MT361/461/5461 Error Correcting Codes: Preliminary Sheet

Solutions to this sheet will appear on Moodle on 16th January so you can check your answers. You are also welcome to ask me about the questions during an office hour.

1. The purpose of this question is to compare the two communication schemes considered in Example 1.2. We shall assume that Alice wants to send Bob the message 'Yes'. (You may check, if you wish, that the behaviour for 'No' is symmetric.)

Suppose that whenever a binary word is sent down the channel, each of its bits flips with probability $p > 0$, so a 0 becomes a 1 and a 1 becomes a 0.

(a) Why is it reasonable to assume that $p < 1/2$?

(b) Suppose that Alice and Bob use Scheme 1, so Alice sends 11 down the channel. If Bob receives 01 or 10, he requests retransmission.

(i) Explain why the probability that Bob receives 00 is p^2.

(ii) Find the probability that Bob receives 10.

(iii) Find the probability that Bob receives either 01 or 10.

(iv) Let $r \in \mathbf{N}$. Show that the probability that Bob decodes Alice's message as 'No' after r attempts at transmission is $(2p(1-p))^{r-1}p^2$.

(v) Hence show that the probability that Bob decodes Alice's message wrongly is

$$\frac{p^2}{1-2p(1-p)}.$$

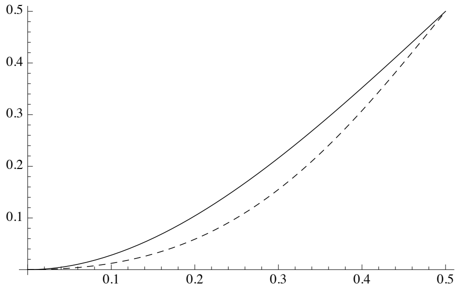
(vi) Let b be the average total number of bits that Alice sends to Bob in Scheme 1. Explain why $b = 2(1-q) + (2+b)q$, where $q = 2p(1-p)$. Hence show that

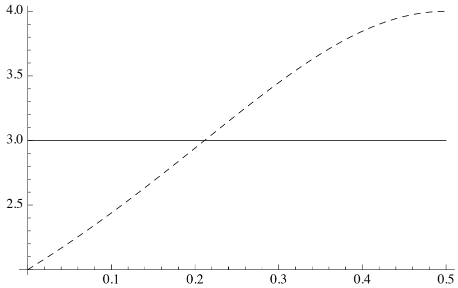
$$b = \frac{2}{1-2p(1-p)}.$$

(c) Using Scheme 2 [**sorry, this was mistyped as 'Scheme 1'**], the probability that Bob decodes Alice's message wrongly was found to be $3p^2 - 2p^3$. How many bits does Alice send to Bob when this scheme is used?

(d) Compare the relative merits of Schemes 1 and 2 when $p = 0.1$ and $p = 0.25$. Why might Scheme 2 be used even when $p = 0.1$?

To confirm your answers in (d) you can use the graphs below which show the probability of incorrect decoding, and the average number of bits sent, for p between 0 and 1/2. The dashed line shows Scheme 1, the solid line shows Scheme 2.





2. Let $\mathbf{Z}_2$ denote the alphabet of binary digits $\{0, 1\}$ with addition modulo 2. So

$$0 + 0 = 0, \quad 0 + 1 = 1, \quad 1 + 0 = 1, \quad 1 + 1 = 0.$$

The *square code* is the binary code of length 8 with codewords

$$\{(u_1, u_2, u_3, u_4, u_1 + u_2, u_3 + u_4, u_1 + u_3, u_2 + u_4) : u_1, u_2, u_3, u_4 \in \mathbf{Z}_2\}.$$

The name comes from the representation of the codewords as a square of four message bits, (u_1, u_2, u_3, u_4), surrounded by four check bits.

$$\begin{array}{cc|c} u_1 & u_2 & u_1 + u_2 \\ u_3 & u_4 & u_3 + u_4 \\ \hline u_1 + u_3 & u_2 + u_4 & \end{array}$$

In general, a received word $(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8)$ may usefully be represented by the square diagram

$$\begin{array}{cc|c} v_1 & v_2 & v_5 \\ v_3 & v_4 & v_6 \\ \hline v_7 & v_8 & \end{array}$$

(a) Check that 11000011 is a codeword in the square code. Draw it as a square diagram.

(b) Suppose that Alice sends 11000011 down a noisy channel, and Bob receives 01000011, which he represents by the square

$$\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 1 & 1 & \end{array}$$

(i) Explain why Bob knows that an error has occurred in the channel.

(ii) Suppose Bob assumes that exactly one error has occurred. Explain how he can work out that Alice sent 11000011.

(c) Assume that the channel behaves as in Question 1. For each of the received words 00100110, 01001100 and 01101110, decide which codeword you think Alice is most likely to have sent.

MT361/461/5461 Error Correcting Codes: Sheet 1

Hand in your answers to Questions 2, 3, 4 and 5.
If you are an MSc or MSci student please also do Question 6. (This question is highly recommended to everyone else.) All other questions are optional for everyone. I will be happy to discuss any of the questions in office hours.

To be returned to McCrea 240 by 1pm on Thursday 26th January 2012 or handed in at the Thursday lecture.

1. Find all words $v \in \{0,1\}^5$ such that $d(v, 11011) = 3$.

2. Let C be the binary code with codewords 001, 010, 100 and 111.

(a) What are the length and size of C?

(b) Draw a diagram showing C as a subset of $\{0,1\}^3$. (For one way to draw $\{0,1\}^3$, see page 6 of the lecture notes.)

(c) Explain why C is 1-error detecting, referring to Definition 2.7 in your answer.

(d) What is the probability that, when a codeword $u \in C$ is sent, a different codeword $u' \in C$ is received? Assume that each of the three bits in a transmitted word flips independently with probability p.

3. Let $m \in \mathbf{N}$ and let C be the repetition code of length $2m$ over a q-ary alphabet A, where $q \geq 2$. Show that C is $(m-1)$-error correcting but not m-error correcting.

4. Let C be a code such that if u, $u' \in C$ and $u \neq u'$ then $d(u, u') \geq 3$. Arguing directly from Definition 2.7, show that

(a) C is 2-error detecting.

