This question paper consists of 2 printed pages, each of which is identified by the reference MATH-3152

Only approved basic scientific calculators may be used.

Mock Examination for the Module MATH-3152

(January 2004)

Coding Theory

Time allowed: 2 hours

Attempt no more than **four** questions. All questions carry equal marks.

1. (a) Let Σ_q be an alphabet of size q and $C \subset \Sigma_q^n$ be a q-ary block code of length n. Define:

   (i) The *Hamming distance* d on Σ_q^n.

   (ii) The *minimum distance* $d(C)$ of the code C.

   (iii) The parameter $A_q(n, d)$.

   (b) State and prove the ball-packing bound on $A_q(n, d)$.

   (c) Prove that $A_q(n, d) \geq A_q(n + 1, d)/q$.

   (d) In each of the following cases *either* construct a code with the specified parameters *or* explain why no such code exists.

   (i) A 7-ary $(5, 550, 3)$ code.

   (ii) A 5-ary $(7, 26, 6)$ code.

   (iii) A 5-ary $(8, 130, 6)$ code.

2. (a) Let C be the ternary linear code with generator matrix

$$G = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

   (i) List the codewords of C and find the minimum distance $d(C)$.

   (ii) Construct a standard array for C. Use your array to decode the received vectors 112 and 120

   (b) Suppose that a binary linear $[9, 4, 3]$ code C is transmitted down a binary symmetric channel with symbol error probability $p < \frac{1}{2}$. Show that $P_{\text{corr}}(C)$, the probability of any transmitted codeword being *correctly* decoded, satisfies

$$P_{\text{corr}}(C) \geq (1 - p)^9 + 9p(1 - p)^8 + 13p^7(1 - p)^2 + 9p^8(1 - p).$$

Continued ...

Given that $p = 0.01$, find an upper bound on the word error rate $P_{\text{err}}(C)$ of the code. Compare your answer with $P_{err}(C_0)$, where $C_0 = \mathbb{Z}_2^4$.

3. Let C be the binary linear code with generator matrix

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

(a) Write down a parity check matrix H for C.

(b) Explain how the minimum distance of C may be deduced from H. Find $d(C)$.

(c) How many cosets does C have? How many cosets are led by weight 1 vectors? Does any coset have a weight 2 coset leader?

(d) Construct a syndrome look-up table for C. Hence, or otherwise, decode the received vectors 100110, 011101 and 101001.

4. (a) Construct the projective equivalence class of the vector $3152 \in \mathbb{Z}_{11}^4$, listing your answer in lexicographical order. [Denote the digit "10" by X, the last digit lexicographically.]

(b) Write down the parity check matrix for the standard Hamming code $C = \mathrm{Ham}(\mathbb{Z}_3^3)$. Determine the parameters $[n, k, d]$ of C. Decode the received vector $1110 \cdots 0 \in \mathbb{Z}_3^n$.

(c) Let $\widehat{C}$ be the following subcode of C:

$$\widehat{C} = \{\mathbf{x} \in C \mid \sum_{i=1}^{n} 2^i x_i = 0 \bmod 3\}.$$

Prove that $\widehat{C}$ is also a linear code. Write down a parity check matrix $\widehat{H}$ for $\widehat{C}$, and determine the parameters $[\widehat{n}, \widehat{k}, \widehat{d}]$ of $\widehat{C}$. Decode the received vector $1110 \cdots 0 \in \mathbb{Z}_3^{\widehat{n}}$.

5. (a) Define the term *cyclic code*.

(b) Determine whether the following codes are cyclic. Briefly explain your answers.

(i) The binary code $\{0000, 1010, 0101, 1110, 1101, 1011, 0111\}$.

(ii) The ternary code $\{000, 011, 101, 110\}$.

(iii) The 7-ary code $\{\mathbf{x} \in \mathbb{Z}_7^5 | \sum_{i=1}^5 ix_i = 0 \bmod 7\}$.

(iv) $E_n \subset \mathbb{Z}_2^n$, the set of even weight binary words of length n.

(v) $O_n \subset \mathbb{Z}_2^n$, the set of odd weight binary words of length n.

(c) (i) Factorize $p(x) = x^5 - 1$ over $\mathbb{Z}_{31}$ into irreducible factors. (Hint: what is $p(2^n)$?)

