Some bridges between codes and designs

Peter J. Cameron
School of Mathematical Sciences
Queen Mary and Westfield College
London E1 4NS
U.K.
p.j.cameron@qmw.ac.uk

Abstract

The best-known connections between codes and designs are the constructions of various t-designs whose blocks are the sets of coordinate positions. In this paper, I will examine some rather different connections, relating codes more directly to certain kinds of experimental designs. I begin with a problem posed by a geneticist, which appears to be an extremal problem related to biplanes, but on closer examination is solved by coding theory. I turn next to the origins of coding theory: it appears that results of R. A. Fisher in 1942 and 1945 anticipate the sphere-packing bound and the construction of Hamming codes. Fisher's work was in factorial design, and leads to several questions which could be investigated. The connection between designs and codes was first noticed via projective geometry. Here too there is much subsequent work and several open problems.

1 Hamming and Assmus

In 1998, two great pioneers of coding theory, Richard Hamming and Ed Assmus, died. This section recalls briefly a small part of their contributions.

The binary Hamming code of length 7 is the null space C (over GF(2)) of the matrix

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

whose columns are the integers from 1 to 7 written in the base 2. It has 16 vectors, of which the weights (numbers of non-zero entries) are 0, 3 (seven times), 4 (seven times) and 7. The dual code (the subspace $C^\perp$ orthogonal to C) consists of the vectors of even weight in C. It has only one non-zero weight. It follows from the Assmus–Mattson theorem [1] that the supports of the words of each weight in C are the blocks of a 2-design whose points are the seven coordinate positions. The words of weight 3

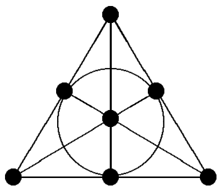

Figure 1: The Fano plane

give the familiar Fano plane, which also happens to be part of the 23ACCMCC logo (Figure 1).

The words of weight 4 form the complementary 2-$(7,4,2)$ design, whose blocks are the complements of lines in the Fano plane. A typical block is the set of positions of the Cs in the conference logo.

This connection between codes and designs is well understood, though new and interesting problems continue to arise. In the remainder of the paper, I will discuss some connections of a different sort.

2 An extremal problem

I begin with a problem, which appears to have little direct connection with codes. To begin, recall that a *biplane* is a set of subsets (blocks) of $\{1,\ldots,n\}$ such that

(a) any two blocks meet in two points;

(b) any two points lie in two blocks.

It is easy to see that, for some integer k, every block of a biplane contains k points, and every point lies in k blocks; the number of points and the number of blocks are both equal to $\binom{k}{2}+1$. Thus, a biplane is just a symmetric 2-design (SBIBD) with $\lambda = 2$. The design at the end of the last section is a biplane.

Only finitely many biplanes are known at present. The known examples have $k = 3$, 4, 5, 8, 9, 11 and 13. With current methods, there seems to be no hope of deciding whether or not infinitely many biplanes exist. In view of the difficulty of this question and the scarcity of examples, we can turn it into an extremal problem:

> What is the smallest number m of subsets (blocks) of $\{1,\ldots,n\}$ such that
>
> (a) any two blocks meet in *at most* two points;
>
> (b) any two points lie in *at least* two blocks?

It is the opposing inequalities which give this problem its particular subtlety. Note that, if we have a configuration which satisfies the conditions, then removing a point leaves one which still satisfies them; so the extremal m is a monotonic increasing function of n.

The following result gives some bounds for m.

Theorem 2.1 *Let m be the least number of subsets of $\{1, \ldots, n\}$ satisfying conditions (a) and (b) above.*

(i) $m \geq n$, with equality if and only a biplane with n points exists.

(ii) $m \leq (2 + o(1))n$.

Proof (i) Count incidences between point-pairs and block-pairs. If i is the number of such incidences, then $2\binom{n}{2} \leq i$, by (a), and $2\binom{m}{2} \geq i$, by (b); so the inequality follows. If equality holds, then both bounds are tight, so we have equality in both (a) and (b); that is, we have a biplane.

