MATHEMATICAL TRIPOS PART II 2013
CODING AND CRYPTOGRAPHY
EXAMPLE SHEET 3

The first 15 examples are minimal to cover the course; you are also encouraged to try questions 16–17.

1 Write down the weight enumerators of the trivial code (that is to say, $\mathbb{F}_2^n$), the zero code (that is to say, $\{\mathbf{0}\}$), the repetition code and the simple parity code.

2 List the codewords of the Hamming (7,4) code and its dual. Write down the weight enumerators and verify that they satisfy the MacWilliams identity.

3 (a) Show that if C is linear, then so are its extension C^+, truncation C^- and puncturing C', provided the symbol chosen to puncture by is 0. Give an example to show that C' may not be linear if we puncture by 1.

(b) Show that extension followed by truncation does not change a code. Is this true if we replace 'truncation' by 'puncturing'?

(c) Give an example where puncturing reduces the information rate and an example where puncturing increases the information rate.

(d) Show that the minimum distance of the parity extension C^+ is the least even integer n with $n \geq d(C)$.

(e) Show that the minimum distance of the truncation C^- is $d(C)$ or $d(C) - 1$ and that both cases can occur.

(f) Show that puncturing cannot decrease the minimum distance, but give examples to show that the minimum distance can stay the same or increase.

4 Show that if $2^k \sum_{i=0}^{d-2} \binom{n-1}{i} < 2^n$ then $A(n, d) \geqslant 2^k$. Compare this with the GSV bound in the case $n = 10$ and $d = 3$. [Hint: construct a parity check matrix for a linear code by choosing one row at a time.]

5 The Mariner mission to Mars[1] used the $\mathrm{RM}(5, 1)$ code. What is its information rate? What proportion of errors could it correct in a single codeword? How does it compare to the Hamming code of length 31 ($= 2^5 - 1$)?

6 Show that the weight enumerator of $\mathrm{RM}(d, 1)$ is

$$t^{2^d} + (2^{d+1} - 2)s^{2^{d-1}}t^{2^{d-1}} + s^{2^d}.$$

7 (i) Show that every codeword in $\mathrm{RM}(d, d-1)$ has even weight.

(ii) Show that $\mathrm{RM}(m, m-r-1) \subseteq \mathrm{RM}(m, r)^\perp$.

(iii) By considering dimension, or otherwise, show that $\mathrm{RM}(m, r)$ has dual code $\mathrm{RM}(m, m-r-1)$.

[1] Launched by NASA from Cape Canaveral on 30 May 1971, Mariner 9 was the first spacecraft to orbit another planet, narrowly beating Soviet Mars 2 and Mars 3, which both arrived within a month. After 349 days in orbit, Mariner 9 had transmitted 7,329 images, covering 100% of Mars' surface. It still orbits Mars in an orbit which will eventually decay in 2022.

8 If there is a perfect e-error correcting binary code of length n, show that $V(n,e)$ divides 2^n. This condition is not sufficient for such a code to exist. We prove this by establishing the following results.

(i) Verify that $\frac{2^{90}}{V(90,2)} = 2^{78}$.

(ii) Suppose that C is a perfect 2-error correcting binary code of length 90 and size 2^{78}. Explain why we may suppose, without loss of generality, that the zero word $\mathbf{0} \in C$.

(iii) Let C be as in (ii) with $\mathbf{0} \in C$. Consider the set

$$X = \{\mathbf{x} \in \mathbb{F}_2^{90} : x_1 = 1,\ x_2 = 1, d(\mathbf{0}, \mathbf{x}) = 3\}.$$

Show that, corresponding to each $\mathbf{x} \in X$, we can find a unique $\mathbf{c}(\mathbf{x}) \in C$ such that $d(\mathbf{c}(\mathbf{x}), \mathbf{x}) = 2$. Show that $d(\mathbf{c}(\mathbf{x}), \mathbf{0}) = 5$.

(iv) Continuing with the argument of (iii), show that $c_i(\mathbf{x}) = 1$ whenever $x_i = 1$. If $\mathbf{y} \in X$, find the number of solutions to the equation $\mathbf{c}(\mathbf{x}) = \mathbf{c}(\mathbf{y})$ with $\mathbf{x} \in X$ and, by considering the number of elements of X, obtain a contradiction.

