Finite geometry and coding theory

Peter J. Cameron
School of Mathematical Sciences
Queen Mary and Westfield College
London E1 4NS

Socrates Intensive Programme
Finite Geometeies and Their Automorphisms
Potenza, Italy
June 1999

Abstract

In these notes I will discuss some recent developments at the interface between finite geometry and coding theory. These developments, are all based on the theory of quadratic forms over GF(2), and I have included an introduction to this material. The particular topics are bent functions and difference sets, multiply-resolved designs, codes over $\mathbb{Z}_4$, and quantum error-correcting codes.

Some of the material here was discussed in the Combinatorics Study Group at Queen Mary and Westfield College, London, over the past year. This account is quite brief; a more detailed version will be published elsewhere. Many of the participants of the study group contributed to this presentation; to them I express my gratitude, but especially to Harriet Pollatsek and Keldon Drudge.

1 Codes

There are many good accounts of coding theory, so this section will be brief. See MacWilliams and Sloane [18] for more details.

Let A be an alphabet of q symbols. A *word* of length n over A is simply an n-tuple of elements of A (an element of A^n). A *code* of length n over A

is a set of words (a subset of A^n). A code of length n containing M words is referred to as an (n, M) code.

The *Hamming distance* $d(v, w)$ between two words v and w is the number of coordinates in which v and w differ:

$$d(v, w) = |\{i : 1 \leq i \leq n, v_i \neq w_i\}|.$$

It satisfies the standard axioms for a metric on A^n. The *minimum distance* of a code is the smallest distance between two distinct words in C. A code of length n having M codewords and minimum distance d is referred to as an (n, M, d) code.

The basic idea of coding theory is that, in a communication system, messages are transmitted in the form of words over a fixed alphabet (in practice, usually the binary alphabet $\{0, 1\}$). During transmission, some errors will occur, that is, some entries in the word will be changed by random noise. The number of errors occurring is the Hamming distance between the transmitted and received words.

Suppose that we can be reasonably confident that no more than e errors occur during transmission. Then we use a code whose minimum distance d satisfies $d \geq 2e + 1$. Now we transmit only words from the code C. Suppose that u is transmitted and v received. By assumption, $d(u, v) \leq e$. If u' is another codeword, then $d(u, u') \geq d \geq 2e + 1$. By the triangle inequality, $d(u', v) \geq e+1$. Thus, we can recognise the transmitted word u, as the codeword nearest to the received word. For this reason, a code whose minimum distance d satisfies $d \geq 2e + 1$ is called an *e-error-correcting code.*

Thus, good error correction means large minimum distance. On the other hand, fast transmission rate means many codewords. Increasing one of these parameters tends to decrease the other. This tension is at the basis of coding theory.

Usually, it is the case that the alphabet has the structure of a finite field $\mathrm{GF}(q)$. In this case, the set of words is the n-dimensional vector space $\mathrm{GF}(q)^n$, and we often require that the code C is a vector subspace of $\mathrm{GF}(q)^n$. Such a code is called a *linear code.* A linear code of length n and dimension k over $\mathrm{GF}(q)$ is referred to as a $[n, k]$ code; it has q^k codewords. If its minimum distance is d, it is referred to as an $[n, k, d]$ code. Almost always, we will consider only linear codes.

The *weight* $\mathrm{wt}(v)$ of a word v is the number of non-zero coordinates of v. The *minimum weight* of a linear code is the smallest weight of a nonzero codeword. It is easy to see that

$$d(v,w) = \mathrm{wt}(v-w),$$

and hence that the minimum distance and minimum weight of a linear code are equal.

Two linear codes C and C' are said to be *equivalent* if C' is obtained from C by a combination of the two operations:

(a) multiply the coordinates by non-zero scalars (not necessarily all equal); and

(b) permute the coordinates.

Equivalent codes have the same length, dimension, minimum weight, and so on. Note that, over $\mathrm{GF}(2)$, operation (a) is trivial, and we only need operation (b).

A linear $[n,k]$ code C can be specified in either of two ways:

- A *generator matrix* G is a $k \times n$ matrix whose row space is C. Thus,

  $$C = \{xG : x \in \mathrm{GF}(q)^k\},$$

  and every codeword has a unique representation in the form xG. This is useful for encoding: if the messages to be transmitted are all k-tuples over the field $\mathrm{GF}(q)$, then we can encode the message x as the codeword xG.

- A *parity check matrix* H is a $(n-k) \times n$ matrix whose null space is C: more precisely,

  $$C = \{v \in \mathrm{GF}(q)^n : vH^\top = 0\}.$$

  This is useful for decoding, specifically for *syndrome decoding*. The *syndrome* of $w \in \mathrm{GF}(q)^n$ is the $(n-k)$-tuple $wH^\top$. Now, if C corrects e errors, and w has Hamming distance at most e from a codeword v, it can be shown that the syndrome of w uniquely determines $w - v$, and hence v.

If $h_1, \ldots, h_n$ are the columns of the parity check matrix H of a code C, then a word $x = (x_1, \ldots, x_n)$ belongs to C if and only if $x_1h_1 + \cdots + x_nh_n = 0$, that is, the entries in x are the coefficients in a linear dependence relation between the columns of H. Thus, we have:

Proposition 1.1 *A code C has minimum weight d or greater if and only if any $d-1$ columns of its parity check matrix are linearly independent.*

There is a natural inner product defined on $\mathrm{GF}(q)^n$, namely the *dot product*

$$v \cdot w = \sum_{i=1}^{n} v_i w_i.$$

If C is an $[n, k]$ code, we define the *dual code*

$$C^{\perp} = \{v \in \mathrm{GF}(q)^n : (\forall w \in C)\ v \cdot w = 0\};$$

it is an $[n, n-k]$ code. Then a generator matrix for $C^{\perp}$ is a parity check matrix for C, and *vice versa.*

Sometimes, in the case when q is a square, so that the field $\mathrm{GF}(q)$ admits an automorphism σ of order 2 given by $x^{\sigma} = x^{\sqrt{q}}$, we will use instead the *Hermitian inner product*

$$v \circ w = \sum_{i=1}^{n} v_i w_i^{\sigma}.$$

We now give a family of examples.

A 1-error-correcting code should have minimum weight at least 3. By Proposition 1.1, this is equivalent to requiring that no two columns of its parity check matrix are linearly dependent. Thus, the columns should all be non-zero, and should span distinct 1-dimensional subspaces of $V = \mathrm{GF}(q)^k$, where k is the codimension of the code. Multiplying columns by non-zero scalars, or permuting them, gives rise to an equivalent code. So linear 1-error-correcting codes correspond in a natural way to sets of points in the projective space $\mathrm{PG}(k-1, q)$.

Many interesting codes can be obtained by choosing suitable subsets (ovoids, unitals, etc.). But the simplest, and optimal, choice is to take all the points of the projective space. The code thus obtained is the *Hamming code* $\mathcal{H}(k, q)$ of length $n = (q^k - 1)/(q-1)$ and dimension $n-k$ over $\mathrm{GF}(q)$. To reiterate: the parity check matrix of the Hamming code is the $k \times n$ matrix whose columns span the n one-dimensional subspaces of $\mathrm{GF}(q)^k$. It is a $[n, n-k, 3]$ code.

One further code we will need is the famous *extended binary Golay code.* This is a $[24, 12, 8]$ code over $\mathrm{GF}(2)$, and is the unique code (up to equivalence) with this property. The *octads*, or sets of eight coordinates which support words of weight 8 in the code) are the blocks of the Steiner system $S(5, 8, 24)$ (or, in other terminology, the 5-$(24, 8, 1)$ design).

We discuss briefly some operations on codes. Let C be a linear code.

- *Puncturing* C in a coordinate i is the process of deleting the ith coordinate from each codeword.
- *Shortening* C in a coordinate i is the process of selecting those codewords in C which have entry 0 in the ith coordinate, and then deleting this coordinate from all codewords.
- *Extending* C, by an *overall parity check*, is the process consisting of adding a new coordinate to each codeword, the entry in this coordinate being minus the sum of the existing entries (so that the sum of all coordinates in the extended code is zero). We denote the extension of C by $\overline{C}$.
- The *direct sum* of codes C_1 and C_2 is the set of all words obtained by concatenating a word of C_1 with a word of C_2.

The *weight enumerator* of a linear code is an algebraic gadget to keep track of the weights of codewords. If C has length n, its weight enumerator is

$$W_C(x, y) = \sum_{i=0}^{n} a_i x^{n-i} y^i,$$

where A_i is the number of words of weight i in C. Note that $W_C(1, 0) = 1$, and $W_C(1, 1) = |C|$.

The weight enumerator has many important properties. For example, the weight enumerator of the direct sum of C_1 and C_2 is the product of the weight enumerators of C_1 and C_2. For our purposes, the most important result is *MacWilliams' Theorem*:

Theorem 1.2 *Let C be a linear code over* $\mathrm{GF}(q)$*, and $C^\perp$ its dual. Then*

$$W_{C^\perp}(x, y) = \frac{1}{|C|} W_C(x + (q-1)y, x - y).$$

We illustrate this theorem by calculating the weight enumerators of Hamming codes. It is easier to find the weight enumerators of their duals:

Proposition 1.3 *Let C be the q-ary Hamming code $\mathcal{H}(k,q)$ of length $n = (q^k-1)/(q-1)$ and dimension $n-k$. Then every non-zero word of $C^\perp$ has weight q^{k-1}.*

We say that $C^\perp$ is a *constant-weight code.*

Proof Let $h_1, \ldots, h_n$ be the columns of the parity check matrix H of C. We claim that each word of $C^\perp$ has the form $(f(h_1), \ldots, f(h_n))$, where f belongs to the dual space of the k-dimensional space V of column vectors of length k; and every element of V^* gives rise to a unique word of $C^\perp$. This holds because H is a generator matrix of $C^\perp$, so the words of $C^\perp$ are linear combinations of the rows of H. Now the ith row of H has the form $(e_i(h_1), \ldots, e_i(h_n))$, where e_i is the ith dual basis vector. So the claim is proved.

Now for any non-zero $f \in V^*$, the kernel of f has dimension $k-1$, and so contains $(q^{k-1}-1)/(q-1)$ one-dimensional subspaces, and so it vanishes at this many of the columns of H. So the corresponding word of $C^\perp$ has weight

$$(q^k-1)/(q-1) - (q^{k-1}-1)/(q-1) = q^{k-1}.$$

It follows that the weight enumerator of $C^\perp$ is

$$x^{(q^k-1)/(q-1)} + (q^k-1)x^{(q^{k-1}-1)/(q-1)}y^{q^{k-1}},$$

and so the weight enumerator of C is

$$\frac{1}{q^k}\left((x+(q-1)y)^{(q^k-1)/(q-1)} + (q^k-1)(x+(q-1)y)^{(q^{k-1}-1)/(q-1)}(x-y)^{q^{k-1}}\right).$$

Finally on this topic, we mention that the weight enumerator of the extended binary Golay code is

$$x^{24} + 759x^{16}y^8 + 2576x^{12}y^{12} + 759x^8y^{16} + y^{24}.$$

A code C is *self-orthogonal* if $C \subseteq C^\perp$, and is *self-dual* if $C = C^\perp$. The extended binary Golay code just mentioned is self-dual; other examples are the extended binary Hamming code of length 8 (a $[8,4,4]$ code with weight enumerator $x^8 + 14x^4y^4 + y^8$), and the binary repetition code of length 2 (a $[2,1,2]$ code with weight enumerator $x^2 + y^2$).

Using MacWilliams' Theorem, we see that the weight enumerator of a self-dual code C of length n over $\mathrm{GF}(q)$ satisfies

$$W_C(x,y) = \frac{1}{q^{n/2}} W_C(x+(q-1)y, x-y). \tag{1}$$

This gives a system of equations for the coefficients of W_C, but of course not enough equations to determine it uniquely.

Gleason [13] found a simple description of all solutions of these equations, using classical invariant theory. We describe his technique for self-dual binary codes.

Let G be a finite group of 2×2 matrices over $\mathbb{C}$. Let $f(x,y)$ be a polynomial of degree n. We say that f is an *invariant* of G if

$$f(ax+by, cx+dy) = f(x,y) \text{ for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G.$$

Since the sum and product of invariants is invariant, the set of G-invariants is a subalgebra of the algebra $\mathbb{C}[x,y]$ of all polynomials in x and y over $\mathbb{C}$. We denote this subalgebra by $\mathbb{C}[x,y]^G$.

If $f(x,y)$ is a G-invariant, then its homogeneous component of degree k (the sum of all terms $a_{ij}x^iy^j$ with $i+j=k$) is also G-invariant. So the algebra $\mathbb{C}[x,y]^G$ is *graded*, according to the following definition:

Let $A = \bigoplus_{k\geq 0} A_k$ be an algebra over $\mathbb{C}$. We say that A is *graded* if $A_i \cdot A_j \subseteq A_{i+j}$ for all $i,j \geq 0$. If $\dim(A_k)$ is finite for all $k \geq 0$, then the *Hilbert series* of A is the formal power series

$$\sum_{k\geq 0} \dim(A_k) t^k.$$

Molien's Theorem gives an explicit formula for the Hilbert series of $\mathbb{C}[x,y]^G$ for any finite group G:

Theorem 1.4 *Let G be a finite group of 2×2 matrices over $\mathbb{C}$. Then the Hilbert series of $\mathbb{C}[x,y]^G$ is given by*

$$\frac{1}{|G|}\sum_{A\in G} (\det(I-tA))^{-1}.$$

Now let C be a self-dual binary code. Since all words in C have even weight, the weight enumerator of C satisfies

$$W_C(x,-y) = W_C(x,y).$$

Also, since $W_C(x,y)$ is homogeneous of degree n, we can rewrite Equation 1 as

$$W_C\left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right) = W_C(x,y).$$

These two equations assert that the polynomial W_C is an invariant of the group $G = \langle A_1, A_2\rangle$, where

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$

Now it is easily checked that

$$A_1^2 = A_2^2 = (A_1A_2)^8 = I,$$

so G is a dihedral group of order 16. Now Molien's Theorem, and some calculation, shows that the Hilbert series of $\mathbb{C}[x,y]^G$ is

$$\frac{1}{(1-t^2)(1-t^8)}.$$

¿From this we see that the dimension of the nth homogeneous component is equal to the number of ways of writing n as a sum of 2s and 8s.

We know some examples of self-dual codes: among them, the repetition code of length 2 and the extended Hamming code of length 8, with weight enumerators respectively

$$\begin{aligned} r(x,y) &= x^2+y^2, \\ h(x,y) &= x^8+14x^4y^4+y^8. \end{aligned}$$

Moreover, any polynomial of the form r^ih^j is a weight enumerator (of the direct sum of i copies of the repetition code and j copies of the extended Hamming code). It can be shown that, for fixed n, these polynomials (with $2i+8j=n$) are linearly independent; thus they span the nth homogeneous component of the algebra of invariants of G. This proves *Gleason's Theorem*:

Theorem 1.5 *A self-dual binary code has even length $n = 2m$, and its weight enumerator has the form*

$$\sum_{j=0}^{\lfloor n/8 \rfloor} a_j (x^2+y^2)^{(n-8j)/2} (x^8+14x^4y^4+y^8)^j$$

for some $a_j \in \mathbb{Q}$, $j \in \{0,\ldots,\lfloor n/8 \rfloor\}$.

The technique has other applications too. We give one of these. A self-orthogonal binary code has the property that all its weights are even. Such a code is called *doubly even* if all its weights are divisible by 4.

If C is a doubly even self-dual code, then the weight enumerator of C is invariant under the group $G^* = \langle A_1^*, A_2 \rangle$, where A_2 is as before and

$$A_1^* = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{pmatrix}.$$

It can be shown that G^* is a group of order 192, and the Hilbert series of its algebra of invariants is

$$\frac{1}{(1-t^8)(1-t^{24})}.$$

Now there exist doubly even self-dual codes which have lengths 8 and 24, namely the extended Hamming code and the extended Golay code. The weight enumerator of the extended Hamming code is given above. The weight enumerator of the extended Golay code is

$$g(x,y) = x^{24} + 759x^{16}y^8 + 2576x^{12}y^{12} + 759x^8y^{16} + y^{24}.$$

Again, these two polynomials are independent, and we have Gleason's second theorem:

Theorem 1.6 *A doubly even self-dual code has length n divisible by 8, say $n = 8m$, and its weight enumerator has the form*

$$\sum_{j=0}^{\lfloor n/24 \rfloor} a_j (x^8+14x^4y^4+y^8)^{(n-24j)/8} \times$$
$$\times (x^{24} + 759x^{16}y^8 + 2576x^{12}y^{12} + 759x^8y^{16} + y^{24})^j$$

for some $a_j \in \mathbb{Q}$, $j \in \{0,\ldots,\lfloor n/24 \rfloor\}$.

Several further results of the same sort are given in Sloane's survey [26].

The final topic in this section concerns the covering radius of a code. This is a parameter which is in a sense *dual* to the *packing radius*, the maximum number of errors which can be corrected.

Let C be a code of length n over an alphabet A. The *covering radius* of C is

$$\max_{v \in A^n} \min_{c \in C} d(v, c).$$

That is, it is the largest value of the distance from an arbitrary word to the nearest codeword. Said otherwise, it is the smallest integer r such that the spheres of radius r with centres at the codewords cover the whole of A^n.