(ii) Let $n = q^2 + q + 1$, q a prime power, and let D be a planar difference set in $\mathbb{Z}/(n)$. This is a subset of $\mathbb{Z}/(n)$ of size $q + 1$, having the property that any non-zero element of $\mathbb{Z}/(n)$ has a unique representation as the difference of two elements of D. Equivalently, the translates of D are the lines of a projective plane on the point set $\mathbb{Z}/(n)$. Now it is a standard result that $-D$ (and hence any translate of $-D$) is an *oval* in this projective plane; that is, meets any line in at most two points. Moreover, $-D$ is itself a difference set, so its translates form another projective plane.

Now take all translates of D and $-D$. Any two points of $\mathbb{Z}/(n)$ lie in one translate of D and one of $-D$. The above remarks show that any two of these sets meet in at most two points.

Now the gap between consecutive primes p_n and p_{n+1} is known to be $o(p_n)$ (indeed, $O(p_n^c)$ for some $c < 1$). So by choosing q to be the smallest prime (power) such that $q^2 + q + 1 \geq n$, we obtain the stated result.

The next result shows that, if n is just a little smaller than the number of points in a biplane, then a biplane with some points removed is optimal.

Theorem 2.2 *Suppose that $k \geq 4$ and $\binom{k-1}{2} + 1 < n < \binom{k}{2} + 1$. Then the number m of sets required satisfies*

$$m \geq \min \left\{ \binom{k}{2} + 1, \left\lfloor 2\binom{n}{2} \Big/ \binom{k-1}{2} \right\rfloor + 1 \right\}.$$

Proof. Suppose that some block contains l points. Any two points of this block lie in at least one further block, and no two further blocks contain the same pair. So $m \geq \binom{l}{2} + 1$.

So we are done if some block contains k or more points. Suppose that the block sizes $l_1, \ldots, l_m$ are all smaller than k. Then $\sum_{i=1}^{m} \binom{l_i}{2}$ counts the number of incidences between a pair of points and a block. Since each point-pair lies in at least two blocks, we have

$$2\binom{n}{2} \leq \sum_{i=1}^{m} \binom{l_i}{2} \leq m\binom{k-1}{2},$$

as required.

I conclude this section with some values. All are obtained from the above theorem except for $n = 8$ and $n = 12$. In the first case, an example is obtained with ten blocks, by taking translates of the sets $\{0, 2, 4, 6\}$ and $\{0, 1, 3, 4\}$ mod 8. For $n = 12$, the non-existence of a set of 15 blocks was shown, partly with the help of a computer. I have no decent bounds for $n = 17$.

n	$3, 4$	$5 \ldots 7$	8	$9 \ldots 11$	$12 \ldots 16$
m	4	7	10	11	16

3 A search problem

The story I tell in this section is mostly true and (I think) entirely plausible.

The problem discussed in the last section was posed by an experimental scientist. When I asked about the background, I found that the 'real' problem of interest was the following:

> We are given a set of n objects, and we know that one pair is 'active'. We can test any subset of the objects; the test result will be positive if and only if the active pair is contained in the set being tested.
>
> *How many tests are required?*

The objects may, for example, be genes, and the active pair a pair of genes with the property that only in combination do they have an effect. (This was the actual context.)

The connection with the earlier problem arises from the following scheme which the scientist had developed. Choose a family of sets as in the earlier problem, and test these sets. Condition (b) guarantees at least two positive results, and condition (a) guarantees that these determine the active pair. Thus $(2 + o(1))n$ tests are required.

It is possible to do better, at least for $n \geq 8$: we can find the active pair in $2\lceil \log_2 n \rceil + 1$ tests. This is a variant of binary search. Divide the set of n objects into two nearly equal subsets A and B. Test A and B. If either test is positive, we have reduced the problem from a set of size n to a set of size $\lceil n/2 \rceil$, and we proceed inductively. If both tests give negative results, we know that the active pair has the form $\{a, b\}$, with $a \in A$, $b \in B$. We identify a by binary search (combining subsets of A with the whole of B) in $\log_2 |A|$ tests; then we identify b in the same way with $\log_2 |B|$ tests.

This is within two of the information-theoretic lower bound $\log_2 \binom{n}{2}$ for the number of binary tests required to locate one of $\binom{n}{2}$ outcomes.

However, on having this scheme explained, the scientist says that the tests have to be performed simultaneously. So the sets tested must be chosen in advance, and not depend on the results of other tests. His original scheme, though expensive, does satisfy this requirement.

It is possible to meet this requirement at the cost of doubling the number of tests. This scheme uses coding theory: here is a brief review of the concepts. See Hill [11] for more details.