This result, obtained by Marcel Golay, shows that there is no perfect $(90, 2^{78})$-code. He found another case when $2^n/V(n,e)$ is an integer and there *does* exist an associated perfect code (the *Golay code*) - see question 18 below[2].

9 **[The MacWilliams identity for binary codes]**

Let $C \subseteq \mathbb{F}_2^n$ be a linear code of dimension k.

(i) Show that

$$\sum_{\mathbf{x} \in C} (-1)^{\mathbf{x}.\mathbf{y}} = \begin{cases} 2^k & \text{if } \mathbf{y} \in C^\perp \\ 0 & \text{if } \mathbf{y} \notin C^\perp. \end{cases}$$

(ii) If $t \in \mathbb{R}$, show that

$$\sum_{\mathbf{y} \in \mathbb{F}_2^n} t^{w(\mathbf{y})} (-1)^{\mathbf{x}.\mathbf{y}} = (1-t)^{w(\mathbf{x})} (1+t)^{n-w(\mathbf{x})}.$$

(iii) By using parts (i) and (ii) to evaluate

$$\sum_{\mathbf{x} \in C} \left(\sum_{\mathbf{y} \in \mathbb{F}_2^n} (-1)^{\mathbf{x}.\mathbf{y}} \left(\frac{s}{t}\right)^{w(\mathbf{y})} \right)$$

in two different ways, obtain the MacWilliams identity

$$W_{C^\perp}(s,t) = 2^{-\dim C} W_C(t-s, t+s).$$

10 An *erasure* is a digit which has been made unreadable in transmission. Why are they easier to deal with than errors? Find a necessary and sufficient condition on the parity check matrix for it to be always possible to correct t erasures. Find a necessary and sufficient condition on the parity check matrix for it never to be possible to correct t erasures (ie whatever message you choose and whatever t erasures are made the recipient cannot tell what you sent).

[2] The deep connections between the Golay code and certain Mathieu groups (a class of sporadic finite simple groups) is beyond the scope of this course. See the great little book *From error correcting codes through sphere packings to simple groups* by (I kid you not) Thomas Thompson of Walla Walla College (Carus Mathematical Monographs, 1983).

11 (i) Consider the collection K of polynomials $a_0 + a_1\omega$ with $a_j \in \mathbb{F}_2$ manipulated subject to the usual rules of polynomial arithmetic and to the further condition $1+\omega+\omega^2 = 0$. Show by direct calculation that $K^\times = K \setminus \{0\}$ is a cyclic group under multiplication and deduce that K is a finite field.

(ii) Repeat (i) where this time the collection K of polynomials is $a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3$ and the further condition is replaced by $1 + \omega + \omega^4 = 0$.

[Of course, this question follows directly from general theory but such direct calculations are not uninstructive.]

12 (i) Identify the cyclic codes of length n corresponding to each of the polynomials 1, $X-1$ and $X^{n-1} + X^{n-2} + \cdots + X + 1$.

(ii) Factor the polynomials $X^3 - 1$ and $X^5 - 1$ into irreducibles in $\mathbb{F}_2[X]$. Hence find all cyclic codes of length 3 or 5 and relate them to codes you have already met.

(iii) Show that there are three cyclic codes of length 7 corresponding to irreducible polynomials of which two are versions of Hamming's original code. What are the other cyclic codes of length 7? You should relate them to codes you have already met.

13 Prove the following results:

(i) If K is a field containing $\mathbb{F}_2$, then $(a+b)^2 = a^2 + b^2$ for all $a, b \in K$.

(ii) If $P \in \mathbb{F}_2[X]$ and K is a field containing $\mathbb{F}_2$, then $P(a)^2 = P(a^2)$ for all $a \in K$.

(iii) Let K be a field containing $\mathbb{F}_2$ in which $X^7 - 1$ factorises into linear factors. If β is a root of $X^3 + X + 1$ in K, then β is a primitive root of unity and β^2 is also a root of $X^3 + X + 1$.

(iv) We continue with the notation of (iii). Show that the BCH code with $\{\beta, \beta^2\}$ as defining set is Hamming's original (7,4) code.

14 Let $\omega \in \mathbb{F}_{16}$ be a root of $X^4 + X + 1$. Let C be the BCH code of length 15 and design distance 5, with defining set $\omega, \omega^2, \omega^3, \omega^4$.