We saw that, if the number of errors is at most the packing radius, then nearest-neighbour decoding correctly identifies the transmitted codeword. The covering radius has a similar interpretation: if the number of errors is greater than the covering radius, then nearest-neighbour decoding will certainly give the wrong codeword.

We give one result on the covering radius of binary codes which will be used in Section 5. We say that a code C has *strength* s if, given any s coordinate positions, all possible s-tuples over the alphabet occur the same number of times in these positions. The *maximum strength* is the largest integer s for which the code has strength s.

Theorem 1.7 *Let C be a code of length n over an alphabet A of size q and v an arbitrary word in A^n.*

(a) If C has strength 1, then the average distance of v from the words of C is $n(q-1)/q$.

(b) If C has strength 2, then the variance of the distances of v from the words of C is $n(q-1)/q^2$.

Proof (a) For $1 \leq i \leq n$, let $d_i(c) = 0$ if v and c agree in the ith coordinate, 1 otherwise. Then

$$d(v, c) = \sum_{i=1}^{n} d_i(c).$$

So the average distance from v to C is

$$\frac{1}{|C|} \sum_{i=1}^{n} \sum_{c \in C} d_i(c).$$

Now since C has strength 1, for any i we have $d_i(C) = 0$ for $|C|/q$ words $c \in C$, and $d_i(c) = 1$ for the remaining $(q-1)|C|/q$. So the inner sum is $(q-1)|C|/q$, and the result follows.

(b) Similarly, we have

$$d(v,c)(d(v,c)-1) = \sum_{i \neq j} d_i(c)d_j(c),$$

and if C has strength 2, then

$$\sum_{c \in C} d_i(c)d_j(c) = (q-1)^2|C|/q^2.$$

Thus, the average value of $d(v,c)(d(v,c)-1)$ is equal to $(q-1)^2 n(n-1)/q^2$. Now simple manipulation gives the result.

Theorem 1.8 *Let C be a linear binary code of length n containing the all-1 word.*

(a) The covering radius of C is at most $n/2$.

(b) If C has maximum strength at least 2, then its covering radius is at most $(n - \sqrt{n})/2$.

Proof (a) The hypothesis guarantees that C has strength at least 1. (For this, the all-1 word is not necessary; it is enough to assume that the support of C is $\{1, \ldots, n\}$.) Since the average distance from v to C is $n/2$, there is a word of C with distance at most $n/2$ from v.

(b) Suppose that the covering radius is $n/2 - s$. Since $d(v, c + \mathbf{1}) = n - d(v,c)$, all distances from v to C lie in the interval from $n/2 - s$ to $n/2 + s$. Since the variance of these distances is $n/4$, we must have $s \geq \sqrt{n}/2$.

Note that equality in (a) implies that $d(v,c) = n/2$ for all $c \in C$, while equality in (b) implies that $d(v,c) = (n \pm \sqrt{n})/2$ for all $c \in C$.

Exercise 1.1 (a) Show that, if C is a linear binary code of odd minimum weight d, then the minimum weight of $\overline{C}$ is $d+1$.

(b) Investigate how the dimension and minimum weight of codes change under the operations of puncturing, shortening and extending.

Exercise 1.2 If C_1 and C_2 are linear $[n_1, k_1, d_1]$ and $[n_2, k_2, d_2]$ codes over GF(q), prove that the direct sum $C_1 \oplus C_2$ is a linear $[n_1+n_2, k_1+k_2, \min\{d_1, d_2\}]$ code.

Exercise 1.3 (a) Let C be a binary or ternary Hamming code (that is, over GF(2) or GF(3)). Prove that $C \supseteq C^\perp$; that is, $\mathbb{C}^\perp$ is self-orthogonal.

(b) Let C be a Hamming code over GF(4). Prove that $C \supseteq C^\perp$ holds, if $C^\perp$ is calculated with respect to the Hermitian inner product.

Exercise 1.4 (a) Prove that, if all weights in a linear binary code C are divisible by 4, then C is self-orthogonal.

(b) Prove that, if a linear binary code C is self-orthogonal and is generated by a set of words whose weights are divisible by 4, then C is doubly even.

Exercise 1.5 Express the weight enumerator of the Golay code as a combination of r and h.

Exercise 1.6 Show that all doubly even self-dual codes of length 16 have weight enumerator h^2. Find two different examples of such codes.

Exercise 1.7 Can you find a more direct proof that a doubly-even self-dual binary linear code has length divisible by 8?

Exercise 1.8 Fill in the details of the proof of Theorem 1.7.

Exercise 1.9 What is the covering radius of the binary dual Hamming code of length 7?

Exercise 1.10 Let C be an (n, M) code over an alphabet of size q, with packing radius e and covering radius r. Prove that

$$\frac{q^n}{\sum_{i=0}^{e}\binom{n}{i}(q-1)^i} \le M \le \frac{q^n}{\sum_{i=0}^{r}\binom{n}{i}(q-1)^i}.$$

2 Symplectic and quadratic forms

In this section, we describe some of the properties of symplectic and quadratic forms over the field GF(2), and the geometries they define.

Let V be a vector space over a field F. A *quadratic form* on F is a function $Q : V \to F$ which satisfies the conditions

(a) $Q(\lambda v) = \lambda^2 Q(v)$ for all $\lambda \in F$, $v \in V$.

(b) The function $B : V \times V \to F$ defined by

$$Q(v + w) = Q(v) + Q(w) + B(v, w)$$

is *bilinear* (that is, linear in each variable).

We express (b) by saying that the form B is obtained from Q by *polarisation.*

The form B defined in (b) is *symmetric*, that is, $B(v, w) = B(w, v)$. Now, if the characteristic of F is not 2, then it follows from (a) and (b) that $Q(v) = \frac{1}{2}B(v, v)$ for all $v \in V$, so that Q can be recovered from B: that is, quadratic forms and symmetric bilinear forms carry the same information. Things are very different in characteristic 2, however. We are interested in this case, specifically $F = \mathrm{GF}(2)$.

From now on, we assume that $F = \mathrm{GF}(2)$.

Now we find that the form B is *alternating*, that is, $B(v, v) = 0$ for all $v \in V$. In general, an alternating bilinear form is *skew-symmetric*, that is, $B(v, w) = -B(w, v)$ for all $x, y \in V$. Of course, in characteristic 2, this just says that B is symmetric.

Clearly, Q cannot be recovered from B. Instead, we see that, if Q_1 and Q_2 both polarise to B, then $Q = Q_1 - Q_2$ polarises to the zero form, that is,

$$Q(v + w) = Q(v) + Q(w).$$

Also, because $\lambda^2 = \lambda$ for all $\lambda \in F$, we have

$$Q(\lambda v) = \lambda Q(v).$$

Thus, Q is linear. Conversely, two quadratic forms differing by a linear form polarise to the same bilinear form. So each alternating bilinear form corresponds to a coset of the dual space of V in the space of all quadratic forms.

A bilinear form B is said to be *non-degenerate* if it has the properties

(a) if $B(v, w) = 0$ for all $w \in V$ then $v = 0$;

(b) if $B(v, w) = 0$ for all $v \in V$ then $w = 0$.

If B is skew-symmetric (or symmetric), then each of these conditions implies the other, and we need only assume one. A non-degenerate alternating bilinear form on V exists if and only if V has even dimension. For any such form, there is a basis $\{v_1, \ldots, v_n, w_1, \ldots, w_n\}$ for V such that

$$\begin{aligned} B(v_i, v_j) &= 0 = B(w_i, w_j) \text{ for all } i, j, \\ B(v_i, w_i) &= 1 = -B(w_i, v_i) \text{ for all } i, \\ B(v_i, w_j) &= 0 = B(w_j, v_i) \text{ for } i \neq j. \end{aligned}$$

This is called a *symplectic basis*. A linear transformation of V which preserves the form B is called *symplectic*; the *symplectic group* is the group of all such transformations.

A quadratic form on an m-dimensional vector space is *non-singular* if it cannot be written as a form in fewer than m variables by any linear change of variables.

Equivalently, the only subspace W with the property that Q vanishes on W and $B(v, w) = 0$ for all $v \in V$ and $w \in W$ is the zero subspace. (Here B is the bilinear form obtained by polarising Q.) If the field has characteristic different from 2, then Q is non-singular if and only if B is non-degenerate; but this is not true over $\mathbb{Z}_2$, as we will see. In the case of an even-dimensional vector space over $\mathbb{Z}_2$, we will see that a quadratic form Q is non-singular if and only if the bilinear form obtained by polarisation is non-singular.

Given a subspace U of V, we set

$$U^\perp = \{x \in V : B(x, u) = 0 \text{ for all } u \in U\}.$$

The non-singularity of B guarantees that

$$\dim(U) + \dim(U^\perp) = \dim V,$$

but unlike the Euclidean case it is not true in general that $V = U \oplus U^\perp$, since we may have $U \cap U^\perp \neq \{0\}$. A subspace U of V is said to be *totally isotropic* if B vanishes identically on U, in other words, if $U \leq U^\perp$.

A vector x is said to be *singular* for the quadratic form Q if $Q(x) = 0$. A subspace U is *totally singular* if Q vanishes identically on U. By polarising the restriction of Q to U, we see that a totally singular subspace is totally isotropic; but the converse is not true. (Any 1-dimensional subspace is totally isotropic, but the span of a non-singular vector is not totally singular.)

Here is a small example. Take a 2-dimensional vector space over $\mathbb{Z}_2$, with typical vector (x_1, x_2). The four quadratic forms 0, x_1^2, x_2^2 and $x_1^2 + x_2^2 = (x_1 + x_2)^2$ are all singular, and are in fact equal to the four linear forms 0, x_1, x_2 and $x_1 + x_2$. The other four forms x_1x_2, $x_1x_2 + x_1^2 = x_1(x_1 + x_2)$, $x_1x_2 + x_2^2 = x_2(x_1 + x_2)$, and $x_1x_2 + x_1^2 + x_2^2$, are non-singular, and polarise to the bilinear form $x_1y_2 + x_2y_1$. The first three are equivalent under linear change of variable; each has value 0 at three of the four vectors and 1 at the fourth. The last form takes the value 1 at all three non-zero variables.

Let Q be a quadratic form on $V = \mathbb{Z}_2^n$.

A subspace W of V is *anisotropic* if, for all $w \in W$, we have $Q(w) = 0$ if and only if $w = 0$.

A *hyperbolic plane* is a subspace $U = \langle e, f \rangle$ with $Q(e) = Q(f) = 0$ and $B(e, f) = 1$ (So we have $Q(xe + yf) = xy$.)

Two quadratic forms Q_1 on V_1 and Q_2 on V_2 are *equivalent* if there is an invertible linear map $T : V_1 \to V_2$ such that $Q_2(vT) = Q_1(v)$ for all $v \in V_1$.

The next result gives the classification of non-singular quadratic forms.

Theorem 2.1 *(a) An anisotropic space has dimension at most* 2.

(b) Let Q be a quadratic form on V. Then

$$V = W \oplus U_1 \oplus \cdots \oplus U_r,$$

where W is anisotropic, $U_1, \ldots, U_r$ are hyperbolic planes, and the summands are pairwise orthogonal.

(c) If quadratic forms Q_1, Q_2 on V_1, V_2 give rise to decompositions

$$\begin{aligned} V_1 &= W_1 \oplus U_{11} \oplus \cdots \oplus U_{1r}, \\ V_2 &= W_2 \oplus U_{21} \oplus \cdots \oplus U_{2s}, \end{aligned}$$

as in (b), then Q_1 and Q_2 are equivalent if and only if $r = s$ and $\dim(W_1) = \dim(W_2)$.

As a result we see that quadratic forms over $\mathbb{Z}_2$ are determined up to equivalence by two invariants, the number r of hyperbolic planes (which is called the *Witt index*), and the dimension of the anisotropic part. We say that the form has *type* $+1$, 0 or -1 according as $\dim(W) = 0$, 1 or 2. Note that the bilinear form obtained by polarising Q is non-degenerate if and only if Q has non-zero type (that is, if and only if $\dim(V)$ is even).

Proof (a) If W is anisotropic, then the polarisation formula shows that $B(u, v) = 1$ for all distinct non-zero $u, v \in W$. If u, v, w were linearly independent, then

$$1 = B(u, v + w) = B(u, v) + B(u, w) = 0,$$

a contradiction. So $\dim(W) \leq 2$.

(b) The proof is by induction on $\dim(V)$, the case where $V = \{0\}$ being trivial. If V is anisotropic, there is nothing to prove. So we may suppose that there is a vector $u \in V$ with $u \neq 0$ and $Q(u) = 0$. Since Q is non-singular, there is a vector v with $B(u, v) = 1$. Then $Q(v) + Q(u + v) = 1$, and so either $Q(v) = 0$ or $Q(u + v) = 0$. Thus, $U_1 = \langle u, v\rangle$ is a hyperbolic plane. Moreover, $\dim(U_1^\perp) = \dim(V) - 2$, and it is easily checked that the restriction of Q to $U_1^\perp$ is non-singular. By the induction hypothesis, $U_1^\perp$ has a decomposition of the type specified, and we are done.

(c) It is clear that the condition given is sufficient for equivalence; we must show that it is necessary. It is also clear that equivalent quadratic forms are defined on spaces of the same dimension; so we must prove that they have the same Witt index. This follows immediately from the next lemma.

Lemma 2.2 *The Witt index of a quadratic form is equal to the maximum dimension of any totally singular subspace.*

Proof Let

$$V = W \oplus U_1 \oplus \cdots \oplus U_r,$$

where W is anisotropic, $U_1, \ldots, U_r$ are hyperbolic planes, and the summands are pairwise orthogonal. Let $U_i = \langle u_i, v_i\rangle$, where $Q(u_i) = Q(v_i) = 0$ and $B(u_i, v_i) = 1$. Then $X = \langle u_1, \ldots, u_r\rangle$ is totally singular and has dimension r.

We have to show that no larger totally singular subspace exists. This is proved by induction on r; it is true when $r = 0$ (since then V is anisotropic). So let X be a totally isotropic subspace, with $\dim(S) = s > 0$.

Choose a non-zero vector $x \in X$. As in the proof of the theorem, we can take x to lie in one of the hyperbolic planes, say U_1. Now Q induces a non-singular quadratic form $\overline{Q}$ on $\overline{V} = \langle x \rangle^\perp / \langle x \rangle$, and clearly this space has Witt index $r-1$; moreover, $X/\langle x \rangle$ is a totally singular subspace, with dimension $s-1$. By the inductive hypothesis, $s-1 \le r-1$, so $s \le r$.

Finally, if X is maximal totally singular in V, then $\overline{X}$ is maximal totally singular in $\overline{V}$; in this case, the inductive hypothesis shows that $s-1 = r-1$, so that $s = r$, as required.

¿From now on, we consider only non-singular quadratic forms on spaces of even dimension. A form of type $+1$ in $2n$ variables is equivalent to

$$x_1x_2 + x_3x_4 + \cdots + x_{2n-1}x_{2n},$$

while a form of type -1 is equivalent to

$$x_1x_2 + x_3x_4 + \cdots + x_{2n-1}^2 + x_{2n-1}x_{2n} + x_{2n}^2.$$

Theorem 2.3 *For $\epsilon = \pm 1$, let Q be a quadratic form of type ϵ on a vector space V of even dimension $2n$ over $\mathbb{Z}_2$. Then there are $2^{n-1}(2^n + \epsilon)$ vectors $v \in V$ such that $Q(v) = 0$.*

Proof The proof is by induction on n. We begin with $n = 1$. On a 2-dimensional space $\mathbb{Z}_2^2$, the quadratic form x_1x_2 has Witt index 1 (so type $+1$) and has three zeros $(0,0)$, $(1,0)$ and $(0,1)$. The form $x_1^2 + x_1x_2 + x_2^2$ has Witt index 0 (the space is anisotropic), so type -1, and vanishes only at the origin.

Now assume the result for $n-1$. Write $V = U \oplus V'$, where U is a hyperbolic plane and $\dim(V') = 2(n-1)$; the restriction of Q to V' has the same type as Q, say ϵ. So Q has $(2^{n-2}(2^{n-1} + \epsilon)$ zeros in V'. Since U and V' are orthogonal, we have $Q(u+w) = Q(u) + Q(w)$ for $u \in U$, $w \in V'$. Thus, $Q(u+w) = 0$ if and only if either $Q(u) = Q(w) = 0$ or $Q(u) = Q(w) = 1$. So there are

$$3 \cdot 2^{n-2}(2^{n-1} + \epsilon) + 2^{n-2}(2^{n-1} - \epsilon) = 2^{n-1}(2^n + \epsilon)$$

zeros, as required.

This gives an alternative proof of Theorem 2.1(c), since the two types of quadratic form on an even-dimensional vector space have different numbers of zeros, and cannot be equivalent.

Finally, we count the number of maximal totally isotropic or totally singular subspaces.

