{pmatrix} \mathbf{H}_{1,m} & \mathbf{H}_{Rep(2^m)} \\ \mathbf{0} & \mathbf{H}_{Rep(2^m)} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{G}_{1,m}\mathbf{H}_{1,m} + \mathbf{0}\mathbf{G}_{1,m} & \mathbf{G}_{1,m}\mathbf{H}_{Rep(2^m)} + \mathbf{G}_{1,m}\mathbf{H}_{Rep(2^m)} \\ (0 \dots 0)\mathbf{H}_{1,m} + (1 \dots 1)\mathbf{0} & (0 \dots 0)\mathbf{H}_{Rep(2^m)} + (1 \dots 1)\mathbf{H}_{Rep(2^m)} \end{pmatrix} \end{aligned}$$

This equals  $\mathbf{0}$  since

- $\mathbf{G}_{1,m}\mathbf{H}_{1,m} = \mathbf{0}$ ,
- $\mathbf{G}_{1,m}\mathbf{H}_{Rep(2^m)} + \mathbf{G}_{1,m}\mathbf{H}_{Rep(2^m)} = \mathbf{0}$  (in binary arithmetic),
- $(1 \dots 1)\mathbf{H}_{Rep(2^m)} = (0 \dots 0)$  (every column of  $\mathbf{H}_{Rep(2^m)}$  has exactly two ones).

10. We apply Algorithm 3.43 to the received word  $w = 11001000$ . Thus  $\bar{w} = (1, 1, -1, -1, 1, -1, -1, -1)$ .

$$\begin{aligned} w_1 &= \bar{w}\mathbf{L}_3^1 = (2, 0, -2, 0, 0, 2, -2, 0) \\ w_2 &= w_1\mathbf{L}_3^2 = (0, 0, 4, 0, -2, 2, 2, 2) \\ w_3 &= w_2\mathbf{L}_3^3 = (-2, 2, 6, 2, 2, -2, 2, -2) \end{aligned}$$

The largest component of  $w_3$  is 6 occurring in position 2. Since  $v(2) = 010$  and  $6 > 0$ , the presumed message word is 1010 (with corresponding codeword 11001100).

11. We apply Algorithm 3.43 to the received word  $w = 1101$ . Thus  $\bar{w} = (1, 1, -1, 1)$ .

$$\begin{aligned} w_1 &= \bar{w}\mathbf{L}_2^1 = (2, 0, 0, -2) \\ w_2 &= w_1\mathbf{L}_2^2 = (2, -2, 2, 2) \end{aligned}$$

All four components of  $w_2$  have the same magnitude. Since

$$v(0) = 00, \quad v(1) = 10, \quad v(2) = 01 \quad \text{and} \quad v(3) = 11,$$

the following message words are all equally likely:

- 100 (with corresponding codeword 1111)    010 (with corresponding codeword 0101)
- 101 (with corresponding codeword 1100)    111 (with corresponding codeword 1001).

Thus Algorithm 3.43 gives us the four closest (equally likely) codewords to  $w$ .

12. (a) Call the received word  $w$ . The syndrome of  $w$  is

$$s = w\mathbf{H} = w \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix} = 111\ 000\ 100\ 100 + 110\ 011\ 011\ 101 = 001\ 011\ 111\ 001.$$

Since  $wt(s) \not\leq 3$  we look for a row of  $\mathbf{B}$  which differs from  $s$  in at most 2 places. We find that row 8 of  $\mathbf{B}$  differs from  $s$  in two places, so

$$u = [s + b_8, e_8] = 000\ 000\ 100\ 100\ 000\ 000\ 010\ 000.$$

Thus the most likely transmitted codeword is  $v = 111\ 000\ 000\ 000\ 000\ 101\ 011\ 001$ .

(b) Call the received word  $w$ . The syndrome of  $w$  is

$$s = w\mathbf{H} = w \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix} = 101\ 110\ 101\ 010 + 101\ 010\ 101\ 000 = 000\ 100\ 000\ 010.$$

Since  $wt(s) \leq 3$  we have

$$u = [s, 0] = 000\ 100\ 000\ 010\ 000\ 000\ 000\ 000.$$

Thus the most likely transmitted codeword is  $v = 101\ 010\ 101\ 000\ 001\ 001\ 011\ 111$ .

13. (a) Call the received word  $w$ . Since  $wt(w) = 8$ , we add a one to the end of  $w$ . The syndrome of  $w^* = w1$  is

$$s = w^*\mathbf{H} = w^* \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix} = 010\ 000\ 100\ 001 + 101\ 111\ 011\ 111 = 111\ 111\ 111\ 110.$$

This is the last row of  $\mathbf{H}$  so the received word is the most likely transmitted codeword.