(i) Find the minimal polynomial for each element of the defining set, and then compute the generator polynomial of C as the least common multiple of these polynomials.

(ii) If possible, determine the error positions of the following received words:

(a) $r(X) = 1 + X^6 + X^7 + X^8$;

(b) $r(X) = 1 + X + X^4 + X^5 + X^6 + X^9$;

(c) $r(X) = 1 + X + X^2$;

(d) $r(X) = 1 + X + X^7$.

[Your answer to qn 11 (ii) may help with the computations.]

15 Let C be a binary linear code of length n, rank k and distance d.

(i) Show that C contains a codeword $\mathbf{x}$ with exactly d non-zero digits.

(ii) Show that $n \geq d + k - 1$ (the Singleton bound).

(iii) Prove that truncating C on the non-zero digits of $\mathbf{x}$ produces a code C' of length $n - d$, rank $k - 1$ and distance $d' \geq \lceil \frac{d}{2} \rceil$.
[Hint: assume the opposite. Then, given $\mathbf{y} \in C$, and its truncation $\mathbf{y}' \in C'$, consider the coordinates where $\mathbf{x}$ and $\mathbf{y}$ have 1 in common (i.e. where $x_j = y_j = 1$) and where they differ (e.g. $x_j = 1$ and $y_j = 0$).]

(iv) Deduce that

$$n \geq d + \sum_{u=1}^{k-1} \lceil \tfrac{d}{2^u} \rceil$$

(an improved Singleton bound). Why does (iv) imply (ii)? Give an example where $n > d + k - 1$.
[Remark: Codes for which $n - k = d - 1$ are called MDS (*maximum distance separable*) codes. A non-trivial, non-binary example is the Reed-Solomon code and its extensions.]

16 (i) Prove that a binary 2-error correcting code of length 10 can have at most 12 codewords.

Now let p be a prime congruent to 3 modulo 4 and let Q be the set of squares (=quadratic residues) mod p, including 0, so that $|Q| = \frac{p+1}{2}$.

(ii) Show that Q and $Q + 1$ have exactly $\frac{p+1}{4}$ elements in common and deduce that for any pair of elements mod p, there are $\frac{p+1}{4}$ translates (sets of the form $Q + j$) which contain both.

Consider the code of length p and size $p + 1$ whose $(j + 1)$th element is $(x_0, \ldots, x_{p-1})$ where $x_r = 0$ if and only if $r \in Q + j$, $(j = 0, \ldots, p - 1)$, and whose $(p + 1)$th element is $(1, 1, \ldots, 1)$. What is the distance between two distinct codewords? What can one say about the distance between codewords in the truncation of this code?

(iii) Deduce the existence of a $[10, 12]$ 2-error correcting code.[3] See also Sheet 2, question 8. The codes derived this way are not linear.

[3] The related (11,12,6)-code is called the *Paley 2-design.* It is named after Raymond E.A.C. Paley, an MIT mathematician who worked with Norbert Wiener. Paley died in an avalanche in 1933 aged just 26 while skiing in the Canadian Rockies.

17 We construct a perfect 3-error-correcting (23,12)-code, starting from the factorisation

$$X^{23} - 1 = (X - 1)f_1(X)f_2(X)$$

in $\mathbb{F}_2[X]$ where $f_1(X) = 1 + X + X^5 + X^6 + X^7 + X^9 + X^{11}$ and $f_2(X) = X^{11}f_1(1/X)$ is the polynomial obtained from f_1 by reversing the sequence of coefficients.

(i) Show that if $g(X) \in \mathbb{F}_2[X]$ then $g(X)^2 = g(X^2)$. What does this tell you about the roots of g in any field extension of $\mathbb{F}_2$?

(ii) Make a list of the powers of 2 mod 23. Deduce that the cyclic code C with generator polynomial $f_1(X)$ has minimum distance at least 5. [Hint: identify C as a BCH code.]

(iii) Show that $C^{\perp}$ is a subcode of C. Deduce that the parity check extension of C is a self-dual linear code.

(iv) Show that any self-dual linear code, generated by vectors of weight divisible by 4, has minimum distance a multiple of 4.

(v) Deduce that C is a perfect 3-error correcting code.

SM, Lent Term 2013
Comments on and corrections to this sheet may be emailed to `sm@dpmms.cam.ac.uk`