Theorem 2.4 *(a) Let B be a symplectic form on a vector space V of dimension $2n$ over $\mathbb{Z}_2$. Then the number of subspaces of V of dimension n which are totally isotropic with respect to B is*

$$\prod_{i=1}^{n}(2^i + 1).$$

(b) Let Q be a quadratic form of type $+1$ (that is, of Witt index n) on a vector space V of dimension $2n$ over $\mathbb{Z}_2$. Then the number of subspaces of V of dimension n which are totally singular with respect to Q is

$$\prod_{i=0}^{n-1}(2^i + 1).$$

Proof (a) The proof is by induction on n, the result being trivially true when $n = 0$. Suppose that it holds for spaces of dimension $2(n-1)$, and let V have dimension $2n$. For any non-zero vector $v \in V$, the space $v^\perp/\langle v\rangle$ has dimension $2(n-1)$ and carries a symplectic form. By the induction hypothesis, v lies in $N = \prod_{i=1}^{n-1}(2^i+1)$ totally isotropic n-spaces in V. Since there are $(2^n+1)(2^n-1)$ non-zero vectors, and each totally isotropic n-space contains (2^n-1) of them, double counting shows that the number of such spaces is $(2^n+1)N$, as required.

(b) The argument is similar. Assume the result for spaces of dimension $2(n-1)$, and let V have dimension $2n$. By Theorem 2.3, the number of non-zero singular vectors is $(2^{n-1}+1)(2^n-1)$, and each totally singular n-space contains $2^n - 1$ of them, so the induction works as in case (a).

Exercise 2.1 Let Q be a quadratic form in $2n$ variables with Witt index $n-1$. How many totally singular $(n-1)$-subspaces are there for Q?

3 Reed–Muller codes

This section gives a very brief account of Reed–Muller codes, which are very closely connected with affine geometry over $\mathbb{Z}_2$.

Let V be a vector space of dimension n over $\mathbb{Z}_2$. We identify V with $\mathbb{Z}_2^n$, and write a typical vector as $v = (x_1, \ldots, x_n)$. We regard the 2^n vectors in

V as being ordered in some way, say $v_1, v_2, \ldots, v_{2^n}$. Now nay binary word of length $N = 2^n$, say $(c_1, \ldots, c_n)$, can be thought of as a function f from V to $\mathbb{Z}_2$ (where $f(v_i) = c_i$ for $i = 1, \ldots, N$).

Lemma 3.1 *Any function from V to $\mathbb{Z}_2$ can be represented as a polynomial in the coordinates $(x_1, \ldots, x_n)$, in which no term contains a power of x_i higher than the first, for any i.*

Proof It is enough to show this for the function $f = \delta_a$ given by

$$\delta_a(v) = \begin{cases} 1 & \text{if } v = a, \\ 0 & \text{otherwise,} \end{cases}$$

for $a \in V$, since any function is a sum of functions of this form (specifically,

$$f = \sum_{a \in V} f(a)\delta_a.$$

But, if $a = (a_1, \ldots, a_n)$, then we have

$$\delta_a(v) = \prod_{i=1}^{n} (x_i - a_i - 1),$$

where $v = (x_1, \ldots, x_n)$.

Corollary 3.2 *The monomial functions f_I on V are linearly independent for $I \subseteq \{1, \ldots, n\}$, where*

$$f_I(v) = \prod_{i \in I} x_i$$

for $v = (x_1, \ldots, x_n)$.

Proof These 2^n functions span the 2^n-dimensional space of all functions from V to $\mathbb{Z}_2$.

For $0 \leq r \leq n$, the rth order *Reed–Muller code* of length $N = 2^n$ is spanned by the set of polynomial functions of degree at most r on $V = \mathbb{Z}_2^n$. It is denoted by $\mathcal{R}(n, r)$.

The next result summarises the properties of Reed–Muller codes.

Theorem 3.3 *(a) $\mathcal{R}(n,r)$ is a*

$$\left[N = 2^n, k = \sum_{i=0}^{r} \binom{n}{i}, d = 2^{n-r}\right]$$

code.

(b) $\mathcal{R}(n,r)^\perp = \mathcal{R}(n, n-r-1)$.

Proof (i) The code has length $N = 2^n$ by definition, and dimension $k = \sum_{i=0}^{r} \binom{n}{i}$ since this is the number of monomials of degree at most r.

(ii) Next we prove part (b). Since

$$\dim(\mathcal{R}(n,r)) + \dim(\mathcal{R}(n, n-r-1)) = 2^n,$$

it is enough to prove that these codes are orthogonal, and hence enough to prove it for their spanning sets. Now, if f and f' are monomials of degrees at most r, $n-r-1$ respectively, then there is a variable (say x_n) occurring in neither of them, and so the values of f and f' are unaffected by changing x_n from 0 to 1. Thus, the intersection of the supports of f and f' has even cardinality, and so $f \cdot f' = 0$.

(iii) Finally, we establish that $\mathcal{R}(n,r)$ has minimum weight 2^{n-r} by induction on r. This is true for $r = 0$ since $\mathcal{R}(n,0)$ consists of the all-0 and all-1 words only. So assume the result for $r-1$.

Take $f \in \mathcal{R}(n,r)$: we must show that the support of f has size at least 2^{n-r}. By the induction hypothesis, we may assume that $f \notin \mathcal{R}(n, r-1)$. By (b), there is a monomial of degree $n-r$ not orthogonal to f; we may suppose that it is $x_1 \cdots x_{n-r}$. Thus, if S is the support of f, and

$$A = \{(x_1, \ldots, x_n) \in V : x_1 = \cdots = x_{n-r} = 1\},$$

then $|S \cap A|$ is odd. Now A is an affine flat in V of dimension r. So the union of any two translates of A is an affine flat of dimension $r+1$, and supports a word in $\mathcal{R}(n, n-r-1) = \mathcal{R}(n,r)^\perp$; so $|S \cap (A \cup (A+v))|$ is even for all $v \notin A$. Thus, $|S \cap (A+v)|$ is odd for all $v \in V$. In particular, S meets all 2^{n-r} distinct translates of A, so $|S| \geq 2^{n-r}$, and we are done.

Corollary 3.4 *The code $\mathcal{R}(n, n-2)$ is equivalent to the extended Hamming code $\overline{\mathcal{H}(n,2)}$ of length 2^n.*

Proof If we puncture this code in one position, we obtain a $[2^n - 1, 2^n - n - 1, 3]$ linear code. This code is equivalent to a Hamming code: for its minimum weight is 3, so the columns of its parity check matrix are pairwise linearly independent; and the number of columns is $2^n - 1$, so every non-zero n-tuple occurs once.

The weight enumerator of $\mathcal{R}(n, 1)$ is

$$x^{2^n} + (2^{n+1} - 2)x^{2^{n-1}}y^{2^{n-1}} + y^{2^n}.$$

For this code contains the all-0 and all-1 words, and also the linear functions and their complements (each of which have weight 2^{n-1}). This code is equivalent to the dual extended Hamming code.

Note that, if we shorten this code, we obtain the dual Hamming code, which (as we have seen) is a constant-weight code.

We will be particularly interested in the *second-order Reed–Muller code* $\mathcal{R}(n, 2)$. Since $x^2 = x$ for all $x \in \mathbb{Z}_2$, every linear function on V is quadratic, and we have the following description:

$$\mathcal{R}(n, 2) = \{Q + c : Q \text{ a quadratic form on } V,\ c \in \mathbb{Z}_2\}.$$

Recall that two quadratic forms polarise to the same bilinear form if and only if they differ by a linear form. This means that the cosets of $\mathcal{R}(n, 1)$ in $\mathcal{R}(n, 2)$ are in one-to-one correspondence with the alternating bilinear forms on V.

The weight enumerators of these codes are known. They are calculated by the following series of steps:

- Choose $m \leq (n/2)$. Count the number of subspaces W of V of codimension $2m$.

- Count the number of *symplectic forms* (non-degenerate alternating bilinear forms) on the $2m$-dimensional space V/W. Each such form extends uniquely to an alternating form on V with radical W. Let B be such a form.

- There are 2^{2m} quadratic forms on V which polarise to V and are zero on W. Such a form has weight $2^{n-1} + \epsilon 2^{n-m-1}$. Adding the all-1 word, we obtain a word of weight $2^{n-1} - \epsilon 2^{n-m-1}$. So we obtain 2^{2m} words of each such weight.

- Any other quadratic form which polarises to B induces a non-zero linear form on W. Such a form has weight 2^{n-1}. We obtain $2^{n+1}-2^{2m+1}$ forms of weight 2^{n-1}.
- Add the contributions from the last two steps, multiply by the factors coming from the first two steps, and sum over m, to find the weight enumerator of $\mathcal{R}(n,2)$

Exercise 3.1 Show that the code obtained by shortening the extended Golay code on the eight positions of an octad is equivalent to $\mathcal{R}(4,1)$, while the code obtained by puncturing on these positions is equivalent to $\mathcal{R}(4,2)$.

Exercise 3.2 Calculate the weight enumerator of $\mathcal{R}(5,2)$

(a) using the method outlined above;

(b) using Theorem 3.3 and Gleason's Theorem (Theorem 1.6).

Exercise 3.3 Show that a coset of $\mathcal{R}(n,1)$ in $\mathcal{R}(n,2)$ contains words of at most three different weights, and that only two weights occur if and only if the bilinear form indexing the coset is non-degenerate.

(Such a coset is called a *two-weight coset.*)

Exercise 3.4 Prove that the blocks of the design $\mathcal{D}(C)$ formed by the words of minimum weight in the second-order Reed–Muller code $C=\mathcal{R}(2,n)$ are the $(n-2)$-dimensional affine flats in $\mathrm{AG}(n,2)$. Deduce that $\mathcal{D}$ is a 3-design.

4 Self-dual codes

We now apply these results to the problem of counting self-dual and doubly even self-dual binary codes.

A binary self-dual code C of length n has the property that all its words have even weight. This means that the all-1 word $\mathbf{1}$ is orthogonal to every word in C, that is, $C \subseteq \langle \mathbf{1} \rangle^{\perp}$. Since C is self-dual, $\mathbf{1} \in C$.

Now let $W = \langle \mathbf{1} \rangle^{\perp}$, the even-weight subcode of $\mathrm{GF}(2)^n$. Then $x \cdot x = 0$ for all $x \in W$, so the dot product is an alternating bilinear form on W. It is not non-degenerate, since $\mathbf{1}$ lies in its radical; but it induces a non-degenerate bilinear form B on the $(n-2)$-dimensional space $V = W/\langle \mathbf{1} \rangle$. Now a code C containing $\mathbf{1}$ is self-orthogonal if and only if $\overline{C} = C/\langle \mathbf{1} \rangle$ is totally isotropic for B; so C is self-dual if and only if $\overline{C}$ is maximal totally isotropic. Thus, from Theorem 2.4, we have:

Theorem 4.1 *The number of binary self-dual codes of length $n = 2m$ is*

$$\prod_{i=1}^{m-1} (2^i + 1).$$

¿From this theorem, the numbers of binary self-dual codes of length 2, 4, 6, 8 is 1, 3, 15, 135 respectively.

For $n < 8$, the only self-dual codes are direct sums of copies of the repetition code of length 2. The number of codes of this form is equal to the number of partitions of the set of coordinates into subsets of size 2, which is 1, 3, 15, 105 for $n = 2, 4, 6, 8$. So for $n = 8$, there are 30 further codes, which as we shall see are all equivalent to the extended Hamming code of length 8.

For any two binary words x and y, we have

$$\mathrm{wt}(x + y) = \mathrm{wt}(x) + \mathrm{wt}(y) - 2\,\mathrm{wt}(x \cap y).$$

Now $\mathrm{wt}(x \cap y) \equiv (x \cdot y) \bmod 2$. So, if x has even weight, and n is divisible by 4, then we can set $Q(x) = \frac{1}{2}\,\mathrm{wt}(x) \bmod 2$; we have $Q(x) = Q(\mathbf{1} + x)$, so Q is well-defined on V, and we have

$$Q(x + y) = Q(x) + Q(y) + B(x, y).$$

In other words, Q is a quadratic form on V which polarises to B. Furthermore, a code C is doubly even if and only if $\overline{C}$ is totally singular (with respect to Q), and C is doubly even self-dual if and only if $\overline{C}$ is maximal totally singular of dimension $n - 1$.

Thus, from Theorem 2.4(b), we have:

Theorem 4.2 *The number of doubly-even self-dual codes of length $n = 2m$ divisible by 8 is*

$$\prod_{i=0}^{m-2} (2^i + 1).$$

This shows that there are indeed 30 doubly-even self-dual codes of length 8 (all equivalent to the extended Hamming code).

Exercise 4.1 Verify that tthere are 30 codes of length 8 equivalent to the extended Hamming code by showing that the automorphism group of the code has index 30 in the symmetric group S_8.

Exercise 4.2 Count the number of binary words of length n with weight divisible by 4. [Hint: Let a and b be the numbers of words which have weight congruent to 0 or 2 mod 4 respectively. Then $a + b = 2^{n-1}$. Calculate $a - b$ by evaluating the real part of $(1 + \mathrm{i})^n$.]

Hence show that the quadratic form q defined earlier has Witt index $n-1$ if $n \equiv 0 \bmod 8$, and $n - 2$ if $n \equiv 4 \bmod 8$.

This gives an alternative proof that doubly even self-dual codes must have length divisible by 8.

Exercise 4.3 Classify doubly-even self-dual codes of length 16. Use Theorem 4.2 to show that your classification is complete.

5 Bent functions

Let n be even, say $n = 2m$. The code $\mathcal{R}(n, 1)$ has strength 3 (since, by Theorem 3.3, its dual has minimum weight 4). By Theorem 1.8, its covering radius is at most $2^{2m-1} - 2^{m-1}$. This bound is attained. For, if Q is a non-singular quadratic form, then the distances from Q to words of $\mathcal{R}(n, 1)$ are equal to the weights of words in the coset $\mathcal{R}(n, 1) + Q$, and we have seen that these weights are $2^{2m-1} \pm 2^{m-1}$.

Let $n = 2m$, and let $V = \mathbb{Z}_2^n$. A function $f : V \to \mathbb{Z}_2$ is called a *bent function* if its minimum distance from $\mathcal{R}(n, 1)$ is $2^{2m-1} - 2^{m-1}$.

As the name (coined by Rothaus [23]) suggests, a bent function is a function which is at the greatest possible distance from the linear functions.

As we observed, a non-singular quadratic form is a bent function. In fact, a quadratic form is a bent function if and only if it is non-singular; and there are just two such functions up to equivalence.

Bent functions of higher degree exist: see the Exercises. The problem of classifying bent functions appears to be hopeless. Various authors have attacked this problem for reasonably small numbers of variables; see [19, 2].

Bent functions have a range of applications, both theoretical and practical. Here is one example. Let f be a bent function. Then $\mathcal{R}(n, 1) + f$ is a two-weight coset of $\mathcal{R}(n, 1)$. (This follows from the remark following the proof of Theorem 1.8(b): the weights are $2^{2m-1} \pm 2^{m-1}$.) Conversely, any two-weight coset consists of bent functions.

Theorem 5.1 *(a) Let $\mathcal{B}$ be the set of supports of all words which have weight $2^{2m-1} - 2^{m-1}$ in a two-weight coset of $\mathcal{R}(n,1)$. Then the structure $(V, \mathcal{B})$ is a 2-$(2^{2m}, 2^{2m-1} - 2^{m-1}, 2^{2m-2} - 2^{m-1})$ design.*

(b) A design with the parameters given in (a) arises from a two-weight coset of $\mathcal{R}(n,1)$ if and only if it has the following property: the symmetric difference of any three blocks of the design is either a block or the complement of a block.

Part (a) of this theorem appears in a number of places. Part (b) is due to Kantor [16], who calls his condition the *symmetric difference property.*

See also [3] for another application of bent functions.

Exercise 5.1 Show that the function

$$x_1x_2 + x_3x_4 + \cdots + x_{2m-1}x_{2m} + x_1x_3 \cdots x_{2m-1}$$

is a bent function.

6 Kerdock codes

We have seen that, if Q is a quadratic form which polarises to a bilinear form of rank $2m$, then the weight of Q is $2^{n-1} \pm 2^{n-m-1}$ or 2^{n-1}; in particular, it is at least $2^{n-1} - 2^{n-m-1}$.

Let $\mathcal{B}$ be a set of alternating bilinear forms on V. Let $K(\mathcal{B})$ denote the set

$$\{Q + c : Q^\pi \in \mathcal{B}, c \in \mathbb{Z}_2\}$$

of functions on V, where Q^π is the bilinear form obtained by polarising Q. Then $K(\mathcal{B})$ is a

$$(N = 2^n, M = 2^{n+1}|\mathcal{B}|, d = 2^{n-1} - 2^{n-m-1})$$

code, where $2m$ is the minimum rank of the difference of two forms in $\mathcal{B}$. In particular, to make d as large as possible, we should require that n is even and the difference of any two forms in $\mathcal{B}$ is non-degenerate. (We call such a set $\mathcal{B}$ a *non-degenerate set.*) Furthermore, the code $K(\mathcal{B})$ is linear if and only if $\mathcal{B}$ is closed under addition.

Lemma 6.1 *A non-degenerate set of alternating bilinear forms on a $2m$-dimensional vector space has cardinality at most 2^{2m-1}.*

Proof Each form can be represented by a skew-symmetric matrix with zero diagonal: if $\{e_1, \ldots, e_{2m}\}$ is a basis for V, the (i,j) entry of the matrix representing B is $B(e_i, e_j)$. Now, if $B - B'$ is non-degenerate, then the first rows of the matrices representing B and B' are unequal. Since there are at most 2^{2m-1} possible first rows (remember that the diagonal entry is zero), there are at most 2^{2m-1} forms in a non-degenerate set.