(b) Call the received word  $w$ . Since  $wt(w) = 9$ , we add a zero to the end of  $w$ . The syndrome of  $w^* = w0$  is

$$s = w^*\mathbf{H} = w^* \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix} = 001\ 000\ 110\ 010 + 100\ 000\ 100\ 001 = 101\ 000\ 010\ 011.$$

Since  $wt(s) \not\leq 3$  we look for a row of  $\mathbf{B}$  which differs from  $s$  in at most 2 places. No such row exists. We now calculate the second syndrome

$$s\mathbf{B} = 000\ 100\ 010\ 010.$$

Since  $wt(s\mathbf{B}) \leq 3$ , we have

$$u = [0, s\mathbf{B}] = 000\ 000\ 000\ 000\ 000\ 100\ 010\ 010.$$

Thus the corresponding codeword in  $C_{24}$  is  $v^* = 001\ 000\ 110\ 010\ 001\ 001\ 000\ 110$ , so the most likely transmitted codeword is  $v = 001\ 000\ 110\ 010\ 001\ 001\ 000\ 11$ .

These questions are based on the material in Section 4: Cyclic Codes. You do not need to submit your answers to any of these questions.

1. Let  $f(x) = x + x^2 + x^3 + x^6 + x^7$  and  $p(x) = 1 + x + x^4$ .
  - (a) Find the quotient  $q(x)$  and the remainder  $r(x)$  so that  $f(x) = q(x)p(x) + r(x)$ .
  - (b) Determine the binary word of length 4 corresponding to  $f(x) \bmod p(x)$ .
2. Let  $C$  be the smallest cyclic linear code containing the word 011011.
  - (a) List the codewords of  $C$ , both as binary words and as polynomials.
  - (b) Determine the generator  $g(x)$  of  $C$ .
  - (c) Find a generating matrix for  $C$  using Corollary 5.28.
  - (d) Use  $g(x)$  to encode the message word 11. Use  $g(x)$  to determine the message word corresponding to the codeword 011011.
3. Show that a cyclic linear code of length  $n$  corresponds to an ideal in the ring of polynomial residues modulo  $x^n + 1$ .
4.
  - (a) How many cyclic linear codes of length 7 have 8 codewords? Explain your answer.
  - (b) How many cyclic linear codes of length 7 have 32 codewords? Explain your answer.
  - (c) Determine the most number of codewords in a *proper* cyclic linear code of length 7.
5. Let  $C$  be the cyclic linear code of length 15 generated by

$$g(x) = 1 + x^2 + x^4 + x^5.$$

- (a) Use Theorem 4.27 to find a generating matrix in standard form for  $C$ .
- (b) Find a parity check matrix for  $C$ .
- (c) Show that  $C$  is a 2 cyclic burst error correcting code.
- (d) Is  $C$  a 2-error correcting code? Explain your answer.
- (e) Assuming that codewords are being sent through a channel where burst errors are likely and assuming that a cyclic burst error of length at most 2 has occurred in each of the received words, use the method of Example 4.40 to decode each of the following received words.
  - (i)  $w = 000\ 101\ 000\ 011\ 101$
  - (ii)  $w = 100\ 010\ 001\ 101\ 000$

*continues overleaf*

6. The codes  $C_1$  and  $C_2$  of Example 4.48 were used to encode a message by cross interleaving and then having the codewords of  $C_2$  interleaved to depth 3. The following string of digits is received:

100011001111101010011001111010100110100100011101000100...

Determine the most likely messages  $m_1$ ,  $m_2$  and  $m_3$ .

7. Theorem 4.46 says that if  $C$  is an  $l$  burst error correcting code, interleaved to depth  $s$ , then all bursts of length at most  $sl$  will be corrected, provided that each codeword is affected by at most one burst of errors. Determine the required length of error free transmission between burst of errors to ensure that each codeword is affected by at most one burst of errors.

1. (a) The quotient and remainder are  $q(x) = 1 + x^2 + x^3$  and  $r(x) = 1 + x^3$ .

$$\begin{array}{r}
 x^4 + x + 1 \overline{) \begin{array}{r} x^7 + x^6 + x^3 + x^2 + x + 1 \\ x^7 + x^4 + x^3 \\ \hline x^6 + x^4 + x^2 + x \\ x^6 + x^3 + x^2 \\ \hline x^4 + x^3 + x \\ x^4 + x + 1 \\ \hline x^3 + 1 \end{array} \\
 \hline
 \end{array}$$

(b) By part (a) we have  $f(x) = 1 + x^3 \pmod{p(x)}$ , so the binary word corresponding to  $f(x)$  modulo  $p(x)$  is 1001.