We wish to send words of length n over an alphabet A with $|A| = q$ over a noisy channel where errors can occur. We *assume* that, with high probability, not too many errors occur during transmission of a word.

The strategy is to send words from a *code* C, a subset of A^n. We require:

(a) *large minimum distance* d: if $d \geq 2e + 1$, we can correct up to e errors;

(b) *many codewords* (subject to (a)): the transmission rate is $\log_q |C|/n$;

(c) *computationally efficient encoding and decoding* (subject to (a) and (b)).

It is common in coding theory for the alphabet A to be the two-element set $\{0, 1\}$, which we interpret as the binary field GF(2). More generally, we suppose that q is a prime power, and that the alphabet is the field GF(q). Then the set A^n of all words of length n is a vector space. If C is a subspace of A^n, it is called a *linear code*. Such a code has the property that its minimum distance is equal to its *minimum weight*, the least number of non-zero entries in a non-zero codeword.

A linear code C of length n and dimension r can be specified either by a *generator matrix* G (an $r \times n$ matrix whose row space is C), or by a *parity check matrix* H (an $(n - r) \times n$ matrix whose null space is C, or, more precisely, such that $C = \{x \in A^n : xH^\top = 0\}$). The generator matrix is useful for encoding: an r-tuple u of data is encoded as the codeword uG. The parity check matrix is useful for decoding. One basic technique is *syndrome decoding*. Suppose that x is transmitted and $y = x + a$ is received, where $x \in C$ and the *error pattern* a has weight at most e. The *syndrome* of y is $yH^\top$; since $xH^\top = 0$, the syndrome depends only on the error pattern. If C has minimum weight at least $2e + 1$, then distinct error patterns have distinct syndromes; so knowledge of the syndrome determines the error pattern, and hence the transmitted codeword.

Now the identification scheme for the active pair is as follows. Imagine that the zero codeword was transmitted, and two errors were made in the positions of the active pair. Using a 2-error-correcting code, we can correct the errors, i.e. identify the active pair. Let H be a $k \times n$ parity check matrix for the code. Use syndrome decoding, i.e. calculate $vH^\top$, where v is the characteristic vector of the active pair. Then v can be recovered from its syndrome.

The ith bit in the syndrome is the inner product of v with the ith row of H, i.e. it is 1 if the active pair is separated by the support of the ith row. We can determine this with two tests A_i^0 and A_i^1, where A_i^ϵ is the set of positions where ϵ occurs in the ith row. So $2k$ tests are required, the corresponding sets being the positions of zeros and ones in the rows of H.

Suppose that $n = 2^d - 1$. (For an arbitrary value of n, choose a scheme based on the smallest number above n of this form, and discard as many points as necessary.) There is a 2-error-correcting *BCH code* of length n and codimension $k = 2d$ (see Hill [11]). So we can identify the active pair in $2k = 4\lceil \log_2 n \rceil$ tests.

At this point, the scientist says that, later, it will be necessary to identify active sets of size greater than 2. Unlike the first two methods, the coding theory scheme is easily adapted. To identify an active e-tuple, just choose an e-error-correcting code. For fixed e and large n, a BCH code would be suitable; but any code whatever could be used.

Also, he says that there is a possibility that some of the test results are incorrect. Now such errors would completely invalidate the first two schemes. However, the coding theory method is easily adapted to deal with the possibility of incorrect results.

Choose a code C of length m and dimension k that will correct the maximum number of errors likely to occur in the tests. Let G be its $k \times m$ generator matrix, that is, C is the row space of G. Then use $G^\top H$ instead of H previously used. This requires $2m$ tests. Why does this work? The 'new syndrome' that we compute is $v(G^\top H)^\top = vH^\top G$ instead of $vH^\top$. But multiplication by G is the operation of encoding $vH^\top$ using the code C. This can be decoded, and errors corrected, using a decoding algorithm for this code. Then we have the correct syndrome $vH^\top$, and we can find the active pair as before.

For example, suppose that we are trying to identify an active pair from a set of size 1000. The 2-error-correcting BCH code of length 1023 has codimension 20, so that 40 tests are required.

If we suspect that some tests will give the wrong result, but (say) not more than 3% of all tests in a sequence will be wrong, we could use a 2-error-correcting shortened BCH code of length 30 and dimension 20, which will yield the required information in 60 tests, correctly if at most two test results are wrong.