(b) C is 1-error correcting.

Show that the condition on distances is satisfied by the binary code of length 5 with codewords

00000, 11100, 00111, 11011.

Is this code 3-error detecting? Is it 2-error correcting?

5. (a) Which of the following are ISBNs:

1-84628-040-0, 1-84628-400-0, 0-486-68735-X?

(b) Show that if two unequal adjacent digits are interchanged when writing down an ISBN then the result is not an ISBN.

6. Suppose that C is a binary code of length 5 such that $d(u, u') \geq 3$ for all distinct codewords u, $u' \in C$. Show that C has size at most 4.

7. I have an important decision 'Yes' or 'No' that I wish to communicate to a friend across a crowded room. I can shout to him up to three times. The probability that he mishears on the first shout is p, and on the second and third it is $r > p$. Assume that $r < 1/2$.

(a) Explain a three-shout strategy, making it clear how my friend will decode what he hears.

(b) Calculate the probability that the three-shout strategy will successfully communicate the message.

(c) If $p = 1/5$ show that the three-shout strategy is superior to a single shout if and only if $r < 1/3$.

(d) Show more generally that there is a function $f : [0, 1/2] \to [0, 1/2]$ such that three shouts are superior to a single shout if and only if $r < f(p)$. Sketch the graph of f.

8. Consider the binary code $C = \{00, 01\}$ of length 2. Show that, if we mistakenly write $d(u, v) = t$, rather than $d(u, v) \leq t$, in Definition 2.7, then C is 2-error detecting but not 1-error detecting.

9. In Lewis Carroll's 'Doublets Game', the aim is to turn one word into another, changing one letter at a time, while staying within the English language.

(a) Show that $d(\text{WARM}, \text{COLD}) = 4$ and find a solution to the Doublet puzzle starting at WARM and ending at COLD using just 3 intermediate words.

(b) Find $r \in \mathbf{N}$ and English words u and w such that $d(u, w) = r$ but there is no r-step solution to the Doublet puzzle.

(c) Suggest an efficient algorithm for solving Doublet puzzles.

10. Do there exist binary words u, v, w of the same length such that $d(u, v) = 3$, $d(v, w) = 4$ and $d(w, u) = 6$? **Misprinted as $d(w, u) = 5$, which makes the question less interesting.** Now answer the same question for words over the ternary alphabet $\{0, 1, 2\}$.

MT361/461/5461 Error Correcting Codes: Sheet 2

Hand in your answers to Questions 1, 2, 3 and 4.
If you are an MSc or MSci student please also do Question 10. All other questions are optional for everyone. I will be happy to discuss any of the questions in office hours.

To be returned to McCrea 240 by 1pm on Thursday 2nd February 2012 or handed in at the Thursday lecture.

1. Let C be the binary code with codewords

$$00000,\ 11100,\ 00111,\ 11011.$$

Decode (a) 11111, (b) 10010 and (c) 11000 using nearest neighbour decoding.

2. Let C be the binary code of length 9 whose codewords are all binary words of the form

$$(u_1, u_2, u_3, u_4, u_1, u_2, u_3, u_4, x)$$

where x is chosen to make the total number of 1s in (u_1, u_2, u_3, u_4, x) even.

(a) Find codewords $u, w, u', w' \in C$ such that $d(u, w) = 3$ and $d(u', w') = 4$.

(b) Let

$$\begin{aligned} u &= (u_1, u_2, u_3, u_4, u_1, u_2, u_3, u_4, x) \\ v &= (v_1, v_2, v_3, v_4, v_1, v_2, v_3, v_4, y) \end{aligned}$$

be codewords in C.

(i) Show that if $d(u_1u_2u_3u_4, v_1v_2v_3v_4) = 1$ then $d(u, v) = 3$.

(ii) Show that if $d(u_1u_2u_3u_4, v_1v_2v_3v_4) \geq 2$ then $d(u, v) \geq 4$.

(iii) Hence find the minimum distance of C.

(c) Deduce that C is 1-error correcting and 2-error detecting. [*You may use Theorem 3.4.*]

3. Let C be a t-error correcting code over the alphabet A. Suppose that you receive the word $v \in A^n$ and that $w \in C$ is the nearest codeword to v in the Hamming distance. Arguing from Definition 2.7, show that if w is not the sent codeword, then at least $t + 1$ errors have occurred in the channel.

4. Using the Square Code (see preliminary problem sheet) find an eight question strategy for the Liar Game (see the second exercise on page 7), in which you can write down, in advance, the eight questions you will ask during the game.

[*Hint: in Lecture 4 the Square Code was used to send numbers between* 0 *and* 15 *down a noisy channel. The slides are available on Moodle.*]

Please state your questions so that they will be intelligible to non-mathematicians.

5. In the ternary repetition code of length 7 (considered in Example 2.5), what is the maximum distance of a word $v \in \{0, 1, 2\}^7$ from a codeword?

6. As in Example 2.12, say that a word of length 10 over the alphabet $\{0, 1, \ldots, 9, \mathrm{X}\}$ is an *ISBN* if it satisfies

$$10u_1 + 9u_2 + \cdots + 2u_9 + u_{10} \equiv 0 \mod 11,$$

where any X in the word is interpreted at 10.

(a) Show, by example, that the code consisting of all ISBNs is not 2-error detecting or 1-error correcting. [*Hint: There is no need to use ISBNs that are assigned to books, so one possible starting codeword is the all-zeroes word* $u = 00\ldots0$.]

(b) What is the minimum distance of the ISBN code?

(c) Show that if an ISBN is written down backwards then it is still an ISBN.

7. The purpose of this question is to give a more algebraic proof of the triangle inequality for Hamming distance (proved in lectures in Theorem 2.3). Let A be an alphabet and let u, v, w be words over A of length n.

(a) Let $i \in \{1, 2, \ldots, n\}$. Thinking of u_i, v_i, w_i as words of length 1, prove that

$$d(u_i, w_i) \le d(u_i, v_i) + d(v_i, w_i).$$

(b) Deduce the triangle inequality by summing the inequality in (a) over all i.

8. Let C be the length 12 binary code in Lemma 2.11. Show that no matter how many errors occur in the channel, there is always a unique nearest codeword to any received word. (This shows that nearest neighbour decoding never fails for C: of course it might give the wrong answer.)