2. (a) Since  $C$  is cyclic it must contain all cyclic shifts of 011011, these are 101101 and 110110. Since  $C$  is linear, it must contain all sums of two or more of these words. However  $011011 + 101101 = 110110$  so we do not obtain any new codewords apart from the zero word. Thus  $C = \{000000, 011011, 101101, 110110\}$  or, equivalently  $C = \{0, x + x^2 + x^4 + x^5, 1 + x^2 + x^3 + x^5, 1 + x + x^3 + x^4\}$ .

(b) The generator of  $C$  is  $g(x) = 1 + x + x^3 + x^4$ .

(c) Using Corollary 4.24, a generating matrix for  $C$  is

$$\mathbf{G} = \begin{pmatrix} g(x) \\ xg(x) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

(d) The message word 11 is equivalent to the message polynomial  $a(x) = 1 + x$ . Thus it encodes to 101101 since

$$a(x)g(x) = (1+x)(1+x+x^3+x^4) = 1+x+x+x^2+x^3+x^4+x^4+x^5 = 1+x^2+x^3+x^5.$$

The codeword 011011 is equivalent to the polynomial  $c(x) = x + x^2 + x^4 + x^5$ . The corresponding message word is found by determining  $c(x)/g(x)$ . Since

$$\frac{x + x^2 + x^4 + x^5}{1 + x + x^3 + x^4} = x$$

the corresponding message word is 01.

3. Let  $C$  be a cyclic linear code of length  $n$ . Each codeword of  $C$  corresponds to a unique polynomial in the ring of polynomial residues modulo  $1 + x^n$ . Call this ring  $R$  with operations addition and multiplication (modulo  $1 + x^n$ ).

$C$  is a linear code, so the set of polynomials in  $C$  with the operation addition is a subgroup of the abelian group  $(R, +)$ . The set is closed under addition, contains the identity (zero word) and each codeword is its own inverse.

3. Let  $r = \alpha_0 + \alpha_1x + \alpha_2x^2 + \dots + \alpha_{n-1}x^{n-1}$  be an arbitrary element of the ring  $R$ , and let  $f(x) \in C$ . Then we must show that  $rf(x)$  is a codeword of  $C$ . (Note that multiplication is commutative so  $rf(x) = f(x)r$ .) Now  $C$  is a cyclic code, so if  $f(x) \in C$ , then  $xf(x) \in C$ . Thus for any  $x^i$  ( $0 \leq i \leq n-1$ ) and  $f(x) \in C$ , the polynomial  $x^i f(x) \in C$ . Thus

$$rf(x) = \alpha_0 f(x) + \alpha_1 x f(x) + \alpha_2 x^2 f(x) + \dots + \alpha_{n-1} x^{n-1} f(x)$$

is a linear combination of polynomials in  $C$  and hence, since  $C$  is linear,  $rf(x) \in C$ . Thus the conditions of Definition 4.18 are satisfied.

4. The factorisation of  $1 + x^7$  into irreducible polynomials is

$$1 + x^7 = (1 + x)(1 + x + x^3)(1 + x^2 + x^3).$$

Thus the possible generators for a cyclic linear code of length 7 are:

$$\begin{array}{ccccccc} 1 & (1 + x) & (1 + x + x^3) & (1 + x^2 + x^3) & (1 + x)(1 + x + x^3) & & \\ (1 + x)(1 + x^2 + x^3) & (1 + x + x^3)(1 + x^2 + x^3) & (1 + x)(1 + x + x^3)(1 + x^2 + x^3) & & & & \end{array}$$

(a) A linear code with 8 codewords has dimension  $k = 3$ . A cyclic linear code of length  $n = 7$  and dimension  $k = 3$  would have a generator of degree  $n - k = 4$ . There are two possible generators of degree 4 listed above, thus there are two cyclic linear codes of length 7 with 8 codewords.

(b) A linear code with 32 codewords has dimension  $k = 5$ . A cyclic linear code of length 7 and dimension  $k = 5$  would have a generator of degree  $n - k = 2$ . None of the generators listed above has degree 2, thus no cyclic linear code of length 7 has 32 codewords.

(c) To find the most number of codewords in a proper cyclic linear code of length 7, we need to identify the generator (not 1) with the lowest degree (since that will give a large value for  $k$ ). The cyclic linear code of length 7 with generator  $1 + x$  has dimension  $k = 6$ , thus the most number of codewords in a proper cyclic linear code of length 7 is  $2^6 = 64$ .