(ii) For each $k \in \{0, 1, 2, \ldots, 5\}$ let N_k denote the number of distinct 31-ary cyclic codes of length 5 and dimension k. Determine the numbers $N_0, N_1, \ldots, N_5$.

(iii) Choose any one of the cyclic codes of dimension 3 C say. Write down the generator polynomial $g(x)$, the check polynomial $h(x)$, a generator matrix G and a parity check matrix H for C. Determine $d(C)$. Write down the $g^{\perp}(x)$ the generator polynomial of the dual code.

End.

Coding Theory Mock Final Exam: Solution Notes

1. (a) (i) See Defn 3.

   (ii) See Defn 5.

   (iii) See Defn 8.

   (b) See Theorem 11 and its proof.

   (c) This is just question 7 from Problem Set 1. See solutions thereto.

   (d) (i) A 7-ary $(5, 550, 3)$ code cannot exist by ball-packing, for if it did the collection of 550 1-balls centred on codewords would have to be disjoint. But then this collection would contain $550|B_1| = 550(1 + 5 \times 6) = 17050$ words, while the whole space Σ_7^5 contains only $7^5 = 16807$.

   (ii) A 5-ary $(7, 26, 6)$ code cannot exist, though we cannot rule it out by ball-packing ($26|B_2| = 9490 < 5^7$). To show it can't exist we appeal to the Singleton bound (Theorem 9). Since all codewords differ from one another in at least 6 places, the 2-digit words we get by deleting the last 5 digits of each codeword must all be distinct. Hence a 5-ary $(7, M, 6)$ code must have $M \leq 5^2 = 25$.

   (iii) A 5-ary $(8, 130, 6)$ code does not exist by (ii) and part (c).

2. (a) (i) $C = \{000, 101, 202, 012, 110, 211, 021, 122, 220\}$. Since C is linear $d(C)$ is the minimum **weight** of nonzero codewords, which is 2.

   (ii) We list the $|Z_3^3|/|C| = 27/9 = 3$ cosets of C in the obvious order, choosing minimum weight coset leaders. There are 9 different ways to choose the non-zero coset leaders, hence 9 different standard arrays are possible. Here is one of them:

000	101	202	012	110	211	021	122	220
100	201	002	112	210	011	121	222	020
200	001	102	212	010	111	221	022	120

   Note that both the nonzero coset leaders have weight 1, but that their cosets in each case contain 2 other weight 1 vectors – either of these would be a valid choice of coset leader too. We decode a received vector as the codeword at the top of its column, so $112 \mapsto 012$ and $120 \mapsto 220$ with my choice of array. You'll get different answers if you choose the coset leaders differently. Note that with my choice, the code always corrects by changing only the first digit.

(b) Let γ_r, $r = 0, 1, \ldots, 9$ denote the number of coset leaders of weight r. As always, $\gamma_0 = 1$. Further, since $d = 3 \geq 2 \times 1 + 1$, we know that every weight 1 vector is a coset leader, so $\gamma_1 = 9$. There are $2^9/2^4 = 32$ cosets in total, so

$$\sum_{r=2}^{9} \gamma_r = 22.$$

Recall that

$$P_{\text{corr}}(C) = \sum_{r=0}^{9} \gamma_r p^r (1-p)^{9-r}.$$

Now since $p < \frac{1}{2}$, we know that $(p-1) > p$ and hence that $r \geq s$ implies,

$$p^r(1-p)^{7-r} \leq p^2(1-p)^{7-2}$$

So we obtain a *lower* bound on $P_{\text{corr}}(C)$ by assuming that the remaining 22 coset leaders have the *largest* weights possible. Clearly 111111111 cannot lead its coset (it has *maximal* weight) so $\gamma_9 = 0$. Bearing in mind that there $\left(\begin{array}{c} 9 \\ r \end{array} \right)$ words of weight r, we see that the "worst case scenario" is $\gamma_9 = 0$, $\gamma_8 = 9$, $\gamma_7 = 13$, $\gamma_6 = \gamma_5 = \cdots = \gamma_2 = 0$. Hence

$$P_{\text{corr}}(C) \geq (1-p)^9 + 9p(1-p)^8 + 13p^7(1-p)^2 + 9p^8(1-p).$$

If $p = 0.01$ this gives $P_{\text{corr}}(C) \geq 0.99656427$ and hence

$$P_{\text{err}}(C) = 1 - P_{\text{corr}}(C) \leq 0.0034357300.$$

Using the trivial code $C_0 = \mathbb{Z}_2^4$, a received vector is correctly decoded if and only if no errors occur. Hence $P_{\text{err}}(C_0) = 1 - (1-p)^4 = 0.03940399$. So C is much more reliable, but less than half as fast to transmit as C_0

3. (a)

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

(b) $d(C) = d$ if and only if no set of $(d-1)$ columns of H is linearly dependent, but some set of d columns is. In this case, no pair of columns is linearly dependent (there are no parallel pairs) but the first 3 columns sum to 000, hence are linearly dependent. So $d(C) = 3$.