9. Let $p < 1/2$. A jury consists of three people. Two of them get the verdict wrong with probability p, while the third flips a coin to decide. What is the probability that the majority verdict of the jury is correct? Comment on your answer.

10. **(MSc/MSci)** Let C be the Reed–Solomon code with alphabet $\mathbf{F}_5$ defined in Example 2.2(2) of the MSc lecture notes. So $n = 4$ and the codewords are

$$u(f) = (f(0), f(1), f(2), f(3))$$

for $f(x) = bx + c$, with b, $c \in \mathbf{F}_5$.

Suppose that the word $v = 1312$ is received. Show that $v = 1312$ is not in C and decode v using nearest neighbour decoding.

MT361/461/5461 Error Correcting Codes: Sheet 3

Hand in your answers to Questions 2, 3, 4, 5.
If you are an MSc or MSci student please also do Question 10, and only do Question 4 if you want to try it again. All other questions are optional for everyone. I will be happy to discuss any of the questions in office hours.

To be returned to McCrea 240 by 1pm on Thursday 9th February 2012 or handed in at the Thursday lecture.

1. Let C be the code over the ternary alphabet $\{0, 1, 2\}$ with codewords

$$0000 \quad 0111 \quad 0222 \quad 1012 \quad 1120 \quad 1201 \quad 2021 \quad 2102 \quad 2210$$

Decode the received words 1201, 2121, 2222 using nearest neighbour decoding. In each case, write down how many errors must have occurred in the channel if your answer is not the sent word.

2. Let C a binary code. For each codeword $u = u_1u_2 \ldots u_n \in C$, let $D(u)$ be the word of length $2n$ defined by

$$D(u) = u_1u_2 \ldots u_n \, u_1u_2 \ldots u_n.$$

Let $D(C) = \{D(u) : u \in C\}$. Prove that if C is an (n, M, d)-code then $D(C)$ is a $(2n, M, 2d)$-code.

[*Hint: please do not assume that C is a repetition code, or any other special type of code. The only thing you may assume about C is that it is a binary code with parameters (n, M, d).*]

3. Let u and w be binary words of length n.

(a) Let $1 \le i \le n$ and let u' and w' be the binary words obtained by flipping the bit in position i of u and w, respectively. Show that $d(u, w) = d(u', w')$.

(b) Let $1 \le i < j \le n$ and let u' and w' be the binary words obtained by swapping the bits in positions i and j of u and w, respectively. Show that $d(u, w) = d(u', w')$.

4. The aim of this question is to show that the maximum size of a binary code of length 5 and minimum distance 3 is 4.

(a) Suppose that C is a binary $(5, M, 3)$-code where $M \ge 4$. Using Question 3, explain why we can assume that $00000 \in C$.

(b) Show that C contains at most one word with four or five 1s.

(c) Explain why we can assume that $11100 \in C$.

(d) Find all binary words w with exactly three 1s such that $d(11100, w) \ge 3$.

(e) Hence show that C has size 4.

5. Let C be a code and let $t \in \mathbf{N}$. Prove that the minimum distance of C is at least $2t+1 \iff$ the Hamming balls of radius t about distinct codewords in C are disjoint.

[*Note: The result is already known from Theorem 4.6* **(corrected from 4.1 after lecture on 2nd February)***, but the point of this question is to give a direct proof of (b)* $\iff$ *(d) that does not go* via *condition (a) that C is t-error correcting.*]

6. Show that the maximum size of a ternary 1-error correcting code of length 4 is 9.

7. Suppose that Alice wishes to send a message 'Yes' or 'No' to Bo through the binary symmetric channel with crossover probability p. She sends 'Yes' with probability $3/4$ and 'No' with probability $1/4$. They agree to use the length 3 binary repetition code, encoding 'Yes' as 111 and 'No' as 000.

(a) Find $\mathbf{P}[001 \text{ received} \mid 000 \text{ sent}]$ and $\mathbf{P}[001 \text{ received} \mid 111 \text{ sent}]$.

(b) Hence find $\mathbf{P}[000 \text{ sent} \mid 001 \text{ received}]$.

(c) Find $\mathbf{P}[111 \text{ sent} \mid 001 \text{ received}]$.

(d) Show that if $p > 1/4$ then

$$\mathbf{P}[111 \text{ sent} \mid 001 \text{ received}] > \mathbf{P}[000 \text{ sent} \mid 001 \text{ received}].$$

(e) Comment on the implications for nearest neighbour decoding.

8. Alice knows a polynomial f with coefficients in the natural numbers, of unknown degree. Bob can pick any number $x \in \mathbf{Z}$ and ask Alice to tell him $f(x)$. After hearing Alice's answer, Bob may then pick $y \in \mathbf{Z}$ and ask for $f(y)$, and so on. Find a strategy for Bob that will determine f in as few questions as possible.

9. **(MSc/MSci)** Let p be a prime, let $p \geq n \geq k$ and let $a_1, a_2, \ldots, a_n \in \mathbf{F}_p$.

(a) Using Lemma 1.6(iii) prove that if f, $g \in \mathbf{F}_p[x]$ are distinct polynomials of degree $< k$ then $u(f) \neq u(g)$.

(b) Hence give an alternative proof of Lemma 2.3, that the Reed–Solomon code $RS_{p,n,k}$ has size p^k.

10. **(MSc/MSci)** Let C be the Reed–Solomon code with parameters $p = 7$, $n = 5$, $k = 3$ where polynomials are evaluated at $0, 1, 2, 3, 4 \in \mathbf{F}_7$.

(a) Using polynomial interpolation, or otherwise, find a codeword $u \in C$ such that the first three positions of u are $(1, 0, 4)$.

(b) Suppose that the word $v = (1, 4, 1, 1, 2)$ is received. Explain why there is at most one codeword within distance 1 of v. Find such a codeword.

MT361/461/5461 Error Correcting Codes: Sheet 4

Hand in your answers to Questions 1, 2 and 3(a), (b).
If you are an MSc or MSci student please also do Questions 3(c) and 4. All other questions are optional. I will be happy to discuss any of the questions in office hours.

To be returned to McCrea 240 by 1pm on Thursday 18th February 2012 or handed in at the Thursday lecture.