A *Kerdock set* is a non-degenerate set of bilinear forms on a $2m$-dimensional vector space V over $\mathbb{Z}_2$, having cardinality 2^{2m-1}, that is, attaining the upper bound. A *Kerdock code* is a code of the form $K(\mathcal{B})$, where $\mathcal{B}$ is a Kerdock set. Thus, it is a $(2^{2m}, 2^{4m}, 2^{2m-1} - 2^{m-1})$ code.

It can be shown that, for $m > 1$, a Kerdock code must be non-linear. (The largest additively closed non-singular set has cardinality 2^m; we will construct it in the next section. In the case $m = 2$, the unique example of a Kerdock code is the *Nordstrom–Robinson code*, a $(16, 256, 6)$ code. The first construction for all m was given by Kerdock [17]. A simplified construction by Dillon, Dye and Kantor is presented in [8], Chapter 12.

Although Kerdock codes are non-linear, they have recently been "linearised" in a remarkable way by Hammons *et al.* [15]. This is the subject of the last section.

Since non-quadratic bent functions exist, it is natural to ask whether 'Kerdock sets' of higher degree can exist too. So far, no examples have been found.

Exercise 6.1 Let $O = \{1, 2, \ldots, 8\}$ be an octad in the extended Golay code $\mathcal{G}_{24}$. Consider the set of words of $\mathcal{G}_{24}$ whose supports intersect O in one of the following eight sets:

$$\emptyset, \{1,2\}, \{1,3\}, \ldots, \{1,8\}.$$

Now restrict these words to the complement of O. Show that the result is a $(16, 256, 6)$ code, and identify it with a Kerdock code.

Show that, if we use instead the set

$$\emptyset, \{1,2\}, \{1,3\}, \{2,3\},$$

we obtain a linear $[16, 7, 6]$ code.

7 Some resolved designs

Some infinite families of systems of linked symmetric BIBDs (or SLSDs, for short) were constructed by Cameron and Seidel [9]. The smallest of these systems was used by Preece and Cameron [21] to construct certain resolvable designs (which they called *fully-balanced hyper-graeco-latin Youden 'squares'*). For example, they gave a 6×16 rectangle, in which each cell contains one letter from each of three alphabets of size 16, satisfying a number of conditions, including:

- No letter occurs more than once in each row or column of the rectangle.
- The sets of letters from each of the three alphabets in the columns of the rectangle form a 2-$(16, 6, 2)$ design.
- Each pair of alphabets carry a 2-$(16, 6, 2)$ design, where two letters are incident if they occur together in a cell of the rectangle.
- The number of columns containing a given pair of letters from distinct alphabets is 1 if the two letters are incident, 3 otherwise.

In this section we construct an infinite sequence of such designs.

A symmetric balanced incomplete-block design (SBIBD) can, like any incidence structure, be represented by a graph (its *incidence graph* or *Levi graph*). The vertex set of the graph Γ is the disjoint union of two sets X_1 and X_2, and each edge has one end in X_1 and the other in X_2. If the design is a 2-(v, k, λ) design, the graph has the properties

- $|X_1| = |X_2| = v$;
- for $\{i, j\} = \{1, 2\}$, any point in X_i has exactly k neighbours in X_j;
- for $\{i, j\} = \{1, 2\}$, any two points in X_i have exactly λ neighbours in X_j.

¿From such a design, we obtain a resolved design with $r = 2$ classes of blocks as follows: the treatments are the $t = vk$ edges of Γ; for $i = 1, 2$, the blocks in the ith class consist of the sets of edges on each of the vertices of X_i (so that there are v blocks of k in each class).

Any regular bipartite graph has a 1-*factorisation*, a partition of the edge set into k classes of v edges each, where the edges of each class partition the vertices. (This follows from Hall's Marriage Theorem.) This partition of the edge set (treatment set) is orthogonal to the two block partitions. Using it, we can represent the design by a Latin rectangle as follows. Number the elements of X_i from 1 to v for $i = 1, 2$, and number the 1-factors from 1 to k; then the (i, j) entry in the $k \times v$ rectangle is the number of the vertex in X_2 joined to the vertex j of X_1 by an edge of the 1-factor numbered i.

In the case of a SBIBD arising from a difference set in a group A, we have an action of A on the graph Γ so that the orbits are X_1 and X_2 and the action on each orbit is regular. In this case, A permutes the edges in k orbits each of size v, forming the desired 1-factorisation.

A *system of linked SBIBDs*, or SLSD for short, can be represented by a multipartite graph Γ with r classes $X_1, \ldots, X_r$, satisfying the conditions

- for any distinct i, j, the induced subgraph on $X_i \cup X_j$ is the incidence graph of a SBIBD (with parts X_i and X_j), having parameters 2-(v, k, λ) independent of i and j;
- there exist integers x and y such that, for any distinct i, j, k, and any vertices $p_i \in X_i$ and $p_j \in X_j$, the number of common neighbours of p_i and p_j in X_k is equal to x if p_i and p_j are adjacent, and to y otherwise.

We cannot construct a resolved design from a SLSD unless an extra condition holds. A *full clique* in a SLSD is a set of vertices, containing one from each of the sets X_i, whose vertices are pairwise adjacent. (So a full clique contains r vertices.) A *full clique cover* is a set of full cliques with the property that every edge is contained in exactly one full clique in the set. (So the number of full cliques in a full clique cover is vk.) Now if we have a full clique cover of a SLSD, we construct a design as follows: the treatments are the $t = vk$ full cliques; for $i = 1, \ldots, r$, the blocks in the ith class are the sets of full cliques in the cover containing each of the vertices in X_i. Each of the r block classes contains v blocks of size k.

A 1-*factor* is a set of v full cliques covering all vertices just once; a 1-*factorisation* is a partition of the full cliques into 1-factors. (Thus, it is a partition of the treatments into k sets of v, which is orthogonal to each block partition.) I do not know whether 1-factorisations always exist. However, if there is a group A of automorphisms whose orbits are $X_1, \ldots, X_r$ and which acts regularly on each orbit, then the orbits of A on full cliques form a 1-factorisation.

If we have a 1-factorisation, then we can represent the design by a $k \times n$ rectangle whose entries are $(r-1)$-tuples, similarly to before. We number the elements of each set X_i from 1 to n, and the 1-factor from 1 to k; then the (i,j) entry of the rectangle is the $(r-1)$-tuple $(l_2, \ldots, l_r)$, where l_i is the point of X_i lying in a full clique of the ith 1-factor with the jth point of X_1. This is the representation used in [21].

The construction of the designs is based on properties of bilinear and quadratic forms over $F = \mathrm{GF}(2)$. Let B be any alternating bilinear form on a $2n$-dimensional vector space over F. The set $\mathcal{Q}(B)$ of quadratic forms which polarise to B has 2^{2n} members. If Q is one member of this set, then all others can be obtained by adding linear forms to Q. Suppose that B is non-degenerate. Then any linear form can be written as $L(x) = B(v,x)$ for some vector $v \in V$. So

$$\mathcal{Q}(B) = \{Q(x) + B(v,x) : v \in V\} = \{Q(x+v) + Q(v) : v \in V\}.$$

Let $X = \{x \in V : Q(x) = 0\}$ be the set of zeros of Q. Then the set of zeros of $Q(x) + B(v,x)$ is obtained by translating X by v, and complementing this set in V if $Q(v) = 1$. So any quadratic form in $\mathcal{Q}(B)$ has either N or $2^{2n} - N$ zeros, for some N. We can take $N = 2^{2n-1} + \epsilon 2^{n-1}$, where $\epsilon = \pm 1$; the form Q has type ϵ if it has $2^{2n-1} + \epsilon 2^{n-1}$ zeros (Theorem 2.3).

Now the set X of zeros of Q is a difference set in the additive group of the vector space V, and so gives rise to a symmetric BIBD, whose points are the vectors in V and whose blocks are the translates of X; as we have seen, these are the zero sets of the quadratic forms in $\mathcal{Q}(B)$, complemented in the case of forms of type opposite to that of B.

This design has a more symmetrical description, as follows. (The proof that this is the same is an exercise, or is given in [9].) Let B_1 and B_2 be two alternating bilinear forms on V, whose difference $B_1 - B_2$ is non-degenerate. Then the points and blocks of the SBIBD are the sets $\mathcal{Q}(B_1)$ and $\mathcal{Q}(B_2)$ respectively; a point Q_1 and block Q_2 are incident in the design D_ϵ if and only if the form $Q_1 - Q_2$ (which is non-singular) has type ϵ.

The design D_ϵ has $v = 2^{2n}$, $k = 2^{2n-1} + \epsilon 2^{n-1}$ and $\lambda = 2^{2n-2} + \epsilon 2^{n-1}$.

Let V be a vector space of dimension $2n$ over the field $F = \mathrm{GF}(2)$. Given a non-degenerate set $\mathcal{B}$ of alternating bilinear forms and a value $\epsilon = \pm 1$, we define a SLSD $S_\epsilon(\mathcal{B})$ as follows: the elements are the quadratic forms in the sets $\mathcal{Q}(B)$ for $B \in \mathcal{B}$; forms $Q_i \in \mathcal{Q}(B_i)$ and $Q_j \in \mathcal{Q}(B_j)$ are incident if $Q_i - Q_j$ has type ϵ. It follows from the description of the designs that the first condition in the definition of a SLSD is satisfied; see [9] for a proof that the second condition holds too.

The largest non-degenerate sets are the Kerdock sets; but these do not have full clique covers in general. However, there is a construction which produces sets of cardinality 2^n; it is these which we use.

Let $K = \mathrm{GF}(2^n)$. There is a F-linear map from K onto F, the *trace map*, given by

$$\mathrm{Tr}(x) = x + x^2 + x^{2^2} + \cdots + x^{2^{n-1}}.$$

(Note that $x^{2^n} = x$ for all $x \in K$.)

Let V be a 2-dimensional vector space over K. By restricting scalars from K to F, V becomes a $2n$-dimensional vector space over F. If b is an alternating bilinear form on V as K-space, then $B = \mathrm{Tr}(b)$ is an alternating bilinear form on V as F-space; and B is non-degenerate if and only if b is. Similarly, the traces of the quadratic forms (on the K-space V) polarising to b are precisely the quadratic forms (on the F-space V) polarising to B.

Now take b to be any non-degenerate alternating bilinear form on the K-space V (for example, take $b((x_1, x_2), (y_1, y_2)) = x_1y_2 - x_2y_1$). Then αb is also a non-degenerate alternating bilinear form, for any non-zero $\alpha \in K$. We have

$$\mathrm{Tr}(\alpha_1 b) - \mathrm{Tr}(\alpha_2 b) = \mathrm{Tr}((\alpha_1 - \alpha_2)b)$$

for $\alpha_1 \neq \alpha_2$. So the 2^n forms

$$\{\mathrm{Tr}(\alpha b) : \alpha \in K\}$$

comprise a non-degenerate set of cardinality 2^n, and so give rise to a SLSD with $r = 2^n$.

We must now produce the full clique cover and its 1-factorisation. The argument uses a little group theory.

The explicit form of b given in the last section is the determinant of the matrix $\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$. It follows from this that the *special linear group* $\mathrm{SL}(2, 2^n)$ of 2×2 matrices of determinant 1 over K preserves b, and hence each of the forms $\mathrm{Tr}(\alpha b)$. Thus the product $A \cdot \mathrm{SL}(2, 2^n)$, where A is the additive group of V, acts on the SLSD, fixing each of the sets $X_1, \ldots, X_r$. (In fact it acts doubly transitively on each X_i.)

The subgroup fixing a point p_i of X_i is a complement to A in this product, and so is isomorphic to $\mathrm{SL}(2, 2^n)$; it is transitive on the remaining points of

X_i and has two orbits on X_j for all $j \neq i$, namely, the points incident and non-incident to p_i. If $p_j \in X_j$ is a point incident with p_i, then the stabiliser of p_i and p_j is a dihedral group of order $2(2^n - \epsilon)$. Now all such dihedral groups in our group $A \cdot \mathrm{SL}(2, 2^n)$ are conjugate (they are the normalisers of Sylow p-subgroups, where p is a prime divisor of $2^n - \epsilon$); so this subgroup fixes one point p_l in each set X_l. Moreover, these points p_l are pairwise incident. For, if p_l and p_m were not incident, their stabiliser would be a dihedral group of order $2(2^n + \epsilon)$; but this number does not divide $2(2^n - \epsilon)$.

Now the set of all these points p_l is a full clique. It is the unique full clique containing p_i and p_j which is stabilised by a dihedral group of order $2(2^n - \epsilon)$. So we have constructed a full clique cover.

Now the orbits of the group A on these full cliques form the required 1-factorisation, as we described earlier.

Exercise 7.1 Complete the proof that a non-degenerate set of alternating bilinear forms gives rise to a SLSD.

8 Extraspecial 2-groups

An *extraspecial 2-group* is a 2-group whose centre, derived group, and Frattini subgroup all coincide and have order 2. Such a group E has order 2^{2n+1} for some n. If $\zeta(E)$ is the centre, then we can identify $\zeta(E)$ with the additive group of $F = \mathbb{Z}_2$. Since squaring (the map $e \mapsto e^2$) is a function from E to $\zeta(E)$, the factor group $\overline{E}$ is elementary abelian of order 2^{2n}, and can be identified with the additive group of a $2n$-dimensional vector space over F.

Now the structure of the group can be defined in terms of the vector space. Commutation (the map $(e, f) \mapsto [e, f] = e^{-1}f^{-1}ef$) is a function from $E \times E$ to $\zeta(E)$. Observing that $[ez, f] = [e, fz] = [e, f]$ for $z \in \zeta(E)$, we see that it induces a map from $\overline{E} \times \overline{E}$ to F. It is readily checked that this map is a non-singular alternating bilinear form B. (This explains why $|\overline{E}|$ is an even power of 2.) We also have that $(ez)^2 = e^2$ for $z \in \zeta(E)$, so squaring induces a quadratic form $Q : \overline{E} \to F$, which polarises to B.

The vector space $\overline{E}$, with the bilinear form B and quadratic form Q, determine the structure of the extraspecial group E. From the classification of quadratic forms, we conclude that there are just two extraspecial groups of order 2^{2n+1} for any n (up to isomorphism).

A subgroup S of E is normal in E if and only if it contains $\zeta(E)$. For such a subgroup, $\overline{S} = S/\zeta(E)$ is a subspace of $\overline{E}$. If we start with $\overline{S} < \overline{E}$ we could have the corresponding $S < E$ containing $\zeta(E)$ or not, but normality of S does not matter in what follows. The following are immediate:

(a) S is abelian if and only if $\overline{S}$ is totally isotropic;

(b) S is elementary abelian if and only if $\overline{S}$ is totally singular.

Consider the case $n = 1$. The quadratic form x_1x_2 corresponds to a group generated by two elements of order 2 with product of order 4; this is the *dihedral group* D_8. The form $x_1x_2 + x_1^2 + x_2^2$ corresponds to a group in which all six non-central elements have order 4; this is the *quaternion group* Q_8. The singular forms correspond to groups which, while having a central subgroup of order 2 with elementary abelian quotient, are not extraspecial; x_1^2 corresponds to $C_2 \times C_4$, and 0 to $C_2 \times C_2 \times C_2$.

Two extraspecial 2-groups are isomorphic if and only if the corresponding quadratic forms are equivalent. So our classification of quadratic forms (Theorem 2.1) gives the following result:

Theorem 8.1 *For each $m \geq 1$, there are (up to isomorphism) just two extraspecial 2-groups of order 2^{2m+1}.*

The groups are determined by the quadratic forms. They can also be described in a more group-theoretic manner as follows.

Let G_1, G_2 be groups, Z_1, Z_2 subgroups of $\zeta(G_1)$ and $\zeta(G_2)$ respectively, and $\theta : Z_1 \to Z_2$ an isomorphism. The *central product* $G_1 \circ G_2$ of G_1 and G_2 with respect to θ is obtained from the direct product $G_1 \times G_2$ by identifying each element $z \in Z_1$ with its image $z\theta \in Z_2$; in other words, it is the group

$$G_1 \circ G_2 = (G_1 \times G_2)/N,$$

where $N = \{(z^{-1}, z\theta) : z \in Z_1\}$.

Now if the quadratic forms Q_1 and Q_2 on V_1 and V_2 give rise to groups E_1 and E_2 as above, and we take θ to be the unique isomorphism from $\zeta(E_1)$ to $\zeta(E_2)$, then the group associated with the form $Q_1 + Q_2$ on $V_1 \oplus V_2$ is the central product $E_1 \circ E_2$.

Hence the two extraspecial groups of order 2^{2m+1} can be written as

$$\begin{array}{ll} D_8 \circ D_8 \circ \cdots \circ D_8 \circ D_8 & (m \text{ factors}) \text{ and} \\ D_8 \circ D_8 \circ \cdots \circ D_8 \circ Q_8 & (m \text{ factors}). \end{array}$$

Exercise 8.1 Let Q be a non-singular quadratic form on a space of odd dimension V over $\mathbb{Z}_2$. Show that Q vanishes on half of the vectors in V.