(c) C has $|\mathbb{Z}_2^6|/|C| = 2^6/2^3 = 8$ cosets. Since $d(C) = 3$, every weight 1 vector is a coset leader (see Corollary 57), so 6 of the cosets have weight 1 leaders. C itself is led by 000000, so that leaves 1 coset unaccounted for. There are $\left(\begin{array}{c} 6 \\ 2 \end{array} \right) = 15$ weight 2 vectors in Z_2^6, and the question is whether all of these are contained in the cosets with weight 1 leaders. But in order to lie in coset $\mathbf{v}+C$ with $w(\mathbf{v}) = 1$, a vector must lie distance 1 from one of the codewords of C. Now

$$C = \{000000, 100101, 010110, 110011, 001011, 101110, 011101, 111000\}$$

so the only weight 2 vectors with this property are

$$000101, 100001, 100100, 000110, 010010, 010100, 000011, 001001,$$

$$001010, 011000, 101000, 110000.$$

Hence the remaining weight 2 vectors, namely 100010, 010001 and 001100 must lie in the 8th coset, and any of these 3 may be chosen as the coset leader.

(d) We compute $S(\mathbf{v}_r) = \mathbf{v}_r H^T$ for each of the coset leaders $\mathbf{v}_1, \ldots, \mathbf{v}_8$. Since there are 3 different choices for $\mathbf{v}_8$ 3 different tables are possible – they differ only in the last row. Choosing 100010 as our sole weight 2 coset leader, we obtain:

coset leader	syndrome
000000	000
100000	101
010000	110
001000	011
000100	100
000010	010
000001	001
100010	111

$S(100110) = 011$ so 100110 lies in $001000+C$. Hence we correct it by subtracting 1 from its 3rd digit: $100110 \mapsto 101110$.
$S(011101) = 000$ so 011101 is a codeword and needs no correction.
$S(101001) = 111$ so 101001 lies in $100010 + C$. Hence we correct it $101001 \mapsto 101001 - 100010 = 001011$. Note that your answer will be different if you chose a different weight 2 coset leader.

4. (a) $[3152] = \{\lambda(3, 1, 5, 2) \mid \lambda \in \mathbb{Z}_{11}, \lambda \neq 0\}$. Hence

$$[3152] = \{1498, 2875, 3152, 453X, 5917, 62X4, 7681, 8X69, 9346, X723\}$$

in lexicographical order.

(b) Choose the unique vector from each projective equivalence class in $\mathbb{Z}_3^3$ whose first nonzero digit is 1 and assemble these in lexicographical order as the columns of H:

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{bmatrix}$$

It follows that $n = 13$ (number of columns of H), the redundancy of C is 3 (number of rows of H), so $k = 10$, and since C is a Hamming code, $d = 3$ by

construction.
1110000000000 has syndrome $1110000000000H^T = 022 = 2\mathbf{c}_3$ where $\mathbf{c}_i$ are the columns of H. Hence we decode it as

$$1110000000000 \mapsto 11100000000000 - 0020000000000 = 1120000000000.$$

(c) Let $\mathbf{x}, \mathbf{y} \in \widehat{C}$. Then $\sum 2^i x_i = \sum 2^i y_i = 0$, and $\mathbf{x}, \mathbf{y} \in C$, so $\mathbf{x} + \mathbf{y} \in C$ since C is linear. Further

$$\sum 2^i (x_i + y_i) = \sum 2^i x_i + \sum 2^i y_i = 0 + 0 = 0$$

so $\mathbf{x} + \mathbf{y} \in \widehat{C}$.
Let $a \in \mathbb{Z}_3$. Then $a\mathbf{x} \in C$ since C is linear, and