1. Consider the five binary codes below:

$$\begin{aligned} C_1 &= \{0000, 1100, 1010, 0110\} \\ C_2 &= \{0111, 0100, 0010, 0001\} \\ C_3 &= \{1000, 0100, 0010, 0001\} \\ C_4 &= \{0000, 1100, 0110, 0011\} \\ C_5 &= \{0110, 1100, 1001, 0011\} \end{aligned}$$

(a) Show that C_1 is equivalent to C_2.

(b) Show that C_1 is not equivalent to C_3. Is C_2 equivalent to C_3?

(c) Show that C_4 and C_5 are not equivalent. Is either equivalent to C_1, C_2 or C_3?

(d) (Optional.) Classify all binary codes of length 4 and minimum distance 2 up to equivalence.

2. (a) Use Lemma 6.6 to show that $A_2(n, d) = 2$ whenever $d > 2n/3$.

(b) Show that if C is a binary code of length 9 and minimum distance 6 then $|C| \leq 4$. Hence show that $A_2(9, 6) = 4$.

3. A binary code C of length n is said to be *perfect* if there exists $e \in \mathbf{N}$ such that

$$\{0, 1\}^n = \bigcup_{u \in C} B_e(u)$$

where the union is disjoint. (In words: the Hamming balls of radius e about codewords are disjoint, and every binary word of length n is in one of these balls.)

(a) Show that if n is odd then the binary repetition code of length n is perfect.

(b) Show that if C is a perfect binary code of length n with $e = 1$ then C is 1-error correcting and $n = 2^m - 1$ for some $m \in \mathbf{N}$. Express $|C|$ in terms of m. [*You may use any general results proved earlier in the course.*]

(c) (**MSc, MSci**) Show that a perfect binary code has odd minimum distance.

4. (**MSc, MSci**) Consider the Reed–Solomon code $RS_{5,4,2}$ over $\mathbf{F}_5$ where polynomials are evaluated at $a_1 = 0$, $a_2 = 1$, $a_3 = 2$, $a_4 = 3$. Suppose that you receive the words (i) 2413, (ii) 1033, (iii) 1032. In each case solve the Key Equation for $Q(x)$ and $E(x)$ and decode (where possible) the received word.

5. Let $n, d \in \mathbf{N}$ where $n \geq d$. Let C be a code of length $n+1$ and minimum distance $d+1$.

(a) Suppose that two codewords at distance $d+1$ in C differ in position i. Show that the code $C^\star$ of length n obtained by removing position i from all codewords in C has minimum distance d.

(b) Deduce that $A_q(n+1, d+1) \leq A_q(n, d)$ for any $q \geq 2$.

6. Let $n, d \in \mathbf{N}$ where $n \geq d$. Suppose that C is a binary code of length n and minimum distance d. Define a new code C^+ of length $n+1$ by appending a final bit to each codeword in C so that each codeword in C^+ has an even number of 1s.

(a) Show that the distance between any two codewords in C^+ is even.

(b) Use the previous question to show that if d is odd then

$$A_2(n, d) = A_2(n+1, d+1).$$

7. Let $q \geq 2$ and let A be a q-ary alphabet.

(a) Show that if $u \in A^n$ then the number of words in the Hamming ball of radius t about u is

$$\sum_{k=0}^{t} \binom{n}{k} (q-1)^k.$$

(b) Hence generalize Theorem 5.4 by showing that if C is a q-ary (n, M, d)-code then

$$M \leq q^n \Big/ \sum_{k=0}^{e} \binom{n}{k} (q-1)^k$$

where $e = \lfloor (n-1)/2 \rfloor$.

8. Suppose that we use a binary code C of length n to send messages down the binary symmetric channel C described on page 19 of the notes.

(a) Suppose that we receive a word $v \in \{0,1\}^n$. Show that for each $u \in C$,

$$\mathbf{P}[u \text{ sent} \mid v \text{ received}] = \frac{\mathbf{P}[v \text{ received} \mid u \text{ sent}]\mathbf{P}[u \text{ sent}]}{\mathbf{P}[v \text{ received}]}.$$

(b) Show that $\mathbf{P}[v \text{ received} \mid u \text{ sent}] = p^{d(u,v)}(1-p)^{n-d(u,v)}$.

(c) Hence prove Theorem 4.5.

MT361/461/5461 Error Correcting Codes: Sheet 5

Hand in your answers to Questions 1, 2 and 3.
If you are an MSc or MSci student then Question 3 is optional, but please do Questions 1, 2, 5 and 6. All other questions are optional for everyone. Question 4 is in the style of previous exam questions. I will be happy to discuss any of the questions in office hours.

1. Suppose that C is a $(4, q^2, 3)$-code over the alphabet $A = \{0, 1, \ldots, q-1\}$.

(a) Show that if $u = (u_1, u_2, u_3, u_4)$ and $u' = (u'_1, u'_2, u'_3, u'_4) \in C$ are distinct codewords then $(u_1, u_2) \neq (u'_1, u'_2)$.

(b) Deduce that for all i, $j \in A$ there is a unique codeword, say (i, j, X_{ij}, Y_{ij}), whose first two positions are (i, j). [*Hint: C has size q^2.*]

(c) Explain in one sentence why it follows that

$$C = \{(i, j, X_{ij}, Y_{ij}) : i, j \in A\}$$

(i) Prove that the rows of the matrix X have distinct entries. [*Hint: suppose row i has a repeated entry, so $X_{ij} = X_{ij'}$ where $j \neq j'$. What does this imply about the codewords whose first two positions are (i, j) and (i, j')?*]

(ii) Prove that the columns of X have distinct entries.

(iii) Deduce that X is a Latin square.

(iv) Prove that X and Y are MOLs.

2. Let C be the ternary code with codewords $000, 111, 222, 012, 021, 120, 102, 201, 210$. Let $C^\star$ be the code obtained from C by puncturing it in its final position. Write down the codewords in $C^\star$ and find the length, size, and minimum distance of $C^\star$.

3. Let $q \geq 2$ and let $A = \{0, 1, \ldots, q-1\}$. Suppose that C is a q-ary code of length $n \geq 2$ and minimum distance $n - 1$ **(corrected from 'minimum distance 3' after lecture on Thursday 16th)**.