Exercise 8.2 Let Q be a (possibly singular) quadratic form on a vector space V of dimension $2n$ over $\mathbb{Z}_2$, which polarises to B. The *radical* of B is defined to be the set

$$\{v \in V : B(v,w) = 0 \text{ for all } w \in V\}.$$

(a) Show that the radical has even dimension $2d$.

(b) Show that the number of zeros of Q is

$$2^{2n-1} + \epsilon 2^{n+d-1},$$

for some $\epsilon \in \{+1, 0, -1\}$.

Exercise 8.3 Prove that the quadratic forms

$$x_1x_2 + x_3x_4$$

and

$$x_1^2 + x_1x_2 + x_2^2 + x_3^2 + x_3x_4 + x_4^2$$

are equivalent.

Deduce that $D_8 \circ D_8 \cong Q_8 \circ Q_8$.

9 Quantum computing

We give a brief review of the concept of public-key cryptography, the RSA system (its most popular realisation), and the relevance a fast quantum algorithm for factorising large integers would have for this system. No attempts at rigorous definitions of complexity classes or proofs of the assertions will be given.

In any cryptosystem, the *plaintext* to be transmitted is encrypted by some algorithm depending on additional data called the *key*, to produce *ciphertext*. The recipient uses another algorithm to recover the plaintext from the ciphertext and key. The simplest example is the *one-time pad*, the only provably secure cipher system. The plaintext is first encoded as a string of bits of length n, say $a_1a_2\ldots a_n$. The key consists of a string $b_1b_2\ldots b_n$ of n random bits, produced by some physical randomising process such as tossing coins. The encryption algorithm is bitwise addition; so the ciphertext is $c_1c_2\ldots c_n$, where $c_i = a_i + b_i$ (addition mod 2). The decryption algorithm in this case is identical to the encryption algorithm: add the key bitwise (since $c_i + b_i = a_i$). The ciphertext is itself a random string of bits, so an interceptor without knowledge of the key is unable to gain any information. (The security of this system was proved by Shannon.)

Note that both the sender and the recipient must have the key, which must be kept secret from the interceptor. The key must either be shared on a previous occasion, or conveyed by a channel which is known to be secure.

Public-key cryptography was invented by Diffie and Hellman in 1975. (In fact, the same idea had been invented six years earlier by James Ellis, an employee of GCHQ, who was unable to publish it because of his employment.) The idea is that the encryption algorithm and the key are made public, but the decryption is so demanding of computational resources that this knowledge is of no use to the interceptor. However, the recipient (who publishes the key) has some additional information, the *secret key*, which makes the decryption much easier.

More formally, let M be the set of plaintext messages, C the set of ciphertext messages, and K the set of keys. An *encryption system* consists of a pair of functions, *encryption* $e : M \times K \to C$, and *decryption* $d : C \times K \to M$, such that $d(e(m, k), k) = m$ for all $m \in M$ and $k \in K$. A *public-key system* also has a set S of secret keys, and an inverse pair of functions $p : S \to K$ and $q : K \to S$ such that

- computation of $e(m, k)$ and $p(s)$ are easy;
- computation of $d(c, k)$ and $q(k)$ are difficult;
- if $q(k)$ is known, then computation of $d(c, k)$ is easy (in other words, computation of $d(c, p(s))$ is easy).

Each user i of the system selects an element $s_i \in S$, computes $k_i = p(s_i)$, and publishes the result. If user j wants to send a message m to user i, she looks up k_i in the public directory, computes $c = e(m, k_i)$, and transmits this ciphertext. Now i computes $d(m, q(s_i)) = d(m, k_i) = m$. An interceptor knows c and k_i but is faced with the difficult tasks of either computing $d(m, k_i)$ directly or computing $s_i = q(k_i)$.

The RSA system works as follows. Each secret key consists of a pair p, q of large prime numbers (of hundreds of bits), and an integer a coprime to $(p-1)(q-1)$. From this, by Euclid's algorithm, one computes b such that $ab \equiv 1 \pmod{(p-1)(q-1)}$. Now the public key is the pair (N, a), where $N = pq$. The encryption algorithm takes the message m, which is encoded as an integer less than N, and computes the ciphertext $c = m^a \bmod N$. The possessor of the secret key can decrypt this by raising it to the power b mod N, since

$$c^b = m^{ab} \equiv m^1 = m \bmod N.$$

(We use the fact that $\phi(N) = (p-1)(q-1)$, and Fermat's Little Theorem asserting that $m^{\phi(N)} \equiv 1 \pmod{N}$.

Of the interceptor's two strategies, the second (calculating the secret key) involves factorising N, so that b can also be calculated. It is thought that, for most choices of secret key, the first strategy (decrypting using the public key) is also equivalent to factorising N. An intermediate strategy would be to find b such that $m^{ab} \equiv m \pmod{N}$ for all m. This amounts to finding u such that $m^u \equiv 1 \pmod{N}$, for then b can be found by the Euclidean algorithm. But the smallest such u is the least common multiple of $(p-1)$ and $(q-1)$; knowledge of this determines $(p-1)(q-1)$ and hence p and q.

So the security of the method depends on the assumption that *factorising large numbers is a hard problem.*

In a dramatic recent development, Peter Shor gave a randomised algorithm for factorising an integer in polynomial time on a quantum computer. We give only a brief account of quantum computing: see [22] for more details.

Classically, a single bit of information can take either the value 0 or 1, which we regard as lying in the set $\mathbb{Z}_2$. By contrast, a quantum state can be a superposition, or linear combination, of these two opposite states, with complex coefficients. Accordingly, a *qubit*, or quantum bit, lives in a 2-dimensional Hilbert space (a vector space over the complex numbers with Hermitian inner product). A state is a 1-dimensional subspace, which we normally represent by a unit vector spanning it. So we take an orthonormal basis of the space to consist of the two vectors e_0 and e_1, corresponding to the values zero and one of the qubit. An arbitrary state of the qubit is represented by $\alpha e_0 + \beta e_1$, where $|\alpha|^2 + |\beta|^2 = 1$.

According to the usual interpretation of quantum mechanics, we cannot observe the state directly. We can make a measurement, corresponding to a Hermitian (self-adjoint) operator on the space. The result of the measurement will be an eigenvalue of the operator. This result is not deterministic; different results will occur with appropriate probabilities. For example, we could measure the qubit in our example. The measurement could correspond to the operator of orthonormal projection onto the space spanned by e_1 (in matrix form, $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$). The eigenvalues of this operator are 0 and 1, corresponding to the eigenvectors e_0 and e_1. If the system is in the state $\alpha e_0 + \beta e_1$, then we obtain the result 0 with probability $|\alpha|^2$, and the result 1 with probability $|\beta|^2$. However, the measurement changes the state of the system; after the measurement, the system is in a state spanned by an eigenvector corresponding to the eigenvalue obtained in the measurement. (If we find the value 0 for the qubit, the system will be in the state e_0, and information about α and β is lost.)

Again, in the classical case, we can represent n bits by an n-tuple of which each entry is zero or one, that is, an element of $V = \mathbb{Z}_2^n$. Correspondingly, an n-tuple of qubits lives in a complex Hilbert space having an orthonormal basis corresponding to V. We write a typical vector in V as v, and denote by e_v the corresponding basis vector of the 2^n-dimensional Hilbert space $\mathbb{C}^{2^n}$.

Note that the Hilbert space is isomorphic to the tensor product of n copies of the 2-dimensional Hilbert space in which a single qubit lives. If $v = (v_1, v_2, \ldots, v_n)$, where $v_i \in \mathbb{Z}_2 = \{0, 1\}$ for $i = 1, 2, \ldots, n$, then

$$e_v = e_{v_1} \otimes e_{v_2} \otimes \cdots \otimes e_{v_n}.$$

Why the tensor product? Peter Shor [25] says:

> One of the fundamental principles of quantum mechanics is that the joint quantum state space of two systems is the tensor product of their individual state spaces.

The *tensor product* of two spaces is the 'universal bilinear product' of the spaces. If $\{e_1, \ldots, e_m\}$ and $\{f_1, \ldots, f_n\}$ are bases for the two spaces V and W, then a basis for the tensor product $V \otimes W$ consists of all symbols $e_i \otimes f_j$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. (Note that $\dim(V \otimes W) = \dim(V) \cdot \dim(W)$.) For $v = \sum_{i=1}^m \alpha_i e_i \in V$ and $w = \sum_{j=1}^n \beta_j f_j \in W$, we set

$$v \otimes w = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j (e_i \otimes f_j).$$

Note that, in contrast to the case of a direct sum, not every vector in the tensor product can be written as a *pure tensor* $v \otimes w$.

A remark on notation. We do not use Dirac's *bra and ket notation* beloved of physicists. However, we do have to keep straight several different vector spaces carrying various forms: we already have an n-dimensional space V over $\mathbb{Z}_2$, and a 2^n-dimensional complex Hilbert space. Shortly we will meet a $2n$-dimensional $\mathbb{Z}_2$-space $\overline{E}$ with a symplectic form! We will use v, a, b for typical vectors of V; its standard basis will be $\{u_1, \ldots, u_n\}$ (where u_i is a vector with 1 in the ith position and 0 elsewhere), and the usual dot product on V will be denoted by

$$a \cdot b = \sum_{j=1}^{n} a_j b_j.$$

We do not give a special name to the Hilbert space. Its vectors have the form $\sum_{v \in V} \alpha_v e_v$ where $\alpha_v \in \mathbb{C}$; the inner product of $\sum \alpha_v e_v$ and $\sum \beta_v e_v$ is $\sum \overline{\alpha_v} \beta_v$.

The theoretical quantum computers which we will discuss will be built from 'quantum circuits'. A *quantum circuit* is built out of "logical quantum wires," each corresponding to one of the n qubits, and *quantum gates*, each acting on one or two wires [25]. A quantum gate is a unitary transformation, since all possible physical transformations of a quantum system are unitary. Shor assumes that each gate acts on either one or two wires; that is, each maps 1 qubit to 1 qubit or 2 qubits to 2 and acts as the identity on the remaining qubits. The reason that we are able to restrict ourselves to one- or two-bit gates is the fact that, as in classical first-order logic, a small number of operations forms a 'universal set' up from which all possible operations can be built. In the case of quantum computing, in the words of Shor [25], "CNOT together with all quantum one-bit gates forms a universal set." CNOT here refers to the two-bit *controlled not* gate which takes the basis vector $e_{(x,y)}$ of $\mathbb{C}^4 = \mathbb{C}^2 \otimes \mathbb{C}^2$ to $e_{(x,x+y)}$. It is clearly a unitary transformation, and is given the name 'controlled not' because the target bit y is negated or not according as the controller bit x equals 1 or 0.

It is the two-qubit gates, and therefore CNOT in particular, which provides a quantum computer with its inherent parallelism. These gates are intrinsically global: there is no way to describe them by restricting attention to a single qubit. Because they accept as input the tensor product of *superpositions* (linear combinations) of e_0 and e_1, they are the gates that makes the computation quantum rather than classical.

As we mentioned, Shor [25] found a probabilistic algorithm which runs in polynomial time on a quantum computer. It depends on the following observation. Suppose that we have computed the order of x mod N, the smallest number t such that $x^t \equiv 1 \pmod{N}$. Suppose that t is even, say $t = 2r$. Then N divides $x^t - 1 = (x^r - 1)(x^r + 1)$. If x^r is not congruent to -1 mod N, then both $x^r - 1$ and $x^r + 1$ have factors in common with N. We can find the g.c.d.s of the pairs $(N, x^r - 1)$ and $(N, x^r + 1)$ by Euclid's algorithm; then we know two different factors of N, and hence the complete factorisation in the RSA case.

The basic idea of quantum computation is to exploit the inherent parallelism of quantum systems. Suppose, for example, that we want to compute 2^n values $f(0), f(1), \ldots, f(2^n - 1)$ of a function simultaneously. We represent the integer i by the binary vector $v \in V = \mathbb{Z}_2^n$ which is its expression in base 2. Now, if n qubit registers are available, we can load them with a superposition of the states e_v, for $v \in V$, as follows: first load each register with 0 (so that we have state e_0); then apply the Hadamard transformation with matrix $(1/\sqrt{2})\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ (corresponding to a 45° rotation of the 2-dimensional Hilbert space) to each qubit. The resulting state is

$$\frac{1}{\sqrt{2^n}} \sum_{v \in V} e_v.$$

Now suppose that some quantum computation replaces e_v by a state representing $f(i)$, where v represents the integer i. Then, by the linearity of the Schrödinger equation, the same computation replaces the above superposition by a superposition of the values $f(i)$ for $i = 0, \ldots, 2^n - 1$. In the factorisation algorithm, we take $f(i) = x^i \pmod{N}$, where $2^n \geq N$ and x is some chosen integer coprime to N.

The last stage involves finding the period of f from the above superposition, which is done by use of the discrete Fourier transform. The Fourier transform of f is concentrated on multiples of the period, so an observation of the result will with high probability give a small multiple of the period.

Let us take as an example the current objective of quantum computing, the factorisation of 15. We take $x = 2$. Suppose that we have 8 qubit registers arranged in two banks of four, so that states of the system have the form

$$\sum_{v,w \in V} a_{vw}(v \otimes w)$$

in the space $\mathbb{C}^{2^4} \otimes \mathbb{C}^{2^4}$. For convenience we write e_i in place of e_v, where i is the integer in the range $[0, 15]$ whose base 2 representation is v. We begin with the state $e_0 \otimes e_0$ (that is, all registers contain zero). Then by applying a Hadamard transform to each of the first four qubits, we obtain (up to normalisation) $\left(\sum_{i\in[0,15]} e_i\right) \otimes e_0$. The crucial part of the computation replaces $e_i \otimes e_0$ with $e_i \otimes e_{2^i \bmod 15}$. Apart from normalisation, the state is now

$$(e_0 + e_4 + e_8 + e_{12}) \otimes e_1 + (e_1 + e_5 + e_9 + e_{13}) \otimes e_2 + \cdots .$$

We extract the period 4 from the first four registers by a discrete Fourier transform. Now we know that 15 divides $(2^4 - 1) = (2^2 - 1)(2^2 + 1) = 3 \cdot 5$, so the two factors of 15 are $(15, 3) = 3$ and $(15, 5) = 5$.

See Shor [24, 25] for further details.

10 Quantum codes

We see that quantum computing is a technology with very great potential uses. What stands in the way of its implementation is the large error rate, caused by the fact that a single bit or qubit is stored by a single electron, instead of by billions of electrons as in a conventional computer. It is widely believed that a quantum computer large enough for real applications will have to be 'fault-tolerant', that is, error-correction must be built in so that the errors introduced by the gates and wires can be corrected faster than they occur. The theory of quantum error-correction will be outlined below; it has not been implemented yet.

The evolution of a quantum system is described by a unitary transformation of the state space. We consider 'errors' to a single qubit represented by the following unitary matrices:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \textit{(bit error)}$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \textit{(phase error)}$$

The effect of X is to interchange the basis vectors e_0 and e_1 (the zero and one states of the qubit). The effect of Z is to change the relative phase of the coefficients α and β of an arbitrary state (the argument of β/α) by 180°.

(The arguments of α and β have no absolute significance, so we could as well use $-Z$, but the given choice is more convenient.)

Along with X and Z, we also allow the identity I (no error) and the product $Y = XZ = -ZX$. (It is more usual in quantum mechanics to set $Y = \mathrm{i}XZ$; then X, Y, Z are the standard *Pauli spin matrices.* Note that X, Z and $\mathrm{i}XZ$ are Hermitian (or self-adjoint). However, $Y = XZ$ is unitary, and then everything can be written with real coefficients. A reason to use $Y = \mathrm{i}XZ$ is that then Y is conjugate to both X and Z, which means a change of basis transforms any one to any other, and we can regard all three non-trivial errors are *equally likely.* However, we shall see that the discrete mathematics is the same whether we make this choice or the simpler real choice, so in what follows, we use $Y = XZ$.)

There is a simple expression for the effect of X and Z on the basis $\{e_0, e_1\}$. We use the convention that $(-1)^0 = 1$ and $(-1)^1 = -1$, where the exponent is taken from the finite field $\mathbb{Z}_2$. Then we have, for $v \in \mathbb{Z}_2$,

$$Xe_v = e_{v+1}, \qquad Ze_v = (-1)^{1\cdot v} e_v.$$

Now we can apply these errors 'coordinatewise' to n qubits. If u_j denotes the jth basis vector for $V = \mathbb{Z}_2^n$, then the errors to the jth qubit act on the 2^n-dimensional Hilbert space with basis $\{e_v : v \in V\}$ as

$$\begin{aligned} X(u_j) &: e_v \mapsto e_{v+u_j}, \\ Z(u_j) &: e_v \mapsto (-1)^{v\cdot u_j} e_v. \end{aligned}$$

Then we can define $X(a)$, $Z(b)$ for any vectors $a, b \in V$ by

$$\begin{aligned} X(a) &: e_v \mapsto e_{v+a}, \\ Z(b) &: e_v \mapsto (-1)^{v\cdot b} e_v. \end{aligned}$$

Then $\{X(a) : a \in V\}$ and $\{Z(b) : b \in V\}$ are groups of unitary transformations of $\mathbb{C}^{2^n}$ which are both isomorphic to the additive group of V. Together they generate the group

$$E = \langle X(a), Z(b) : a, b \in V\rangle = \{\pm X(a)Z(b) : a, b \in V\},$$

a group of order $2 \cdot 2^n \cdot 2^n = 2^{2n+1}$.