4 Who discovered the Hamming codes?

In this section I hope to persuade you that error-correcting codes are almost the same thing as factorial designs in statistics, and that the Hamming codes (whose discovery is attributed to Hamming [10] and Golay [8]) were found five years earlier by R. A. Fisher in the statistical context. In what follows, I will consider fractional factorial designs, which are conceptually simpler than factorial block designs (which Fisher actually wrote about); there is no essential difference and it is conceptually simpler to do so.

We are investigating n factors which can affect the yield of some process. The ith factor can take any one of a set A_i of levels, with $|A_i| = q_i$.

We assume that only the interactions of small numbers of factors affect the yield significantly.

We impose the structure of an abelian group on A_i, and test treatment combinations lying in a subgroup B of $A_1 \times \cdots \times A_n$.

Let C be the annihilator of B in $A_1^* \times \cdots \times A_n^*$. (Here A_i^* is the group of characters of A_i; so C is the set of all characters of $A_i \times \ldots \times A_n$ which are trivial on B.)

Elements of C represent combinations of treatments which are confounded in the experiment. (For example, if an element of C has support in $A_i^* \cup A_j^* \cup A_k^*$, then the interaction of factors j and k cannot be distinguished from the main effect of factor i.)

We want

(a) *Large weight in C* so that potentially significant combinations of factors are not confounded;

(b) *Few trials* (subject to (a)): trials are expensive! This means small B, and so large C: note that

$$|C| = \frac{q_1 \cdots q_n}{|B|}.$$

(c) *simple description* which can be explained to experimenters and for which results can be analysed (subject to (a) and (b)).

Comparing these requirements with those for a code, we see very strong similarities! Design theorists and coding theorists are both looking for subsets C of $A_1 \times \cdots \times A_n$ with large minimum distance and large cardinality.

Coding theorists have n large, all A_i of the same size (almost always 2), and don't insist on group structure (though it does help to use a linear code).

Statisticians have n fairly small, varying alphabet size, and do require group structure.

Hamming codes satisfy both specifications!

Let $V = \mathrm{GF}(q)^k$. Partition the non-zero vectors in V into equivalence classes, where two vectors are equivalent if one is a non-zero scalar multiple of the other. There are $(q^k - 1)/(q - 1)$ equivalence classes.

Choose one vector from each equivalence class, and let H be the matrix of size $k \times (q^k - 1)/(q - 1)$ having these vectors as columns. (For simplicity, take all vectors whose first non-zero entry is 1.) Then any two columns of H are linearly independent.

The code C with parity check matrix H thus has minimum weight 3 and so is 1-error-correcting. This is the *Hamming code $H(k, q)$*.

In Finney's book [5] on experimental design, the following appears:

> Fisher (1942) proved that:
>
> > A 2^n factorial scheme can be arranged in 2^{n-p} blocks of 2^p plots each, without confounding either main effects or 2-factor interactions, provided that $n < 2^p$.
>
> Subsequently (1945), he generalized this theorem and proved that:
>
> > A π^n factorial scheme can be arranged in π^{n-p} blocks of π^p plots each, without confounding either main effects or 2-factor interactions, provided that
> >
> > $$n \leq (\pi^p - 1)/(\pi - 1).$$

(Here π is a prime power.) What Fisher does in these papers is to construct the Hamming codes, as a glance at the actual papers makes clear. The constructions by Hamming and Golay appeared several years later.

5 Codes over mixed alphabets

The comparison of codes and designs in the last section suggests looking again at coding theory in the situation where the alphabet size is permitted to vary from one coordinate to another.

The standard bounds of elementary coding theory generalise immediately to this situation. The proofs require only minimal adjustment to those for constant alphabet size (see Hill [11]).

Theorem 5.1 *Let C be a code of length n over alphabets of sizes $q_1, \ldots, q_n$, where $q_1 \leq \cdots \leq q_n$. Suppose that C has minimum distance d, and let $e = \lfloor (d-1)/2 \rfloor$.*

(a) Sphere-packing bound:

$$|C| \leq \frac{\displaystyle\prod_{i=1}^{n} q_i}{\displaystyle\sum_{k=0}^{e} \sum_{i_1 < \cdots < i_k} \prod_{j=1}^{k} (q_{i_j} - 1)}.$$

(b) Singleton bound:

$$|C| \leq \prod_{i=1}^{n-d+1} q_i.$$

(c) Plotkin bound: *Let*

$$\alpha = \sum_{i=1}^{n} (1 - 1/q_i).$$

If $d > \alpha$ then $|C| \leq d/(d-\alpha)$.