(a) Show that if $u = u_1u_2 \ldots u_n$ and $v = v_1v_2 \ldots v_n$ are codewords in C then

$$d(u_1u_2, v_1v_2) \geq 1.$$

(b) By putting codewords into pigeonholes according to their first two symbols, show that $|C| \leq q^2$.

(c) Deduce that $A_q(n, n-1) \leq q^2$. (This is a special case of the Singleton bound. The case $n = 4$ shows that codes constructed from mutually orthogonal Latin squares are as large as possible.)

4. What does it mean to say that a binary code is an (n, M, d)-code?

For which of the following parameters either give a binary code with these parameters, or show that no such code can exist:

(i) $(5, 2, 5)$; (ii) $(5, 3, 4)$; (iii) $(5, 4, 3)$; (iv) $(5, 16, 2)$.

5. Read the introduction, Section 6, and at least one other section from Hamming's original paper, *Error Detecting and Error Correcting Codes*, Bell Systems Technical Journal, **2** (1950) 147–160. (See `http://www.lee.eng.uerj.br/~gil/redesII/hamming.pdf`.)

Think critically about how Hamming writes and identify at least two strengths or weaknesses of his paper.

6. By generalizing the steps in Question 3, give an alternative proof of the Singleton bound $A_q(n, d) \leq q^{n-d+1}$.

7. (For people who know some basic group theory.) Let G be a finite group of order n with group operation $\circ$. Suppose that $G = \{g_1, g_2, \ldots, g_n\}$. Show that the matrix X defined by $X_{ij} = g_i \circ g_j$ is a Latin square over the alphabet G.

8. The Grand-Vizier and his fifty servants are planning a banquet. Owing to an administrative error, one of the 1000 barrels of wine in his wine-cellar been poisoned with a deadly but slow-acting poison: anyone who drinks from the poisoned barrel will die at a random time in the next day.

(a) Devise a tasting strategy that will identify the poisoned barrel within one day.

(b) One of the Vizier's servants has been replaced with an assassin who is immune to all poisons. Propose a new tasting strategy.

MT361/461/5461 Error Correcting Codes: Sheet 6

Hand in your answers to Questions 1 and 2.

Question 3 is in the style of previous exam questions. I will be happy to discuss any of the questions in office hours.

To be returned to McCrea 240 by 1pm on Thursday 1st March 2012 or handed in at the Thursday lecture.

1. Use the contruction in Lemma 7.6 to write down two mutually orthogonal Latin squares of order 3. Hence construct a $(4, 9, 3)$-code over the alphabet $\{0, 1, 2\}$ containing the codeword 0000.

2. (a) Let H be a Hadamard matrix of order n. Show that the $2n \times 2n$ matrix

$$K = \begin{pmatrix} H & H \\ H & -H \end{pmatrix}$$

is a Hadamard matrix of order $2n$.

(b) Starting from the 2×2 Hadamard matrix

$$\begin{pmatrix} + & + \\ + & - \end{pmatrix}$$

use (a) to construct Hadamard matrices with orders 4 and 8. Hence write down the codewords in (i) a binary $(4, 8, 2)$ code and (ii) a binary $(8, 16, 4)$-code.

(c) Use nearest neighbour decoding to decode (where possible) the received words 01010111, 10011110 and 11000000 sent using the code in (b)(ii).

(d) Use (b)(ii) to show that there is a binary $(7, 16, 3)$-code. Deduce from the Hamming Packing Bound that $A_2(7, 3) = 16$.

(e) Using (b)(ii) and the Plotkin bound, prove that $A_2(7, 4) = 8$.

3. (a) Let u be a binary word of length n and let $t \in \mathbf{N}$. Define the *Hamming ball $B_t(u)$ of radius t about u.*

(b) Let C be a binary code of length n. What does it mean to say that C is *t-error correcting*?

(i) Let u, $w \in C$ be distinct codewords. Show that if C is t-error correcting then the Hamming balls of radius t about u and w are disjoint. [*You may assume, if you wish, that the triangle inequality holds for Hamming distance.*]

(ii) Prove that $|B_t(u)| = \sum_{k=0}^{t} \binom{n}{k}$ for $u \in C$ and $t \in \mathbf{N}$.

(iii) Deduce that if C is t-error correcting then

$$|C| \leq \frac{2^n}{\sum_{k=0}^{t} \binom{n}{k}}.$$

4. Show that a Hadamard matrix of order ≥ 4 has order divisible by 4.

[*Hint: if r, r' and r'' are three rows of H then, by reordering columns, we may assume that r and r' are equal in their first $n/2$ positions and differ in their final $n/2$ positions. What restrictions does this put on r''?*]

5. The purpose of this question is to give a geometric proof of the Plotkin bound (Theorem 9.6). Let C be a binary (n, M, d)-code where $2d > n$. For each codeword $u \in C$ define an associated vector $x(u) \in \mathbf{R}^n$ by

$$x(u)_i = \begin{cases} 1 & \text{if } u_i = 0, \\ -1 & \text{if } u_i = 1. \end{cases}$$

Let $x \cdot y$ be the usual dot product of vectors $x, y \in \mathbf{R}^n$.

(a) Show that $x(u) \cdot x(u) = n$ for all $u \in C$.

(b) Let $u, u' \in C$ be distinct codewords. Show that $x(u) \cdot x(u') = n - 2d(u, u')$ and deduce that $x(u) \cdot x(u') \leq -(2d - n)$.

(c) Let $z = \sum_{u \in C} x(u)$. By considering $z \cdot z$ prove the Plotkin bound.

(d) Find z if C is a binary $(5, 4, 3)$-code.

6. Let C be a binary (n, M, d)-code where $n \geq 2d$. By putting the codewords in C into pigeonholes according to their final $n - 2d$ positions, and then applying Corollary 9.7 to each code of length $2d$ obtained by repeatedly puncturing the codewords in each pigeonhole, prove the asymptotic Plotkin bound

$$|C| \leq 2^{n-2d+1} n.$$

Compare this bound with the Singleton bound for binary codes.

7. Let p be a prime such that $p \equiv 3 \bmod 4$ and let $\mathbf{F}_p$ be the finite field with p elements. We say that $x \in \mathbf{F}_p$ is a *square* if there exists $y \in \mathbf{F}_p$ such that $x = y^2$.