5. (a) Here  $n = 15$  and  $k = 10$ . To find a generating matrix in standard form for  $C$  we first calculate  $r_5, r_6, \dots, r_{14}$ .

$$\begin{array}{ll} x^5 = 1 + x^2 + x^4 & r_5 = 10101 \\ x^6 = x + x^3 + (1 + x^2 + x^4) = 1 + x + x^2 + x^3 + x^4 & r_6 = 11111 \\ x^7 = x + x^2 + x^3 + x^4 + (1 + x^2 + x^4) = 1 + x + x^3 & r_7 = 11010 \\ x^8 = x + x^2 + x^4 & r_8 = 01101 \\ x^9 = x^2 + x^3 + (1 + x^2 + x^4) = 1 + x^3 + x^4 & r_9 = 10011 \\ x^{10} = x + x^4 + (1 + x^2 + x^4) = 1 + x + x^2 & r_{10} = 11100 \\ x^{11} = x + x^2 + x^3 & r_{11} = 01110 \\ x^{12} = x^2 + x^3 + x^4 & r_{12} = 00111 \\ x^{13} = x^3 + x^4 + (1 + x^2 + x^4) = 1 + x^2 + x^3 & r_{13} = 10110 \\ x^{14} = x + x^3 + x^4 & r_{14} = 01011 \end{array}$$

5. Thus a generating matrix for  $C$  is  $\mathbf{G}$  and a parity check matrix (part (b)) is  $\mathbf{H}$  as shown below.

$$\mathbf{G} = \left( \begin{array}{cccccccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{array} \right)$$

$$\mathbf{H} = \left( \begin{array}{ccccc} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

(c) For  $C$  to be a 2 cyclic burst error correcting code, each word in  $C(B_{15}(2))$  must be in a different coset. There are  $2^{n-k} = 32$  different cosets, and there are  $1 + 15 + 15 = 31$  words in  $C(B_{15}(2))$ , so this is possible.

To verify that  $C$  is a 2 cyclic burst error correcting code, we show that every word in  $C(B_{15}(2))$  has a unique syndrome. To calculate syndromes we use  $\mathbf{H}$  from part (b).

Error pattern	Syndrome	Error pattern	Syndrome
0	00000	$1 + x$	01010
1	10101	$x + x^2$	00101
$x$	11111	$x^2 + x^3$	10111
$x^2$	11010	$x^3 + x^4$	11110
$x^3$	01101	$x^4 + x^5$	01111
$x^4$	10011	$x^5 + x^6$	10010
$x^5$	11100	$x^6 + x^7$	01001
$x^6$	01110	$x^7 + x^8$	10001
$x^7$	00111	$x^8 + x^9$	11101
$x^8$	10110	$x^9 + x^{10}$	11011
$x^9$	01011	$x^{10} + x^{11}$	11000
$x^{10}$	10000	$x^{11} + x^{12}$	01100
$x^{11}$	01000	$x^{12} + x^{13}$	00110
$x^{12}$	00100	$x^{13} + x^{14}$	00011
$x^{13}$	00010	$1 + x^{14}$	10100
$x^{14}$	00001		

Each syndrome is different so  $C$  is a 2 cyclic burst error correcting code.

5. (d) For  $C$  to be a 2-error correcting code, each error pattern of weight 0, 1 or 2 would have to be in a distinct coset. There are 32 cosets of  $C$ , but there are

$$\binom{15}{0} + \binom{15}{1} + \binom{15}{2} = 1 + 15 + 105 = 121$$

error patterns of weight up to and including 2. Therefore it is not possible for  $C$  to be a 2-error correcting code.

- (e) (i) We first calculate the syndrome of  $w$ .

$$s = w\mathbf{H} = 01100, \text{ so } s(x) = x + x^2.$$

This syndrome contains a burst error pattern with burst length 2, so we have  $e = 000000000001100$ , and the most likely transmitted codeword is

$$v = w + e = 000\ 101\ 000\ 010\ 001.$$

- (ii) We first calculate the syndrome of  $w$ .

$$s = w\mathbf{H} = 10011, \text{ so } s(x) = 1 + x^3 + x^4.$$

Now apply Lemma 4.38 to calculate  $s_i$  for  $i = 1, 2, \dots$  until a syndrome is found that contains a burst error pattern with burst length at most 2.