For example, what is the best 1-error-correcting code of length 5 over alphabets of size $2, 2, 2, 2, 4$?

- The sphere-packing bound shows that

$$|C| \leq \frac{2 \cdot 2 \cdot 2 \cdot 2 \cdot 4}{1+1+1+1+1+3} = 8.$$

- The Singleton bound gives

$$|C| \leq 2 \cdot 2 \cdot 2 = 8.$$

- The Plotkin bound:

$$\alpha = \tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{3}{4} = \tfrac{11}{4} < 3,$$

  so $|C| \leq 3/(3 - \frac{11}{4}) = 12$.

We conclude that $|C| \leq 8$.

There is a code of length 8, which is a subgroup of $A_1 \times \cdots \times A_5$, if we take $A_1 = A_2 = A_3 = A_4 = \{0, 1\}$ (the cyclic group of order 2) and $A_5 = \{0, a, b, c\}$ with $a + b + c = 0$ (the Klein group of order 4). Let C consist of the following codewords:

$$\begin{array}{ccccc}
0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & a \\
1 & 1 & 0 & 0 & a \\
0 & 1 & 0 & 1 & b \\
1 & 0 & 1 & 0 & b \\
0 & 1 & 1 & 0 & c \\
1 & 0 & 0 & 1 & c
\end{array}$$

It is easily checked that C is a subgroup and has minimum weight 3. So it meets the sphere-packing and Singleton bounds.

This example has a three-parameter generalisation:

Theorem 5.2 *For every $a < b$ and prime power q, there is a perfect code of length $(q^b - q^a)/(q-1) + 1$ over alphabets of sizes q ($(q^b - q^a)/(q-1)$ times) and q^a (once).*

Such a code should have size

$$\frac{q^{(q^b-q^a)/(q-1)} \cdot q^a}{1 + (q-1) \cdot (q^b - q^a)/(q-1) + q^a - 1} = q^{(q^b-q^a)/(q-1)-b+a}.$$

Let C be the binary Hamming code of length $(q^b - 1)/(q-1)$. Its coordinates are indexed by the points of the $(b-1)$-dimensional projective space over $\mathrm{GF}(q)$. Take an $(a-1)$-dimensional subspace S of the $(b-1)$-dimensional vector space. Let C_1 and C_2 be the codes induced on S and its complement, and D_1 and D_2 the subcodes of C consisting of words vanishing on the complement of S and on S, respectively. Then $C_1/D_1 \cong C_2/D_2$, and $C_2 \cong C/D_1$.

Since any point of S is its intersection with a line, and any three collinear points support a word of the Hamming code, C_1 consists of all words, with dimension $(q^a - 1)/(q-1)$. The subcode D_1 is the first-order Hamming code of length $(q^a-1)/(q-1)$, with codimension a in C_1. So D_2 has codimension a in C_2.

Note also that C_2 has no words of weight 1. For, since the Hamming code has covering radius 1, a word of weight 1 would lead to a word of weight 2 in C. Clearly D_2 has minimum weight at least 3.

Let A be an elementary abelian group of order 2^a, and let θ be a homomorphism from C_2 to A with kernel D_2. Now the required code is $\{(v, v\theta) : v \in C_2\}$.

It has minimum weight at least 3. For if $v \in D_2$ then v has weight at least 3, while if $v \notin D_2$ then v has weight at least 2 and $v\theta \neq 0$. Moreover, the code has the same cardinality as C_2, and

$$\dim(C_2) = \dim(C) - \dim(D_1) = \left(\frac{q^b - 1}{q-1} - b\right) - \left(\frac{q^a - 1}{q-1} - a\right),$$

as required.

Note that, in the case when $b = a - 1$, this code also meets the Singleton bound.

6 Codes and projective geometry

One of the first people to notice the connection between codes and factorial designs was R. C. Bose. He related both of these concepts to sets of points in projective spaces. His early work such as [2] included factorial design, but from the early 1960s he was interested in codes as well (for example, [3]). The basic idea is simple.