Let Q be the $p \times p$ matrix Q whose rows and columns are indexed by $i, j \in \{0, 1, \ldots, p-1\}$ and where

$$Q_{ij} = \begin{cases} -1 & \text{if } i - j \text{ is a square in } \mathbf{F}_p \\ 1 & \text{otherwise.} \end{cases}$$

It is known that the matrix H obtained from Q by adding an extra row and an extra column consisting entirely of 1s is a Hadamard matrix of order $p + 1$. Use this to find a Hadamard matrix of order 12.

MT361/461/5461 Error Correcting Codes: Sheet 7

Hand in your answers to Questions 1 and 2.

If you are an MSc or MSci student please also do Question 3. Question 4 is in the style of past exam questions. All other questions are optional. I will be happy to discuss any of the questions in office hours.

To be returned to McCrea 240 by 4pm on Tuesday 13th March 2012 or handed in at the Tuesday lecture.

1. Let C be the square code consisting of all codewords of the form

$$\{(u_1, u_2, u_3, u_4, u_1 + u_2, u_3 + u_4, u_1 + u_3, u_2 + u_4)\}.$$

(a) Show that if $u, w \in C$ then $u + w \in C$, and so C is linear.

(b) Find, with proof, a basis for C. (You should show that the codewords you write down are linearly independent and span C.)

(c) Write down a generator matrix for C. Use your generator matrix to encode a number of your choice.

(d) Let C_{ext} be the parity check extension of C. Write down a general form for the codewords in C_{ext}. Prove that the minimum distance of C_{ext} is 4. What is the maximum number of errors that C_{ext} can detect and correct? [*You may use Lemma 11.4 and any other general results proved in the course.*]

2. Let $n \in \mathbf{N}$. Given binary words $u = u_1u_2 \ldots u_n$ and $v = v_1v_2 \ldots v_n$ of length n, we write $(u \mid v)$ for the word $u_1u_2 \ldots u_nv_1v_2 \ldots v_n$ of length $2n$. Let $\mathbf{1}$ stand for the all-ones word of length n, i.e. $\mathbf{1} = 11 \ldots 1$.

If C is a linear binary code of length n then we let $E(C)$ be the code of length $2n$ consisting of all words of the form $(u \mid u)$ and $(u \mid u + \mathbf{1})$.

(a) Write down the codewords in $E(\{00, 01, 10, 11\})$. Find the length, size and minimum distance of $E(\{00, 01, 10, 11\})$

Let C be a linear binary code

(b) Show that $E(C)$ is linear.

(c) Suppose that $u(1), \ldots, u(m)$ is a basis for C. Show that

$$(u(1)|u(1)), \ldots (u(m)|u(m)), (\mathbf{0}|\mathbf{1})$$

is a basis for $E(C)$.

(d) Show that $|E(C)| = 2|C|$.

(e) Suppose that C has minimum distance $d \leq n/2$. Show that $E(C)$ has minimum distance $2d$. [*You may use Lemma 11.4, or argue directly, as you prefer.*]

3. (**MSc/MSci**)

(a) Let $\bar{f} = 1 + x + x^3$, $\bar{g} = 1 + x^2 + x^3$ be elements of $\mathbf{F}_2[x]/(x^7 - 1)$. Working in this ring, calculate the following products

$$\text{(i) } \bar{f}(x)\bar{g}(x) \quad \text{(ii) } x^5\bar{f}(x); \quad \text{(iii) } \bar{f}(x)\bar{g}(x)(1+x); \quad \text{(iv) } \bar{f}(x)^3.$$

(b) Let C be the ternary code of length 6 defined by

$$C = \{(u_0, u_1, u_2, u_3, u_4, u_5) \in \mathbf{F}_3^6 : u_0 + u_2 + u_4 = 0, u_1 + u_3 + u_5 = 0\}.$$

Show that C is cyclic and find a generating polynomial for C. What are the parameters of C?

4. (a) Define the *weight* of a binary codeword.

(b) Prove that the minimum distance of a linear binary code C is equal to the minimum weight of its non-zero codewords.

(c) Let C be the binary code of length 12 whose codewords are all words of the form $u_1u_2u_3u_4\ u_1u_2u_3u_4\ u_1u_2u_3u_4$.

(i) Find the minimum distance of C.

(ii) Define the parity check extension C_{ext} of C and determine its parameters (length, size, minimum distance).

5. Let C be a linear binary code of length n and let $1 \leq i \leq n$. Show that either all codewords in C have 0 in their ith position, or half of the codewords have 0 in their ith position and half have 1. [*Hint: if u is a codeword with $u_i = 1$, consider the map $C \to C$ defined by $v \mapsto v + u$.*]

6. Suppose that C is a binary code of length n and minimum distance at least δn where $0 < \delta < 1$. Show that if $\delta = 1/2$ then $|C|$ can be arbitrarily large, but if $\delta > 1/2$ then $|C|$ is bounded as n tends to infinity.

7. The *rate* of a binary code C of size M and length n is defined to be $(\log_2 M)/n$.

(a) By generalizing the square code S, define a linear binary $[n^2 + 2n, n^2, 3]$-code for each $n \in \mathbf{N}$. Show that the rate of these codes tends to 1 as $n \to \infty$.

(b) Define a *cube code* by analogy with the square code. By extending these codes, find a family of two-error correcting linear binary codes whose rate tends to 1 as the length tends to infinity.

MT361/461/5461 Error Correcting Codes: Sheet 8

Hand in your answers to Questions 1 and 2.

If you are an MSc student please also do Question 5. All other questions are optional for everyone. I will be happy to discuss any of the questions in office hours. Question 7 is based on the final part of Question 5 on the 2010 exam paper.

To be returned to McCrea 240 by 4pm on Tuesday 19th March 2011 or handed in at the Monday lecture.

1. Let C be the binary code of length 8 with generator matrix

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

(a) Put G into reduced row-echelon form. (Remember that swapping rows is an admissible row operation.)

(b) Find a code C' and a generator matrix G' for C' such that C' is equivalent to C and G' is in standard form.