We call E the *group of allowable errors*, or, for short, the *error group.* It can also be written in the form

$$E = \{\pm A_1 \otimes \cdots \otimes A_n : A_j \in \{I, X, Y, Z\} \text{ for } j = 1, \ldots, n\}.$$

Why do we only consider these particular errors? A quantum code can correct any one-qubit error if and only if it can correct the errors $X(u_j)$, $Y(u_j)$ and $Z(u_j)$ for all j. This follows from the fact that the matrices I, X, Y, Z span the 4-dimensional space of all 2×2 matrices: see [4, 9].

In our set-up, the error group

$$E = \{\pm X(a)Z(b) : a, b \in V\}$$

is an extraspecial 2-group of order 2^{2n+1}. Its centre is $\zeta(E) = \{\pm I\}$, and $\overline{E} = E/\zeta(E)$ is a vector space over F. This is the third, and most important, vector space with which we have to deal. (See Section 8.)

We use the notation $(a|b)$ for the element of $\overline{E}$ which is the coset of $\zeta(E)$ containing $X(a)Z(b)$. (This coset is $\{X(a)Z(b), -X(a)Z(b)\}$.) We use $*$ for the bilinear form on $\overline{E}$ derived from commutation on E. In other words,

$$[X(a)Z(b), X(a')Z(b')] = (-1)^{(a|b)*(a'|b')}I.$$

A short calculation shows that

$$(a|b) * (a'|b') = a \cdot b' - a' \cdot b,$$

where $\cdot$ is the dot product on V. (Of course, we can replace the $-$ sign by a $+$ sign since the characteristic is 2.) Thus, two elements $e, f \in E$ commute if and only if $\overline{e} * \overline{f} = 0$.

The quadratic form will be denoted by Q: that is,

$$(X(a)Z(b))^2 = (-1)^{Q(a|b)}I.$$

Note that in fact $Q(a|b) = a_1b_1 + \cdots + a_nb_n$. (If we had made the choice $Y = \mathrm{i}XZ$ instead of $Y = XZ$, we would need to enlarge our group E to include $\mathrm{i}I$. As a consequence, E would have a larger center, $\zeta(E) = \{(\mathrm{i}^\ell)I : \ell = 0, 1, 2, 3\}$. Nonetheless, the factor group $E/\zeta(E)$ would be "the same" elementary abelian 2-group of order 2^{2n}, and commutation on E would give the same bilinear form on this version of $\overline{E}$. However, we would lose the quadratic form on $\overline{E}$ in this case since e^2 would not be well-defined on $\overline{E}$.)

More on notation. The basis $\{(u_j|0), (0|u_j) : j = 1, \ldots, n\}$ is a symplectic basis for $\overline{E}$. We will reserve the symbol $\perp$ for orthogonality in this space (with respect to the form $*$). We also make another use of $\perp$, derived from this one. If S is a subgroup of E, then we let $S^\perp$ be the centraliser of S in E. (If $S' = C_E(S)$, then $\overline{S'} = \overline{S}^\perp$, so the notation is consistent.)

To avoid confusion with orthogonality in $\overline{E}$, we use a non-standard symbol for orthogonality of vectors in V. For $U \subseteq V$ we'll write $U^{\ddagger} = \{v \in V : v \cdot u = 0 \text{ for all } u \in U\}$.

We will use $\langle \ldots \rangle_V$, $\langle \ldots \rangle_{\overline{E}}$, $\langle \ldots \rangle_{\mathbb{C}}$ for the subspace of V, $\overline{E}$ or $\mathbb{C}^{2^n}$ respectively, spanned by a set $\ldots$; and $\langle \ldots \rangle$ (without adornment) will denote the subgroup generated by $\ldots$.

We are interested in abelian subgroups of E. Note that any set of mutually commuting transformations on the Hilbert space $\mathbb{C}^{2^n}$ which contains the adjoints of all its elements is simultaneously diagonalisable, and has an orthonormal basis consisting of eigenvectors. Said otherwise, the common eigenspaces of the transformations in the set are mutually orthogonal. Any abelian subgroup of E satisfies this.

Choose an abelian subgroup S of E. (Equivalently, choose a subspace $\overline{S}$ of $\overline{E}$ for which $\overline{S} \leq \overline{S}^{\perp}$.) Let $\dim(\overline{S}) = n - k$. As in the previous section, let Q be an eigenspace of S. Then the following hold.

(a) For $s' \in S^{\perp}$, $q \in Q$, we have $s'(q) \in Q$.

(b) Any $e \in E$ permutes the eigenspaces Q_u; it fixes Q if and only if it lies in $S^{\perp}$; so $E/S^{\perp}$ permutes the eigenspaces regularly.

(c) The eigenspaces of S are in one-to-one correspondence with the characters of $\overline{S}$: an eigenspace Q' determines a character χ of S satisfying $s(q) = \chi(s)q$ for $s \in S$ and $q \in Q'$ with $\chi(-I) = -1$ if $-I \in S$; χ thus defines a character of $\overline{S}$.

Now (c) implies that S has 2^{n-k} eigenspaces, one for each character of $\overline{S}$; since $E/S^{\perp}$ permutes these regularly each must have the same dimension, and since they form an orthogonal decomposition of $\mathbb{C}^{2^n}$, this dimension must be 2^k.

Choose Q to be the eigenspace corresponding to the trivial character. Then we have, as in the example above:

(d) For $s \in S$, $q \in Q$, we have $s(q) = q$.

Now Q has dimension 2^k, and is isomorphic to the space of k qubits, but the embedding of Q in the larger space 'smears out' the k qubits over the space of n qubits. (This is exactly analogous to the situation in classical error-correcting codes: there, $V = \mathbb{Z}_2^n$ is the space of n bits, and if G is a generator matrix for C then $C = \{xG : x \in \mathbb{Z}_2^k\}$ is a k-dimensional linear code, which 'smears out' the k information bits of x over the space V.) So Q will be our quantum error-correcting code.

In the classical case, a message consists of n bits, that is, it is a vector in V. Errors are also vectors in V: we have 'received word equals transmitted word plus error'. The zero error has no effect. If we use a code C, then errors in C are undetectable. There will be a set $\mathcal{E}$ of so-called *correctable errors*. They have the property that, if $e, f \in \mathcal{E}$ with $e \neq f$, then $e - f \notin C \setminus \{0\}$, that is, $e - f$ is not an undetectable error. This means that we can detect the addition of an error to a codeword, and we do not confuse the effects of different errors (as long as they lie in $\mathcal{E}$). The commonest situation is that when C has minimum weight at least d, when $\mathcal{E}$ consists of all words of weight at most $\lfloor (d-1)/2 \rfloor$.

Table 1 compares the situation in the classical and quantum cases.

	Classical	Quantum
Message	n bits, in $\mathbb{Z}_2^n = V$	n qubits, in $\mathbb{C}^{2^n}$
Error group	V $e(z) = z + v$	E $e(z)$ as before
Code	$C \leq V$, $\dim(C) = k$	$Q \leq \mathbb{C}^{2^n}$, $\dim(Q) = 2^k$
Undetectable errors	$C \leq V$	$S^\perp \leq E$
Errors with no effect	$\{0\} \leq C$	$S \leq S^\perp$
Correctable errors	$\mathcal{E} \subseteq V$ $e, f \in \mathcal{E} \Rightarrow$ $e - f \notin C \setminus \{0\}$	$\mathcal{E} \subseteq E$ $e, f \in \mathcal{E} \Rightarrow$ $f^{-1}e \notin S^\perp \setminus S$

Table 1: Classical and quantum error correction

We now give the basic result of Calderbank, Rains, Shor and Sloane, which is the quantum analogue of the observation on page 2.

Theorem 10.1 *With the notation as above, assume that the minimal q-weight of $\overline{S}^\perp \setminus \overline{S}$ is $d \geq 2e + 1$. Then Q corrects errors in e qubits.*

Here, the *quantum weight*, or q-weight, of $(a|b) \in \overline{E}$ (or, equivalently, of $\pm X(a)Z(b) \in E$) is the number of coordinates j in which either $a_j \neq 0$ or

$b_j \neq 0$ (in other words, the number of qubits which have suffered a bit error, a phase error, or both). The theorem asserts that, if $\mathcal{E}$ is the set of elements with q-weight at most $\lfloor (d-1)/2 \rfloor$, then we have $e, f \in \mathcal{E} \Rightarrow f^{-1}e \notin S^{\perp} \setminus S$. We will use $[[n, k, d]]$ to denote such a code; the double brackets to distinguish it from a classical $[n, k, d]$ binary code.

Suppose Q is an additive code, that is, Q is the $+1$ eigenspace for $S \leq E$, and $\mathcal{E}$ is the set of correctable errors for Q. Here is how decoding works. Suppose the codeword $q \in Q$ is sent, but the received message is $z = e(q)$ for some $e \in \mathcal{E}$. If $s_1, \ldots, s_{n-k}$ are a set of generators for S, then by calculating $s_j(z) = \chi(s_j)z$ for $1 \leq j \leq n-k$ (the *syndrome* of e) , we can identify the character χ and therefore the eigenspace Q' containing z. We know $Q' = f(Q)$ for some $f \in \mathcal{E}$, and we decode z to $f^{-1}(z) = f^{-1}e(q)$. (Notice that this decoding step does not require knowing e or q.) In order for this procedure to lead us to the correct codeword q it's necessary that $e(Q) = f(Q)$ for $e, f \in \mathcal{E}$ should imply $f^{-1}e(q) = q$. But that is exactly our condition on $\mathcal{E}$: $f^{-1}e$ cannot be in $S^{\perp} \setminus S$.

Now, the extra condition that distinct e and f in $\mathcal{E}$ should produce independent vectors $e(q)$ and $f(q)$ means that $f(Q) = e(Q)$ implies $f = e$. We will say that the code is *non-degenerate* (or *pure*) if it satisfies the stronger condition that $e, f \in \mathcal{E}$ implies $f^{-1}e \notin \overline{S}^{\perp} \setminus \{0\}$. So, in the special case of a non-degenerate code, identifying the coset containing z actually identifies e (rather than just enabling e to be corrected).

Following the theorem in the previous section, to obtain an additive code mapping k qubits to n and correcting errors in $\lfloor (d-1)/2 \rfloor$ qubits, we need to find a totally isotropic $(n-k)$-dimensional subspace $\overline{S}$ of the $2n$-dimensional binary space $\overline{E}$ with $\overline{S}^{\perp} \setminus \overline{S}$ having minimum q-weight d. In this section we construct a specific example of an $[[8, 3, 4]]$ code adapted from [7] and [14].

To begin, we need a 5-dimensional subspace $\overline{S} \leq \overline{S}^{\perp}$ of the 16-dimensional binary space $\overline{E}$, and we would like the minimal non-zero q-weight of vectors in $\overline{S}^{\perp}$ to be 4.

Vectors in $\overline{E}$ have the form $(a|b)$ for $a, b \in V$, where V is a binary space of dimension 8. We base our construction of $\overline{S}$ on the $[8, 4, 4]$ extended Hamming code $\widehat{mH}_3 = \overline{\mathcal{H}(3, 2)}$, the self-dual code obtained by extending the $[7, 4, 3]$ Hamming code $\mathcal{H}(3, 2)$ by adding a parity-check bit to each vector. (We now use the hat notation for the usual dual code, to avoid confusion with our overloaded overbar.)

We construct $\mathcal{H}_3$ by specifying a 3×7 *parity-check matrix* H. The columns of H will consist of all possible non-zero vectors in $\mathbb{Z}_2^3 \simeq \mathrm{GF}(8)$. We do this by choosing a generator $\alpha = \alpha^3 + 1$ of the multiplicative group of GF(8), and letting the columns of H be $\alpha^6, \ldots, \alpha, 1$, (where $x\alpha^2 + y\alpha + z$ is written as the column $[x, y, z]^\top$) giving

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

This construction of H makes it easy to see that $\mathcal{H}_3$ is a *cyclic* code: if $v = (v_1, \ldots, v_7)$ is in $\mathcal{H}_3$, so is its cyclic shift $v' = (v_n, v_1, \ldots, v_{n-1})$. From H we easily obtain a *generator matrix* G for $\mathcal{H}_3$—recall a code is the row space of its generator matrix.

$$G_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Since the vector $\overline{1} = (1, 1, 1, 1, 1, 1, 1)$ is in $\mathcal{H}_3$, we get another generator matrix G_1 by adding $\overline{1}$ to each row of G_0.

$$G_1 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

We obtain the $[8, 4, 4]$ extended Hamming code $\widehat{\mathcal{H}}_3$ by adding a parity check bit to the front of each vector in $\mathcal{H}_3$ so that the resulting vector has even H-weight. The minimum H-weight of nonzero vectors in $\mathcal{H}_3$ is 3, and the minimum H-weight is 4 for $\widehat{\mathcal{H}}_3$. Moreover, $\widehat{\mathcal{H}}_3 = \widehat{\mathcal{H}}_3^\ddagger$.

Finally, we're ready to describe a 5×16 generator matrix for $\overline{S}$. Let a be the last row of G_1 and b be its first. Let a_1 be the extension of a and b_1 be the extension of b, two vectors in $\widehat{\mathcal{H}}_3$. Next a_2 is the extension of the cyclic shift a', and b_2 is the extension of b'. Similarly a_3 and b_3 are extensions of a'' and b''. We take $a_4 = b_5$ to be the 8-tuple $\overline{1}$ and $a_5 = b_4$ to be the 8-tuple $\overline{0}$.

The 5×16 matrix $G^{(0)}$ has for its rows the vectors $(a_1|b_1), \ldots, (a_5|b_5)$.

$$G^{(0)} = \left[\begin{array}{cccccccc|cccccccc} 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$$

The rows of $G^{(0)}$ are linearly independent, so its row space $\overline{S}$ is of dimension 5. Because $\widehat{\mathcal{H}}_3 = \widehat{\mathcal{H}}_3^{\ddagger}$, we see that $(a_i|b_i) * (a_j|b_j) = a_i \cdot b_j + a_j \cdot b_i = 0$ for $i, j = 1, \ldots, 5$, so $\overline{S} \leq \overline{S}^{\perp}$. The dimension of $\overline{S}^{\perp}$ is $16 - 5 = 11$, and so $\overline{S}^{\perp} = \overline{S} \oplus T$ for T of dimension 6. We can choose a basis $\{(a_j|\overline{0}), (\overline{0}|b_j) : j = 1, 2, 3\}$ for T, and from this we can see that the minimum q-weight of $\overline{S}^{\perp}$ is 4.

Our code Q is the $+1$ eigenspace of the subgroup $S < E$ acting on the 2^8-dimensional complex vector space $\mathbf{C}^2 \otimes \ldots \otimes \mathbf{C}^2$. The subspace Q is of complex dimension 2^3; Q thus maps 3 qubits to 8 qubits and corrects errors in E that affect 1 qubit.

We now return briefly to the general theory, to describe a theorem of Calderbank *et al.* [6] which shows how to construct additive quantum codes from certain self-orthogonal codes over GF(4). Recall that we have associated a $[[n, k, d]]$ additive quantum codes to a $n-k$-dimensional subspace $\overline{S}$ of $\overline{E}$ (a $2n$-dimensional vector space over $\mathbb{Z}_2 = \mathrm{GF}(2)$) which is totally isotropic with respect to a symplectic inner product and for which the q-weight of $S^{\perp} \setminus S$ is d.

In order to make the association with codes over GF(4), we construct yet another vector space, an n-dimensional vector space F^n over the field $F = \mathrm{GF}(4)$ We will write F as $\{0, 1, \omega, \overline{\omega}\}$ where $\overline{\omega} = \omega^2 = 1 + \omega$. Note that the cube of every non-zero element is equal to 1, since the multiplicative group has order 3. Since ω and $\overline{\omega}$ form a basis for F as vector space over GF(2), we can write any vector of F^n as $\omega a + \overline{\omega} b$, where a and b are vectors of length n over GF(2). In other words, $a, b \in V$. Now the map $\phi : \overline{E} \to F^n$ defined by

$$\phi((a|b)) = \omega a + \overline{\omega} b$$

is a bijection, and is an isomorphism of GF(2)-vector spaces (if we regard F^n as a GF(2)-space by restricting scalars). Since $\omega a_j + \overline{\omega} b_j \neq 0$ if and only if either $a_j \neq 0$ or $b_j \neq 0$, we have a very important property of ϕ: the q-weight of a vector $(a|b) \in \overline{E}$ is equal to the Hamming weight of its image $\phi((a|b)) = \omega a + \overline{\omega} b \in F^n$.