$$\begin{array}{ll} s_1(x) = xs(x) = x + x^4 + (1 + x^2 + x^4) = 1 + x + x^2 & s_1 = 11100 \\ s_2(x) = xs_1(x) = x + x^2 + x^3 & s_2 = 01110 \\ s_3(x) = xs_2(x) = x^2 + x^3 + x^4 & s_3 = 00111 \\ s_4(x) = xs_3(x) = x^3 + x^4 + (1 + x^2 + x^4) = 1 + x^2 + x^3 & s_4 = 10110 \\ s_5(x) = xs_4(x) = x + x^3 + x^4 & s_5 = 01011 \\ s_6(x) = xs_5(x) = x^2 + x^4 + (1 + x^2 + x^4) = 1 & s_6 = 10000 \end{array}$$

Since  $s_6$  is a burst error pattern with burst of length 1 (less than 2), we have  $e_6 = 000000000010000$  and hence  $e = 000010000000000$ . Thus the most likely transmitted codeword is

$$v = w + e = 100\ 000\ 001\ 101\ 000.$$

6. We write the string of received digits in three rows by writing column 1 first, then column 2 etc... The rows then give us received words for the code  $C_2$ .

$$\begin{array}{lll} c_a = 100110 & c_d = 001011 & c_g = 110101 \\ c_b = 010101 & c_e = 101101 & c_h = 001000 \\ c_c = 011110 & c_f = 111000 & c_i = 001100 \end{array}$$

We now use the parity check matrix  $\mathbf{H}$  to determine which of the words above are codewords of  $C_2$ .

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{array}{lll} c_a \mathbf{H} = 000, & c_d \mathbf{H} = 000, & c_g \mathbf{H} = 110, \\ c_b \mathbf{H} = 000, & c_e \mathbf{H} = 000, & c_h \mathbf{H} = 011, \\ c_c \mathbf{H} = 000, & c_f \mathbf{H} = 000, & c_i \mathbf{H} = 111. \end{array}$$

6. Thus the first six are codewords but the last three are not. We flag all 18 digits of the last three words.

Since the generating matrix for  $C_2$  is in standard form, we obtain the following message words

$$\begin{array}{lll} c_a \rightarrow 100 & c_d \rightarrow 001 & c_g \rightarrow * * * \\ c_b \rightarrow 010 & c_e \rightarrow 101 & c_h \rightarrow * * * \\ c_c \rightarrow 011 & c_f \rightarrow 111 & c_i \rightarrow * * * \end{array}$$

We write these message words in columns, so that the rows give us the received words for the code  $C_1$ .

$$\begin{array}{ll} c_1 = 100011 * * & \text{which corresponds to the codeword } 10001110 \\ c_2 = 011001 * * & \text{which corresponds to the codeword } 01100110 \\ c_3 = 001111 * * & \text{which corresponds to the codeword } 00111100 \end{array}$$

Again, since the generating matrix for  $C_1$  is in standard form, we can determine that the most likely first three message words were

$$m_1 = 1000 \quad m_2 = 0110 \quad m_3 = 0011.$$

7. The process of interleaving to depth  $s$  is illustrated below. Suppose we are transmitting codewords  $c_1, \dots, c_s, c_{s+1}, \dots, c_{2s} \dots$  interleaved to depth  $s$ .

$$\begin{array}{ll} c_1 = (c_{1,1}, c_{1,2}, \dots, c_{1,n}) & c_{s+1} = (c_{s+1,1}, c_{s+1,2}, \dots, c_{s+1,n}) \\ c_2 = (c_{2,1}, c_{2,2}, \dots, c_{2,n}) & c_{s+2} = (c_{s+2,1}, c_{s+2,2}, \dots, c_{s+2,n}) \\ \vdots & \vdots \\ c_s = (c_{s,1}, c_{s,2}, \dots, c_{s,n}) & c_{2s} = (c_{2s,1}, c_{2s,2}, \dots, c_{2s,n}) \end{array}$$

Thus we transmit the digits

$$c_{1,1} c_{2,1} \dots c_{s,1} c_{1,2} c_{2,2} \dots \dots c_{s,n} c_{s+1,1} c_{s+2,1} \dots c_{2s,1} c_{s+1,2} c_{s+2,2} \dots \dots c_{2s,n}$$

Since our code is  $l$  burst error correcting (but not necessarily  $l$  cyclic burst error correcting), the worst case would be if a burst of errors just overlaps into the start of a new set of  $s$  codewords. For example, suppose there were a burst of errors affecting the end of the codewords  $c_1, \dots, c_s$  and the first digit of  $c_{s+1}$ . Then we require a period of error free transmission such that the codeword  $c_{s+1}$  is not affected by another burst of errors. That is we require  $s(n - 1)$  bits of error free transmission.