(c) Using the generator matrix G', encode 7 as a codeword in C'. Which codeword in C' encodes the binary number $b_1b_2b_3b_4$?

(d) Show that C' is not equivalent to the square code by a shuffle of positions. Is C equivalent to the square code by a shuffle of positions?

2. Let $C = \{0000, 1111\}$ be the repetition code of length 4.

(a) Construct a standard array for C.

(b) Use your standard array to decode the received words 0000, 0010, 1010, 0101.

(c) Suppose that 0000 is sent through a noisy channel in which each bit flips independently with probability $p < 1/2$.

(i) Explain why the probability that 0000 is received is $(1-p)^4$.

(ii) Show that the probability that a word of weight 1 is received is $4p(1-p)^3$.

(iii) Which words of weight 2 will be decoded to 0000 using your standard array? Find the probability that one of these words is received.

(iv) Hence find the probability that the receiver decodes the received word as 0000. Evaluate this probability when $p = 1/5$.

3. Show that if C is a linear binary code of length n then the code $C^\star$ obtained by puncturing C in its final position is also linear.

4. **(MSc, MSci)** Find generator and parity check matrices for the Reed–Solomon code $RS_{5,4,2}$ in which polynomials are evaluated at $0, 1, 2, 3 \in \mathbf{F}_5$.

5. **(MSc, MSci)** Let C be the cyclic code of length 4 over $\mathbf{F}_5$ with generator polynomial $g(x) = (x+3)(x+1)$.

(a) Write down a generator matrix for C.

(b) Show that $u \in \mathbf{F}_5^4$ is a codeword in C if and only if $uH^{tr} = \mathbf{0}$ where

$$H = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 1 & 4 & 1 & 4 \end{pmatrix}.$$

(c) Find the minimum distance of C. [*Hint: Lemma 11.4 holds for linear codes over any finite field.*]

(d) Show that C is equal to the Reed–Solomon code $RS_{5,4,2}$ where polynomials are evaluated at $1, 2, 4, 3 \in \mathbf{F}_5$.

6. Let C be a linear binary code of length n. For $i \in \{1, 2, \ldots, n\}$ let $e(i)$ be the word with 1 in position i and 0 in all other positions.

(a) Show that C is one-error correcting if and only if the cosets

$$C,\ C + e(1),\ \ldots,\ C + e(n)$$

are pairwise disjoint.

(b) Suppose that H is a parity check matrix for C. Show that if v, $v' \in \mathbf{Z}_2^n$ then $C + v = C + v'$ if and only if $vH^{tr} = v'H^{tr}$. Hence show that C is one-error correcting if and only if the columns of H are distinct and non-zero.

7. Consider the Hadamard code $C = \{0000, 1111, 1010, 0101, 1100, 0011, 1001, 0110\}$ of length 4.

(i) Prove that C is a linear code and give a basis for it.

(ii) Write down a generator matrix G and find the corresponding parity check matrix H for C.

(iii) Construct a standard array for C.

8. Let C be a linear binary code of length n and dimension m. The purpose of this question is to use linear algebra and dual spaces to show that the dual code $C^\perp$ has dimension $n - m$.

Let $C^\star$ denote the dual space to C consisting of all linear maps from $C^\star$ to $\mathbf{Z}_2$. Let $f : \mathbf{Z}_2^n \to C^\star$ be defined so that $f(v) \in C^\star$ is the map $u \mapsto \langle u, v \rangle$, for $u \in C$.

(a) Show that f is linear.

(b) Show that the image of f is $C^\star$.

(c) Show that the kernel of f is $C^\perp$.

(d) Deduce from the rank-nullity theorem that $\dim C^\perp = n - m$.

MT361/461/5461 Error Correcting Codes: Sheet 9

Questions 1 and 2 cover the material in the final week of lectures. If you are an MSc/MSci student then please also do Question 10. An extra page of questions on the MSc/MSci course will be issued in the lecture later today.

The lecturer will be happy to answer questions over the vacation sent by email to `mark.wildon@rhul.ac.uk`.

1. Let C be the linear binary code of length 6 with generator matrix

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

(a) Find a code C' equivalent to C by a shuffle of positions such that C' has a generator matrix in standard form.

(b) Use Theorem 14.3 to find a parity check matrix H' for C'.

(c) Hence find a parity check matrix for C. [*Hint: undo the shuffle you performed to turn C into C'.*]

(d) Using the parity check matrix found in (c), make a syndrome / coset-leader table for C.

(e) Check that the syndromes of the six errors affecting just one position are distinct. Explain why this shows that C is 1-error correcting.

(f) Decode the received words 011100, 110111, 000011 using syndrome decoding.

2. Let C be the Hamming $[7, 4, 3]$-code defined in Example 14.6.

(a) Explain how C can be used to encode numbers between 0 and 15. Illustrate your answer by encoding the number 9.

(b) Use syndrome decoding to decode the received words 1011010, 0011011, 1100111 as numbers between 0 and 15.

(c) Show that the Hamming balls $B_1(u)$ of radius 1 about codewords are disjoint. Deduce that $\mathbf{Z}_2^7 = \bigcup_{u \in C} B_1(u)$. (This shows that C is perfect, in the sense defined in Question 3 of Sheet 4.)

(d) Suppose that C is used as a 1-error correcting code to send messages through a binary channel in which each bit of a sent word flips independently with probability $p < 1/2$.

(i) Show that the probability that a message is decoded correctly is

$$(1 - p)^7 + 7p(1 - p)^6.$$

(ii) Evaluate this probability when $p = 1/20$. Compare this probability with the probability of successful decoding if messages are sent directly as binary words of length 4.

3. Use the Singleton bound to show that if C is a linear binary $[n,m,d]$-code then $m \leq n-d+1$.

4. (a) Show that there is a linear ternary one-error correcting code of length 12 with a 3×12 ternary parity check matrix all of whose row sums are 0.

(b) You have 12 pennies, one of which *might* be counterfeit, and of a different weight to the others. Using three weighings on a balance find out whether there is a counterfeit penny, and if so, determine whether it is heavy or light.