Let $\circ$ be the Hermitian inner product on F^n,

$$v \circ w = \sum_{j=1}^{n} \overline{v_j} w_j.$$

There is a *trace map* Tr from F to GF(2) defined by

$$\mathrm{Tr}(\alpha) = \alpha + \overline{\alpha},$$

so that $\mathrm{Tr}(0) = \mathrm{Tr}(1) = 0$ and $\mathrm{Tr}(\omega) = \mathrm{Tr}(\overline{\omega}) = 1$. The trace map is linear over GF(2). Now the map $(x, y) \mapsto \mathrm{Tr}(x \circ y)$ takes pairs of vectors in F^n to GF(2). We have:

$$\mathrm{Tr}((\omega a + \overline{\omega} b) \circ (\omega a' + \overline{\omega} b')) = a \cdot b' + a' \cdot b = (a|b) * (a'|b').$$

The proof of this fact is just calculation. Using the fact that the Hermitian inner product is linear in the first variable and semilinear in the second, and the fact that $a \circ b = a \cdot b$ for $a, b \in \mathrm{GF}(2)^n$, we have

$$(\omega a + \overline{\omega} b) \circ (\omega a' + \overline{\omega} b') = a \cdot a' + \overline{\omega} a \cdot b' + \omega b \cdot a' + b \cdot b'.$$

Taking the trace now gives the result (using the linearity of trace and the fact that $\mathrm{Tr}(1) = 0$ and $\mathrm{Tr}(\omega) = \mathrm{Tr}(\overline{\omega}) = 1$).

Next we show that a subspace of F^n is totally isotropic (with respect to $\circ$) if and only if the corresponding subspace of $\overline{E}$ is totally isotropic (with respect to $*$). The forward implication is clear. So suppose that $W \leq F^n$ is the image under ϕ of a totally isotropic subspace of $\overline{E}$. This means that $\mathrm{Tr}(x \circ y) = 0$ for all $x, y \in W$. Take $x, y \in W$, and let $x \circ y = \alpha$. Then $\mathrm{Tr}(\alpha) = 0$, and

$$\mathrm{Tr}(\omega\alpha) = \mathrm{Tr}((\omega x) \circ y) = 0$$

(since $\omega x \in W$); it follows that $\alpha = 0$.

To summarise: the space F^n, on restriction of scalars, becomes isomorphic to $\overline{E}$, by a map ϕ taking q-weight to Hamming weight; and subspace of F^n is the image of a totally isotropic subspace of $\overline{E}$ (with respect to $*$) if and only if it is totally isotropic (with respect to $\circ$). Of course, not every totally isotropic subspace of $\overline{E}$ corresponds to a subspace of W. (The image under ϕ of a subspace of $\overline{E}$ is a GF(2)-subspace but not necessarily a GF(4)-subspace.)

These observations, together with Theorem 10.1, imply the following result of Calderbank *et al.* [6]:

Theorem 10.2 *If W is a totally isotropic (with respect to the hermitian inner product) ℓ-dimensional subspace of F^n such that $W^\perp \setminus W$ has minimum Hamming weight d, then the above associates an additive $[[n, n-2\ell, d]]$ quantum code to W.*

So, in order to construct good quantum error-correcting codes, we need subspaces C of W such that $C^\perp \setminus C$ has large minimum weight (where $C^\perp$ is defined by the Hermitian inner product $\circ$. Examples can be obtained from dual Hamming and BCH codes. We do not give details here.

11 $\mathbb{Z}_4$-codes

We already noted the existence of *Kerdock codes*, which are non-linear $(2^{2n}, 2^{4n})$ codes with weight enumerator

$$x^{2^{2n}} + (2^{4n} - 2^{2n+1})x^{2^{2n-1}+2^{n-1}}y^{2^{2n-1}-2^{n-1}} + (2^{2n+1}-2)x^{2^{2n-1}}y^{2^{2n-1}} \\ +(2^{4n} - 2^{2n+1})x^{2^{2n-1}+2^{n-1}}y^{2^{2n-1}-2^{n-1}} + y^{2^{2n}}.$$

Moreover, these codes are *distance-invariant*; that is, the weight distribution of $u + C$ is the same as that of C for all $u \in C$.

At about the same time, another family of non-linear binary codes, the *Preparata codes*, were discovered. They are also distance-invariant, and their weight enumerators are obtained from those of the Kerdock codes by applying the MacWilliams transformation. Thus, the Kerdock and Preparata codes behave formally like duals of each other. This strange formal duality was not understood for a long time, until the paper of Hammons *et al.* [15] showed that they arise from dual codes over $\mathbb{Z}_4$ by applying the so-called Gray map.

The *Gray map* takes elements of $\mathbb{Z}_4$ to pairs of elements of $\mathbb{Z}_2$, as follows:

$$0 \mapsto 00, \quad 1 \mapsto 01, \quad 2 \mapsto 11, \quad 3 \mapsto 10.$$

Note that it is an isometry between the set $\mathbb{Z}_4$ with the *Lee metric*

$$d(x, y) = \min\{|x-y|, 4-|x-y|\}$$

(so that $d(x, y)$ is the number of places round the cycle separating x from y) and $\mathbb{Z}_2^2$ with the Hamming metric. (More generally, a Gray code is a Hamiltonian cycle in the n-dimensional cube; it is used for analog-to-digital conversion, since adjacent points in the cycle are represented by words differing in only one coordinate. We are interested in the case $n = 2$.)

The Gray map γ can be extended to a map from $\mathbb{Z}_4^n$ to $\mathbb{Z}_2^{2n}$, for any n. It is non-linear, so that it will in general take a linear code in $\mathbb{Z}_4^n$ (a subset closed under addition) to a non-linear code in $\mathbb{Z}_2^{2n}$. Hammons *et al.* showed that the Kerdock and Preparata codes do indeed arise from linear codes over $\mathbb{Z}_4$ in this way; these linear codes are the $\mathbb{Z}_4$ analogues of the dual Hamming and Hamming codes.

In the remainder of this section we outline the construction of the binary and quaternary codes and their connection with symplectic and orthogonal geometry.

11.1 Orthogonal and symplectic geometry

Recall the definition of the error group E of isometries of a real vector space $\mathbb{R}^N$, $N = 2^{m+1}$, m odd. Let $V = GF(2)^{m+1}$, and let $\{e_v : v \in V\}$ be an orthonormal basis of $\mathbb{R}^N$. The isometry $X(a)$ takes e_v to e_{v+a} and $Z(b)$ takes e_v to $(-1)^{b \cdot v} e_v$ for $a, b \in V$; $X(a)$ describes "bit errors" in each qubit for which the corresponding coordinate of a is nonzero, and $Z(b)$ describes "phase errors." The error group for a system of $m+1$ qubits is

$$E = \{(-1)^\ell X(a) Z(b) : a, b \in V, \ell \in \mathbb{Z}_2\}.$$

Then E is an extraspecial 2-group of order $2 \cdot 2^{2(m+1)}$.

From the group structure of E it follows that the quotient $\overline{E} = E/\zeta(E) \simeq \mathrm{GF}(2)^{2(m+1)}$ has an orthogonal geometry. The quadratic form Q on $\overline{E}$ is given by

$$e^2 = (-1)^{Q(\overline{e})} I,$$

and the corresponding symplectic form is given by

$$[e, f] = (-1)^{\overline{e} * \overline{f}} I,$$

where e and f are in E and $\overline{e}$ and $\overline{f}$ are their images in $\overline{E}$. The abelian subgroups

$$X(V) = \{X(a) : a \in V\} \text{ and } Z(V) = \{Z(b) : b \in B\}$$

give *totally singular* $(m+1)$-dimensional subspaces $\overline{X(V)}$ and $\overline{Z(V)}$ of $\overline{E}$ (that is, subspaces on which Q vanishes identically). These totally singular spaces are also totally isotropic: the symplectic form vanishes identically on them. In fact, every maximal totally singular subspace of $\overline{E}$ has dimension $m+1$ and arises as the image of an elementary abelian subgroup of E. Similarly, every maximal totally isotropic subspace of $\overline{E}$ also has dimension $m+1$ and arises as the image of an abelian subgroup of E. Thus, the quadratic form has type $+1$ (or Witt index $m+1$), and the extraspecial group E is a central product of $m+1$ copies of the dihedral group D_8 (see Section 8.)

We don't need the physics, but by way of motivation, recall from Section 10 that we might alternatively have considered an error group consisting of (Hermitian) symmetries of the complex space containing the qubits. In our current setting we need to construct this complex geometry, but now we do it entirely within the group E of symmetries of the real vector space.

First, we need a bit more notation. As before, let $\{u_1, \ldots, u_{m+1}\}$ be the standard basis for V, and let $V' = \langle u_1, \ldots, u_m \rangle \simeq \mathrm{GF}(2)^m$. Let

$$\omega = X(u_{m+1})Z(u_{m+1}) \in E,$$

so $\omega^2 = -I$. This element of order 4 will play the role of i in our construction of a complex vectors space and a group of unitary transformations of it.

Let F be the centraliser in E of ω,

$$F = C_E(\omega) = \{(-1)^\ell X(a)Z(b) : a \cdot u_{m+1} = b \cdot u_{m+1},\ \ell \in \mathbb{Z}_2\}.$$

Exercise 11.1 Show that if $X(a)Z(b)$ is in F, then there are $a', b' \in V'$ so that either $X(a)Z(b) = X(a')Z(b')$ or $X(a)Z(b) = \omega X(a')Z(b')$.

By the previous exercise, an alternative description of F is

$$F = \{\omega^\ell X(a')Z(b') : a', b' \in V', \ell \in \mathbb{Z}_4\},$$

which corresponds to the complex error group of the preceding chapter. It is easily seen that F has order $4 \cdot 2^{2m}$ and that $\overline{F} = F/\zeta(F) \simeq \mathrm{GF}(2)^{2m}$.

Because ω is of order 4, we may think of $\mathbb{R} + \mathbb{R}\omega$ as $\mathbb{C}$. As a consequence, we may regard the 2-dimensional real space $\langle e_v, e_v\omega \rangle$ as a 1-dimensional complex space. Under this identification, we can consider $\{e_{v'} : v' \in V'\}$ as an orthonormal basis of the complex unitary space $\mathbb{C}^{N'}$, for $N' = 2^m = (1/2)N$, and regard F as a subgroup of the unitary group $U(\mathbb{C}^{N'})$.

Since the square of an element $\omega^\ell X(a')Z(b')$ of F depends on ℓ as well as on a' and b', we cannot define a quadratic form on $\overline{F}$. However, commutators depend only on cosets modulo $\zeta(F)$, so $\overline{F}$ does have a symplectic geometry given by the bilinear form $\overline{e} * \overline{f}$ where e and f are preimages in E and

$$[e, f] = (-1)^{\overline{e}*\overline{f}} I.$$

Caution: Our overbar notation is now ambiguous, since $\overline{E}$ and $\overline{F}$ are binary spaces of dimensions $2(m+1)$ and $2m$ respectively; however, context should make clear which we intend.

Finally, following [5], we will need to construct two finite groups $L \leq O(\mathbb{R}^N)$ and $L^\natural \leq U(\mathbb{C}^{N'})$ normalizing E and F respectively for which $L/E \simeq O(2m+2, 2)$ on $\overline{E}$ and $L^\natural/F \simeq Sp(2m, 2)$ on $\overline{F}$.

Exercise 11.2 Let $x_j = X(u_j)$ and $z_j = Z(u_j)$, $j = 1, \ldots, m+1$.

(a) Show that the images of these elements of E in $\overline{E}$ form a symplectic basis of singular vectors, with $\{\overline{x_j} : j = 1, \ldots, m+1\}$ a basis for the maximal totally singular subspace $\overline{X(V)}$ and $\{\overline{z_j} : j = 1, \ldots, m+1\}$ a basis for the maximal totally singular subspace $\overline{Z(V)}$.

(b) Show that $\{x_j, z_j : j = 1, \ldots, m\}$ are in F, and their images in $\overline{F}$ are a symplectic basis with $\{\overline{x_j} : j = 1, \ldots, m\}$ a basis for the maximal totally isotropic subspace $\overline{X(V')}$ and $\{\overline{z_j} : j = 1, \ldots, m\}$ a basis for the maximal totally isotropic subspace $\overline{Z(V')}$.

In our constructions of L and $L^\natural$, we will be guided by the following theorem.

Theorem 11.1 *Let V be a vector space of dimension $2n$ with an alternating bilinear form and a quadratic form of Witt index n polarising to it. If V is a sum of maximal totally isotropic subspaces U and V, then $Sp(V) = \langle Sp(V)_U, Sp(V)_W \rangle$. Further, if U and W are totally singular, then $O(V) = \langle O(V)_U, O(V)_W, T \rangle$, where T is the orthogonal transformation interchanging u_1 and v_1 and fixing the other vectors of a symplectic basis of singular vectors $\{u_1, \ldots, u_n, w_1, \ldots, w_n\}$ formed of bases of U and W.*

The reference for this theorem in [5] is to 43.7 in [1]. Other useful references include the introductory material on the classical groups in [12], the "dictionary" translating between matrix and Lie theoretic descriptions of classical groups in [10], and the explanatory material in [20].

First we construct L. For an invertible binary $(m+1)\times(m+1)$ matrix A, we want an element of $O(\mathbb{R}^N)$ normalizing E and producing $X(A, A^{-\top})$ on $\overline{E}$. Choose $\tilde{A}$ taking the basis vector e_v of $\mathbb{R}^N$ to e_{vA}.

Exercise 11.3 Verify that $\tilde{A}^{-1}X(a)Z(b)\tilde{A} = \pm X(aA)Z(bA^{-\top})$.

The description of the element of $O(\mathbb{R}^N)$ producing $Y(C)$ is more complicated. Choose an $(m+1)\times(m+1)$ binary alternate matrix C. Define an alternate bilinear form B_C on V by $B_C(u,v) = uCv^\top$. Let Q_C be any quadratic form polarising to B_C. Define an element $D(C) \in O(\mathbb{R}^N)$ by

$$D(C)(e_v) = (-1)^{Q_C(v)}e_v \text{ for } v \in V.$$

Exercise 11.4 Verify that $D(C)^{-1}X(a)Z(b)D(C) = \pm X(a+aC)Z(b)$, and that the map on $\overline{E}$ produced by $D(C)$ is independent of the choice of the quadratic form Q_C.

Now $O(V)_W$ is generated by the $X(A, A^{-\top})$ and the $Y(C)$ for A invertible and C alternating.

As in Section 10.7, let H_{m+1} be the tensor product $H \otimes \ldots \otimes H$, so

$$H_{m+1}(e_b) = \frac{1}{\sqrt{2^{m+1}}}\sum_{v\in V}(-1)^{b\cdot v}e_v.$$

Then $H_{m+1}X(V)H_{m+1} = Z(V)$. And $\tilde{H}_2 = I \otimes \ldots \otimes I \otimes H_2$ normalizes E and has the effect on $\overline{E}$ of interchanging x_{m+1} and z_{m+1} and fixing the other basis vectors. Now let

$$L = \langle \tilde{A},\ D(C),\ H_{m+1}, \tilde{H}_2 \ :\ A \text{ invertible, and } C \text{ alternate } (m+1)\times(m+1)\rangle$$

Then we have $L/E \simeq O(\overline{E})$ as desired.

Exercise 11.5 Verify that $L/E \simeq O(\overline{E})$.

The description of $L^\natural$ is almost the same as that of L, except that the basis vectors on which the transformations of $L^\natural$ act are the $e_{v'}$for $v' \in V'$. The difference is in the element of $U(\mathbb{C}^{N'})$ producing what we'll call $Y'(C')$ on $\overline{F}$ for C' a binary $m \times m$ symmetric matrix. Rather than using C' to define a quadratic form on V', we instead define a map $T_{C'} : \mathbb{Z}_4^m \to \mathbb{Z}_4$ as follows. Given $\widehat{v'} \in \mathbb{Z}_4^m$, choose $v' = (v_1, \ldots, v_m) \in \mathbb{Z}_2^m$ with $v' \equiv \widehat{v'} \pmod 2$, and let

$$T_{C'}(\widehat{v'}) = \sum_j C'_{jj} v_j^2 + 2 \sum_{j<k} C'_{jk} v_j v_k,$$

where the C'_{ij} are the entries of C', and the arithmetic is done mod 4. Now we define $D'(C')$ in $U(\mathbb{C}^{N'})$ by

$$D'(C')(e_{v'}) = \mathrm{i}^{T_{C'}(v')} e_{v'} \text{ for } v' \in V'.$$

Exercise 11.6 Verify that

$$D'(C')^{-1} X(a') Z(b') D'(C') = \pm X(a' + a'C') Z(b').$$

Finally, let

$$L^{\natural} = \langle \tilde{A},\ D'(C'),\ H_n \ :\ A \text{ invertible and } C' \text{ symmetric } m \times m \rangle.$$

Exercise 11.7 Verify that $L^{\natural}/F \simeq Sp(\overline{F})$.

11.2 Orthogonal spreads and binary Kerdock codes

In this section we concentrate on the orthogonal geometry of the $2(m+1)$-dimensional GF(2) space $\overline{E}$ and use it to give a definition of a binary Kerdock code of length $N = 2^{m+1}$. Recall that we assume m is odd.

By Theorem 2.3, the space $\overline{E}$ contains

$$2^m(2^{m+1} + 1) - 1 = (2^{m+1} - 1)(2^m + 1)$$

totally singular 1-spaces. Clearly, each maximal totally singular subspace of $\overline{E}$ contains $2^{m+1} - 1$ singular 1-spaces.

If a set Σ of $2^m + 1$ maximal totally singular subspaces of $\overline{E}$ partitions the set of all singular 1-spaces, we call Σ an *orthogonal spread* for $\overline{E}$.