$$\sum 2^i a x_i = a \sum 2^i x_i = a \times 0 = 0$$

so $a\mathbf{x} \in \widehat{C}$. Since $\widehat{C}$ is closed under both vector addition and scalar multiplication, it is a linear code.□
We may construct $\widehat{H}$ by appending an extra row to H representing the extra parity check equation $\sum 2^i x_i = 0$, so the most obvious choice is

$$\widehat{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 \end{bmatrix}$$

The block length of the code remains unchanged, $\widehat{n} = 13$, but its redundancy is greater by 1, so its dimension is $\widehat{k} = 9$. Since $\widehat{C} \subset C$ we know that $d(\widehat{C}) \geq d(C) = 3$. In fact if we sum the 1st, 3rd and 4th columns of $\widehat{H}$ we get 0000, so there is a set of 3 linearly dependent columns and we conclude that $\widehat{d} = d = 3$.
Now 1110000000000 has syndrome $1110000000000\widehat{H}^T = 0222 = \widehat{\mathbf{c}}_2 + \widehat{\mathbf{c}}_4$ so we decode it as

$$1110000000000 \mapsto 1110000000000 - 0101000000000 = 1012000000000.$$

5. (a) See Defn 87.
   (b) (i) Not cyclic since not linear (e.g. $1010 + 0101 = 1111$ is not in the code).
   (ii) Not cyclic since not linear (e.g. $2 \times 011 = 022$ is not in the code).
   (iii) Not cyclic since not closed under cyclic shift (e.g. 02100 is in the code but 21000 is not).
   (iv) E_n is linear, and cyclic shift leaves the weight of a vector unchanged, so it's closed under cyclic shift. Hence E_n is cyclic.
   (v) O_n is not cyclic since it isn't linear (e.g. it doesn't contain the zero vector).

(c) (i) Note that $p(2^n) = 2^{5n} - 1 = 32^n - 1 = 1^n - 1 = 0$ for all n. Hence $p(1) = p(2) = p(4) = p(8) = p(16) = 0$ and we deduce that there are 5 associated linear factors. But since $\deg p(x) = 5$, this completely determines the factorization:

$$x^5 - 1 = (x-1)(x-2)(x-4)(x-8)(x-16)$$

over $\mathbb{Z}_{31}$.

(ii) It follows that there are $2^5 = 32$ distinct cyclic codes of length 5 over $\mathbb{Z}_{31}$, determined by the generator polynomials

$$g(x) = (x-1)^{m_1}(x-2)^{m_2}(x-4)^{m_3}(x-8)^{m_4}(x-16)^{m_5}$$

where each $m_i \in \{0,1\}$. The code $\langle g(x) \rangle$ has dimension $k = 5 - m_1 - m_2 - \cdots - m_5$, so the number of codes of dimension k is the number of ways of choosing $5-k$ nonzero exponents out of 5. Hence

$$N_k = \left(\begin{array}{c} 5 \\ 5-k \end{array} \right) = \left(\begin{array}{c} 5 \\ k \end{array} \right) = \frac{5!}{k!(5-k)!}.$$

(iii) There are $N_3 = 10$ different dimension 3 codes, so lots of answers are possible here. I'll choose the code generated by the degree 2 polynomial

$$g(x) = (x-1)(x-2) = 2 - 3x + x^2.$$

This has check polynomial

$$h(x) = (x-4)(x-8)(x-16) = -16 + 6x - 28x^2 + x^3 = 15 + 6x + 3x^2 + x^3,$$

generator matrix

$$G = \left[\begin{array}{ccccc} 2 & -3 & 1 & 0 & 0 \\ 0 & 2 & -3 & 1 & 0 \\ 0 & 0 & 2 & -3 & 1 \end{array} \right]$$

(see Theorem 107) and parity check matrix

$$H = \left[\begin{array}{ccccc} 1 & 3 & 6 & 15 & 0 \\ 0 & 1 & 3 & 6 & 15 \end{array} \right]$$

(see Theorem 113). Since no pair of columns of H is parallel, and every triple of columns is linearly dependent on dimensional grounds, we deduce that $d = 3$.

The reciprocal of the check polynomial is

$$\overline{h}(x) = 1 + 3x + 6x^2 + 15x^3$$

which is not monic. Note that $15 \times 2 = 30 = -1$ so $15^{-1} = -2$. Hence

$$g^{\perp}(x) = -2 - 6x - 12x^2 + x^3.$$