5. Show that if there is a linear binary $[n,m,d]$-code then there is a linear binary $[n-s, m-s, d]$-code for each $s \leq m-1$. [*Hint: Question 5 on Sheet 7 is relevant.*]

6. The purpose of this question is to generalize the construction in Example 14.6. Let $r \in \mathbf{N}$ and let H be the $(2^r - 1) \times r$ matrix whose columns are all non-zero binary words of length r. Let

$$C = \{u \in \mathbf{Z}_2^{2r-1} : uH^{tr} = 0\}.$$

(a) Show that C is a linear binary code with parity check matrix H.

(b) Show that C is 1-error correcting, and that C contains a codeword u of weight 3. [**Correction:** $u_1 = u_2 = 1$ **can only be assumed if the columns of** H **are ordered in a suitable way.**] Deduce that C is a $[2^r-1, 2^r-1-r, 3]$-code.

(c) Show that C is perfect.

7. Show that if $M \leq 2^n/(1+n)$ then there is a 1-error correcting linear binary code of length n and dimension m, where m is the largest number such that $2^m \leq M$.

8. Let C be a linear binary code of length n and dimension m with generator matrix G. The purpose of this question is to show that the dual code $C^\perp$ is well-defined, and that $C^{\perp\perp} = C$, as suggested by Example 14.5.

(a) Show that if H is a parity check matrix for C then $GH^{tr} = 0$.

(b) Deduce that the dual code $C^\perp$, defined with respect to H, is contained in $\ker G^{tr}$.

(c) Hence show that $C^\perp = \ker G^{tr}$. [*Hint: show both sides have dimension* $n-m$.]

(d) Show that G is a parity check matrix for $C^\perp$ and so $C^{\perp\perp} = C$.

9. Let C be a linear binary code of length n. Show that if a word $u \in C$ is sent, and $v \in \mathbf{Z}_2^n$ is received, then v is correctly decoded under standard array decoding if and only if $u+v$ is the chosen coset leader in the coset $C+v$.

10. Let C be a linear binary $[n,m,d]$-code with parity check matrix H. Show that any $d-1$ columns of the parity check matrix H are linearly independent, and that there exist $i_1, \ldots i_d$ such that the sum of columns $i_1, \ldots, i_d$ of H is zero.

11. (**MSc, MSci**) Let p be a prime and let $a \in \mathbf{F}_p$ be such that the non-zero elements of $\mathbf{F}_p$ are $1, a, a^2, \ldots, a^{p-2}$. (It is a general theorem that any finite field has such a *primitive element.*)

Let $k < p$ and let C be the Reed–Solomon code $RS_{p,p-1,k}$ where polynomials are evaluated at $1, a, a^2, \ldots, a^{p-2} \in \mathbf{F}_p$.

(a) Show that if $f \in \mathbf{F}_p[x]$ has degree $< k$ then the cyclic shift of $u(f)$ is $u(g)$ where $g \in \mathbf{F}_p[x]$ is defined by $g(x) = f(a^{-1}x)$. Deduce that C is cyclic.

(b) Let H be the $(n-k) \times n$ matrix defined by

$$H = \begin{pmatrix} 1 & a & a^2 & \ldots & a^{p-2} \\ 1 & a^2 & a^4 & \ldots & a^{2(p-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a^c & a^{2c} & \ldots & a^{c(p-2)} \end{pmatrix}$$

where $c = n - k$. [**Corrected March 28: powers of a were off by one in each row.**] Show that H is a parity check matrix for C. [*Hint: show that $u(x^i)H^{tr} = 0$ for $0 \leq i < k$.*]

(c) Hence, or otherwise, find a generator polynomial of C.

12. (**MSc, MSci**). Let $n \in \mathbf{N}$ and suppose that r divides n.

(a) Show that $g(x) = x^r + 1 \in \mathbf{F}_2[x]$ divides $x^n + 1$.

(b) Find an explicit basis for the the cyclic binary code C of length n with generator polynomial $g(x)$ and determine its dimension and minimum distance.

(c) Show that the dual code of C is cyclic and find its generator polynomial.

The final two questions are in a similar style to the exam questions on the MSc/MSci course. You should also look at the 2010 paper.

13. (**MSc, MSci**)

(a) What does it mean to say that a binary code is *cyclic*?

(b) Explain how codewords of length n in a binary cyclic code can be represented by polynomials of degree $< n$. If $f(x) \in \mathbf{F}_2[x]/(x^n - 1)$ represents $(u_0, u_1, \ldots, u_{n-1})$, show that $(u_{n-1}, u_0, u_1, \ldots, u_{n-2})$ is represented by $xf(x)$.

(c) What is meant by a *generator polynomial* for a cyclic code?

(d) Let C be the binary cyclic code of length 7 with generator polynomial $g(x) = x^3 + x + 1$.

(i) Write down a generator matrix for C.

(ii) Let C^- be the binary code consisting of all codewords in C whose weight is even. Show that C^- is cyclic and find a generator polynomial for C^-.

You may find it helpful to note that $x^7 + 1 = (x+1)(x^3+x+1)(x^3+x^2+1)$ where the factors are irreducible.

14. (**MSc, MSci**) Let n, $k \in \mathbf{N}$ and let p be a prime such that $p \geq n \geq k$. Let $a_1, \ldots, a_n \in \mathbf{F}_p$ be distinct elements.

(a) Define the *Reed–Solomon code* $RS_{p,n,k}$ associated to these parameters.

(b) Show that the minimum distance of the Reed–Solomon code $RS_{p,n,k}$ is at least $n - k + 1$. (Any general results you use to show that two polynomials are equal should be clearly stated, but need not be proved.)

(c) Suppose that $n - k = 2e$ where $e \in \mathbf{N}$. The Key Equation for a received word v is

$$Q(a_i) = v_i E(a_i) \quad \text{for } 1 \leq i \leq n.$$

(i) State upper bounds on the degrees of the polynomials $Q(x)$ and $E(x)$.

Suppose that $u(f) \in RS_{p,n,k}$ is sent and that $v \in \mathbf{F}_p^n$ is received. Let $Q(x)$, $E(x)$ be a solution to the Key Equation.

(ii) Show that if $d(u, v) \leq t$ then $Q(x) = f(x)E(x)$.

(iii) Show conversely that if $Q(x) = f(x)E(x)$ then $d(u, v) \leq t$. [*Hint: first show that if an error occurs in position i then $E(a_i) = 0$.*]