Let A be an abelian subgroup of E so that $\overline{A}$ is a maximal totally singular subspace of $\overline{E}$. Let $\mathcal{F}(A)$ be the set of eigenspaces for A in $\mathbb{R}^N$. Recall from the preceding chapter that $\mathcal{F}(A)$ is an *orthogonal frame* for $\mathbb{R}^N$—a family of N mutually orthogonal 1-spaces. For an orthogonal spread Σ of $\overline{E}$, let

$$\mathcal{F}(\Sigma) = \cup_{\overline{A} \in \Sigma} \mathcal{F}(A),$$

a set of $(2^m + 1) \cdot 2^{m+1}$ 1-spaces of $\mathbb{R}^N$.

We defined a binary Kerdock code $\mathcal{K}(\mathcal{B})$ of length $N = 2^{m+1}$ in terms of a non-degenerate set $\mathcal{B}$ of alternate bilinear forms on a vector space V of even dimension $m+1$. We will give an alternative description of the code as a set $\mathcal{K}(\Sigma)$ of vectors in $\mathbb{Z}_2^N$ associated with an orthogonal spread Σ of $\overline{E}$ and with the choice of a distinguished element W of Σ.

First note that by replacing Σ by a suitable image under L (which is transitive on ordered pairs of such spaces since it induces $O(V)$ on $\overline{E}$), we may assume that the two maximal totally singular spaces $U = \overline{X(V)}$ and $W = \overline{Z(V)}$ are in Σ. By Proposition 9.5(a), each space $\overline{A} \in \Sigma \setminus \{W\}$ has the form $\overline{A} = UY(C)$ for a unique alternate $(m+1) \times (m+1)$ matrix C; in other words, $A = D(C)^{-1}X(V)D(C)$. ¿From Example 10.3, we know that

$$\mathcal{F}(X(V)) = \left\{ \langle \sum_{v \in V} (-1)^{b \cdot v} e_v \rangle : b \in V \right\},$$

from which it follows that

$$\mathcal{F}(A) = \left\{ \langle \sum_{v \in V} (-1)^{b \cdot v} D(C)(e_v) \rangle : b \in V \right\}.$$

Since $D(C)$ takes e_v to $(-1)^{Q_C(v)} e_v$, where Q_C is a quadratic form on V polarising to the alternate form $B_C(v,w) = vCw^\top$, we may write

$$\mathcal{F}(A) = \left\{ \langle \sum_{v \in V} (-1)^{c_v} e_v \rangle : c_v = Q_C(v) + b \cdot v,\ b \in V \right\}.$$

Somewhat abuse notation and regard $\{e_v : v \in V\}$ as a basis for $\mathbb{Z}_2^N$ as well as for $\mathbb{R}^N$.

Let Σ be an orthogonal spread of $\overline{E}$. We say the following set $\mathcal{K}(\Sigma)$ of vectors in $\mathbb{Z}_2^N$ is a *binary Kerdock code of length* $N = 2^{m+1}$:

$$\mathcal{K}(\Sigma) = \left\{ \sum_{v \in V} c_v e_v : \sum_{v \in V} (-1)^{c_v} e_v \in \mathcal{F}(\Sigma) \right\}.$$

(This notation suppresses the dependence on the distinguished element $W = \overline{Z(V)}$ in Σ.)

Allowing for multiplication by (-1), we see that for each vector $\sum_{v \in V} c_v e_v$ in $\mathcal{K}(\Sigma)$, we have $c_v = Q_C(v) + b \cdot v + \epsilon$ for fixed $b \in V$ and $\epsilon \in \mathbb{Z}_2$. Thus

$$|\mathcal{K}(\Sigma)| = (|\Sigma| - 1) \cdot |V| \cdot |\mathbb{Z}_2| = 2^m \cdot 2^{m+1} \cdot 2 = 2^{2m+2}.$$

Now we connect the two descriptions of the binary Kerdock code. Distinct elements A_1 and A_2 of $\Sigma\backslash\overline{Z(V)}$ correspond to alternating matrices C_1 and C_2 whose difference is invertible. In other words, the set of matrices C occurring in the definition of $\mathcal{K}(\Sigma)$ corresponds to a non-degenerate set of alternating bilinear forms on V of cardinality 2^m. In fact, a *Kerdock set* of matrices is a set of 2^m binary alternating $(m+1)\times(m+1)$ matrices such that the difference of any two is invertible.

The set of quadratic forms polarising to the alternating form B_C on V is given by $\{Q_C+\varphi : \varphi\in V^*\}$, for a fixed choice of Q_C, where V^* is the dual space of V. But we may take $V^*=\{\varphi_b : b\in V\}$, where $\varphi_b(v)=b\cdot v$. Finally, we identify the vector $\sum_{v\in V} c_v e_v$ with the function on V taking v to c_v to complete the equivalence of the two definitions.

11.3 Symplectic spreads and quaternary Kerdock codes

Now we turn to the symplectic geometry of $\overline{F}\simeq\mathbb{Z}_2^{2m}$ and use it to define a $\mathbb{Z}_4$ Kerdock code which, we will show in the next section, maps via the Gray map onto the binary Kerdock code of the previous section.

The space $\overline{F}$ has $2^{2m}-1=(2^m+1)(2^m-1)$ 1-spaces, each of which is totally isotropic (since $v*v=0$ for every $v\in\overline{F}$). Each maximal totally isotropic subspace of $\overline{F}$ has 2^m-1 isotropic 1-spaces.

A set Σ' of 2^m+1 maximal totally isotropic subspaces of $\overline{F}$ is a *symplectic spread* if it partitions the set of all totally isotropic 1-spaces of $\overline{F}$.

Choose an abelian subgroup $A'<F$ so that $\overline{A'}$ is maximal totally isotropic, and let $\mathcal{F}_{\mathbb{C}}(A')$ be the set of eigenspaces of A'; this set of $N'=2^m$ complex 1-spaces forms a unitary frame for $\mathbb{C}^{N'}$ — that is, it is a set of perpendicular 1-spaces with respect to the Hermitian form on $\mathbb{C}^{N'}$.

For a symplectic spread Σ', let

$$\mathcal{F}(\Sigma')=\bigcup_{\overline{A'}\in\Sigma'}\mathcal{F}(A'),$$

a set of $2^m(2^m+1)$ 1-spaces of $\mathbb{C}^{N'}$.

As in the previous section, without loss of generality, we may assume our symplectic spread Σ' contains $U'=\overline{X(V')}$ and $W'=\overline{Z(V')}$. By Proposition 9.6, each $\overline{A'}$ in $\Sigma'\setminus\{W'\}$ has the form $U'Y'(C')$ for a unique symmetric matrix C', and $A' = D'(C')^{-1}X(V')D'(C')$ gives

$$\mathcal{F}(A') = \left\{ \langle \sum_{v' \in V'} (-1)^{b' \cdot v'} D'(C')(e_{v'}) \rangle : b' \in V' \right\}.$$

Now $D'(C')$ takes $e_{v'}$ to $\mathrm{i}^{T_{C'}(v')} e_{v'}$, where

$$T_{C'}(v') = \sum_j C'_{jj} v_j^2 + 2 \sum_{j<k} C'_{jk} v_j v_k \in \mathbb{Z}_4.$$

Using $(-1)^{b' \cdot v'} = (\mathrm{i}^2)^{b' \cdot v'}$, we may write

$$\mathcal{F}(A') = \left\{ \langle \sum_{v' \in V'} \mathrm{i}^{d_{v'}} e_{v'} \rangle : d_{v'} = T_{C'}(v') + 2b' \cdot v',\, b' \in V \right\}.$$

Somewhat abuse notation and regard $\{e_{v'} : v' \in V'\}$ as a basis for $\mathbb{Z}_4^{N'}$ as well as for $\mathbb{C}^{N'}$.

Let Σ' be an symplectic spread of $\overline{F}$. Let

$$\mathbf{K}_4(\Sigma') = \left\{ \sum_{v' \in V'} d_{v'} e_{v'} : \sum_{v' \in V'} \mathrm{i}^{d_{v'}} e_{v'} \in \mathcal{F}(\Sigma') \right\}.$$

We call this set of vectors in $\mathbb{Z}_4^{N'}$ a $\mathbb{Z}_4$-*Kerdock code*; it has length $2^{N'} = (1/2)2^N$, $N' = 2^m$, m odd. (This notation suppresses the dependence on the distinguished element $W' = \overline{Z(V')}$ in Σ'.) We don't call this a quaternary code because $\mathbf{K}_4(\Sigma')$ is not always $\mathbb{Z}_4$-linear.

Allowing for multiplication by i, we see that for each vector $\sum_{v' \in V'} d_{v'} e_{v'}$ in $\mathbf{K}_4(\Sigma')$, we have $d_{v'} = T_{C'}(v') + 2b' \cdot v' + \epsilon'$ for fixed $b' \in V'$ and $\epsilon' \in \mathbb{Z}_4$. Thus

$$|\mathbf{K}_4(\Sigma')| = (|\Sigma'| - 1) \cdot |V'| \cdot |\mathbb{Z}_4| = 2^m \cdot 2^m \cdot 4 = 2^{2m+2}.$$

We would like to relate the $\mathbb{Z}_4$ code $\mathbf{K}_4(\Sigma')$ to the $\mathbb{Z}_2$ code $\mathcal{K}(\Sigma)$ of the previous section. To do this, we will work in E to define a map from symplectic spreads Σ' of $\overline{F}$ to orthogonal spreads Σ of $\overline{E}$.

Recall that ω is in E, and write $\overline{\omega}$ for the corresponding vector in $\overline{E}$ and $\langle \overline{\omega} \rangle^\perp$ for the subspace of vectors orthogonal to $\overline{\omega}$ with respect the symplectic form on $\overline{E}$. Let η be the natural map from $\overline{E}$ to $\overline{E}/\langle \overline{\omega} \rangle$. Since F is the centraliser of ω in E and $\langle \omega \rangle$ is the center of F, we can identify the $2m$-dimensional binary space $\overline{F}$ with $\eta(\langle \overline{\omega} \rangle^\perp)$.

We have symplectic bases

$$\{\overline{x_1}, \ldots, \overline{x_{m+1}}, \overline{z_1}, \ldots, \overline{z_{m+1}}\} \text{ of } \overline{E}$$

and

$$\{\overline{x_1}, \ldots, \overline{x_m}, \overline{z_1}, \ldots, \overline{z_m}\} \text{ of } \overline{F}$$

corresponding to the direct sums $\overline{E} = \overline{X(V)} \oplus \overline{Z(V)}$ and $\overline{F} = \overline{X(V')} \oplus \overline{Z(V')}$. Given a maximal totally isotropic subspace $\overline{A'}$ of $\overline{F}$, there is a unique maximal totally singular subspace $\overline{A}$ of $\overline{E}$ such that $\overline{A} \cap \overline{Z(V)} = \{0\}$ and

$$\eta(\overline{A} \cap \langle\overline{\omega}\rangle^{\perp}) = \overline{A'}.$$

Moreover, if $\overline{A'} = \overline{X(V')}Y'(C')$, then $\overline{A} = \overline{X(V)}Y(C)$, where

$$C = \begin{bmatrix} C' + d(C')^{\top} d(C') & d(C')^{\top} \\ d(C') & 0 \end{bmatrix},$$

and the $(n-1)$-tuple $d(C') = (c'_{11}, \ldots, c'_{n-1,n-1})$ consists of the diagonal entries of C'.

Now, define Σ by

$$\Sigma = \left\{\overline{Z(V)}\right\} \bigcup \left\{\overline{A} : \overline{A} \cap \overline{Z(V)} = \{0\},\ \eta(\overline{A} \cap \langle\overline{\omega}\rangle^{\perp}) = \overline{A'} \in \Sigma'\right\}.$$

We call Σ the *lift* of Σ'; note that Σ is an orthogonal spread of $\overline{E}$.

Exercise 11.8 Show that the 2^m totally singular $m+1$-spaces in Σ intersect pairwise in $\{0\}$, and therefore Σ is an orthogonal spread of $\overline{E}$.

Let us summarize where we are so far. Starting with the symplectic spread Σ' in the $2m$-dimensional binary symplectic space $\overline{F} \simeq \eta(\langle\overline{\omega}\rangle^{\perp})$ we obtain the $\mathbb{Z}_4$-Kerdock code $\mathbf{K}_4(\Sigma')$ of length 2^m. The lift of Σ' is an orthogonal spread Σ in the $2(m+1)$-dimensional binary orthogonal space $\overline{E}$, and from it we obtain the binary Kerdock code $\mathcal{K}(\Sigma)$ of length $2 \cdot 2^m$. In the next section, we show that the Gray map takes $\mathbf{K}_4(\Sigma')$ to $\mathcal{K}(\Sigma)$.

Now we quote without proof the following theorem from [5].

Theorem 11.2 *The Gray map sends* $\mathbf{K}_4(\Sigma')$ *to* $\mathcal{K}(\Sigma)$.

References

[1] M. Aschbacher, *Finite Group Theory*, Cambridge Studies in Advanced Mathematics **10**, Cambridge University Press, Cambridge, 1994.

[2] T. D. Bending, Ph.D. thesis, University of London, 1993.

[3] T. D. Bending and D. G. Fon-Der-Flaass, Crooked functions, bent functions, and distance-regular graphs, *Electronic Journal of Combinatorics* **5** (1998), #R34, 14pp. Available from `http://www.combinatorics.org/Volume_5/v5i1toc.html`.

[4] C. H. Bennett and P. W. Shor, Quantum information theory, *IEEE Transactions on Information Theory* (1998).

[5] A. R. Calderbank, P. J. Cameron, W. M. Kantor and J. J. Seidel, $\mathbb{Z}_4$-Kerdock codes, orthogonal spreads, and extremal Euclidean line systems, *Proceedings of the London Mathematical Society* (3) **75** (1997), 436–480.

[6] A. R. Calderbank, E. M. Rains, P. W. Shor and N. J. A. Sloane, Quantum error correction via codes over GF(4), *IEEE Transactions on Information Theory* **44** (1998), 1369–1387.

[7] A. R. Calderbank, E. M. Rains, P. W. Shor and N. J. A. Sloane, Quantum error correction and orthogonal geometry, *Physical Review Letters*, **78** (1997), 405–409.

[8] P. J. Cameron and J. H. van Lint, *Designs, Graphs, Codes and their Links*, London Mathematical Society Student Texts **22**, Cambridge University Press, Cambridge, 1991.

[9] P. J. Cameron and J. J. Seidel, Quadratic forms over GF(2), *Proc. Kon. Nederl. Akad. Wetensch.* (A) **76** (1973), 1–8.

[10] R. W. Carter, *Simple Groups of Lie Type*, Wiley Interscience, New York, 1972.

[11] R. Cleve and D. Gottesman, Efficient computations of encodings for quantum error correction, *Physical Review* A, **56** (1997) 76–82.

[12] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, $\mathbb{ATLAS}$ *of Finite Groups*, Oxford University Press, Oxford, 1985.

[13] A. M. Gleason, Weight polynomials of self-dual codes and the MacWilliams identities, *Actes du Congrès International de Mathématique* (vol. 3), 211–215.

[14] D. Gottesman, Class of quantum error-correcting codes saturating the quantum Hamming bound, *Physical Review* A, **54** (1996), 1862–1868.

[15] A. R. Hammons Jr., P. V. Kumar, A. R. Calderbank, N. J. A. Sloane and P. Solé, The $\mathbb{Z}_4$-linearity of Kerdock, Preparata, Goethals and related codes, *IEEE Transactions on Information Theory* **40** (1994), 301–319.

[16] W. M. Kantor, Symmetric designs, symplectic groups, and line ovals, *Journal of Algebra* **33** (1975), 43–58.

[17] A. M. Kerdock, A class of low-rate nonlinear binary codes, *Information and Control* **20** (1972), 182–187.

[18] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing Co., Amsterdam, 1977.

[19] C. Parker, E. Spence and V. D. Tonchev, Designs with the symmetric difference property on 64 points and their groups, *Journal of Combinatorial Theory* (A) **67** (1994), 23–43.

[20] H. S. Pollatsek, First cohomology groups of some linear groups over fields of characteristic two, *Illinois Journal of Mathematics* (3) **15** (1971), 393–417.

[21] D. A. Preece and P. J. Cameron, Some new fully-balanced Graeco-Latin Youden 'squares', *Utilitas Math.* **8** (1975), 193–204.

[22] J. Preskill, Lecture notes for Physics 229 at Caltech, `http://theory.caltech.edu/~preskill/ph229`.

[23] O. S. Rothaus, On "bent" functions, *Journal of Combinatorial Theory* (A) **20** (1976), 300–305.

[24] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM Journal on Computing* **26** (1997), 1484–1509.

[25] P. W. Shor, Quantum Computing, *Documenta Mathematica*, special ICM 1998 volume, I, 476–486. Available at `http://www.mathematik.uni-bielefeld.de/DMV-J/xvol-icm/00/Shor.MAN.dvi`.

[26] N. J. A. Sloane, Weight enumerators of codes, pp. 115–142 in *Combinatorics* (ed. M. Hall Jr. and J. H. van Lint), NATO ASI Series **C16**, Reidel, Dordrecht, 1975.