Let A be a $k \times n$ matrix over $\mathrm{GF}(q)$. Assume that no two columns are linearly dependent, and that A has rank k.

(a) A is the parity check matrix of a $[n, n-k]$ code

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

Elementary row operations don't affect C; column permutations and scalar multiplications replace it by an equivalent code (metric properties are unaffected). The code C has minimum weight at least 3, so is 1-error-correcting. The corresponding factorial design has q^k treatments.

(b) The columns of A can be regarded as a set S of n points in the projective space $\mathrm{PG}(k-1, q)$. Elementary row operations induce collineations of the projective space, while column permutations don't change S. The set S spans $\mathrm{PG}(k-1, q)$.

So 1-error-correcting codes (up to equivalence) correspond naturally to spanning subsets of projective space (up to collineations).

The correspondence between codes and projective spaces allows many properties to be transferred back and forth:

(a) The Hamming codes correspond to the entire projective space. The connection between codes and projective spaces can be regarded as a generalisation of the construction of Hamming codes.

(b) Supports of words of the dual code correspond to complements of hyperplane sections of S.

(c) (Bose [2]) MDS codes (those which meet the Singleton bound) correspond to arcs in projective space. (This, and a bound on the size of arcs in projective planes, are in Bose's paper on factorial designs.) This has inspired a great deal of research: see, for example, Bruen *et al.* [4].

(d) (Greene [9]) The weight enumerator of the code is a specialisation of the Tutte polynomial of the matroid represented by the matrix. Hence the MacWilliams identities follow from matroid duality.

The last item is perhaps the most interesting. The Tutte polynomial is a 2-variable polynomial associated with a matroid. If the matroid comes from a graph, then specialisations of the Tutte polynomial count colourings, acyclic orientations, and evaluate knot polynomials and percolation probabilities. As noted, if the matroid comes from a code, then a specialisation gives the weight enumerator of the code.

Presumably the Tutte polynomial carries other important information about the code or the geometric configuration. This is a topic well worth investigating. More about the Tutte polynomial and its evaluation can be found in Welsh [12], and an exposition of the connection between matroids and codes (including a proof of Greene's theorem) is available on the Web, at the address

`http://www.maths.qmw.ac.uk/~pjc/comb/matroid.dvi`

Acknowledgment

I am grateful to Rosemary Bailey and Brendan McKay for their invaluable suggestions which have been incorporated into this paper, and to Gregory Gutin for communicating to me the problem discussed in the second section.

References

[1] E. F. Asmus, Jr. and H. F. Mattson, Jr., New 5-designs, *J. Combinatorial Theory* (A) **6** (1969), 122–151.

[2] R. C. Bose, Mathematical theory of the symmetrical factorial design, *Sankhyā* **8** (1947), 107–166.

[3] R. C. Bose and J. N. Srivastava, On a bound useful in the theory of factorial design and error-correcting codes, *Ann. Math. Statist.* **35** (1964), 408–414.

[4] A. A. Bruen, J. A. Thas and A. Blokhuis, On MDS codes, arcs in PG(n, q) with q even, and a solution of three fundamental problems of B. Segre, *Invent. Math.* **92** (1988), 441–459.

[5] D. J. Finney, *An Introduction to the Theory of Experimental Design*, University of Chicago Press, Chicago, 1960.

[6] R. A. Fisher, The theory of confounding in factorial experiments in relation to the theory of groups, *Ann. Eugenics* **11** (1942), 341–353.

[7] R. A. Fisher, A system of confounding for factors with more than two alternatives, giving completely orthogonal cubes and higher powers, *Ann. Eugenics* **12** (1945), 2283–290.

[8] M. J. E. Golay, Notes on digital coding, *Proc. IEEE* **37** (1949), 657.

[9] C. Greene, Weight enumeration and the geometry of linear codes, *Studies in Applied Math.* **55** (1976), 119–128.

[10] R. W. Hamming, Error detecting and error correcting codes, *Bell Systems Tech. J.* **29** (1950), 147–160.

[11] R. Hill, *A First Course in Coding Theory*, Clarendon Press, Oxford, 1986.

[12] D. J. A. Welsh, *Complexity : Knots, Colourings and Counting*, London Mathematical Society Lecture Note Series **186**, Cambridge University Press, Cambridge, 1993.