On a Class of Constant Weight Codes

Mihai Caragiu

Institute of Mathematics Bucharest
and
Department of Mathematics, Pennsylvania State University

E-mail: caragiu@math.psu.edu

Submitted: July 31, 1995; Accepted: January 2, 1996

Abstract. For any odd prime power q we first construct a certain non-linear binary code $C(q,2)$ having $(q^2-q)/2$ codewords of length q and weight $(q-1)/2$ each, for which the Hamming distance between any two distinct codewords is in the range $[q/2-3\sqrt{q}/2,\ q/2+3\sqrt{q}/2]$ that is, 'almost constant'. Moreover, we prove that $C(q,2)$ is distance-invariant. Several variations and improvements on this theme are then pursued. Thus, we produce other classes of binary codes $C(q,n)$, $n \geq 3$, of length q that have 'almost constant' weights and distances, and which, for fixed n and big q, have asymptotically q^n/n codewords. Then we prove the possibility of extending our codes by adding the complements of their codewords. Also, by using results on Artin $L-$series, it is shown that the distribution of the 0's and 1's in the codewords we constructed is quasi-random. Our construction uses character sums associated with the quadratic character χ of $\mathbf{F}_{q^n}$ in which the range of summation is $\mathbf{F}_q$. Relations with the duals of the double error correcting BCH codes and the duals of the Melas codes are also discussed.

1991 AMS Subject Classification :
Primary 11T71
Secondary 11T23, 94B27

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

1. Introduction

In the present paper we shall first construct, for any odd prime power q, a nonlinear constant weight code $C(q,2)$ with $(q^2-q)/2$ codewords, with the property that each nonzero distance lies in the interval

$$\left[\frac{q}{2}-\frac{3}{2}\sqrt{q}\ ,\ \frac{q}{2}+\frac{3}{2}\sqrt{q}\right]$$

In constructing such codes we shall use character sums associated with the quadratic character χ of $\mathbf{F}_{q^2}$, in which the range of summation is $\mathbf{F}_q$. Sums of this type were considered, for example, by Davenport [5]. He shows, for example, that if θ is any element generating the finite field $\mathbf{F}_{p^k}$ over its prime subfield $\mathbf{F}_p$ and if χ is the quadratic character of $\mathbf{F}_{p^k}$, then

$$\sum_{t=0}^{p-1}\chi(\theta+t)=O\left(p^{\frac{2k+1}{2k+2}}\right)$$

In fact, Weil's theorem shows that the right-hand side of the above estimate can be sharpened to $O(\sqrt{p})$. For references on Weil theorem and related topics (including algebraic geometric codes), one may consult [2], [5], [6], [8], [10], [12], [13], [15]. Other authors have considered as well combinatorial consequences of various results concerning the distribution of the values taken by a multiplicative character of a finite field on a coset of a certain subfield. See, for example, [3]. In the third section of the paper we provide an extension of the basic construction, the result of which will be, for any $n>2$, a class of codes $C(q,n)$ with similar properties as $C(q,2)$, but only with an 'almost' constant weight for their codewords.

Note that whenever we take off the first row and the first column of a normalized Hadamard matrix of order $4t$, the set of all the rows of the remaining matrix can be seen (by replacing each occurrence of a -1 with 0) as a nonlinear code of length $n=4t-1$ having a constant weight $[n/2]=2t-1$, for which the distance between two distinct codewords is $d=2t$. It is well known [1], [9] that the case $4t-1=q$ a prime power will do the job, and thus in this case one can find nonlinear codes of length q, constant weight $(q-1)/2$ and constant distance $(q-1)/2$, having q codewords. A natural question will be, then, what will happen would we give up the requirement for having a constant distance, by permitting a 'small variation' of the parameter d, while keeping a constant weight, say $[n/2]$, for the codewords. Our study of the codes $C(q,2)$ provides a partial answer to this in a special case. Thus, whenever q is an odd prime power, we obtain the lower bound $(q^2-q)/2$ for the maximum number of codewords in a code of length q, constant weight $(q-1)/2$ and nonzero distances within the range $[q/2-3\sqrt{q}/2,\ \ q/2+3\sqrt{q}/2]$. In particular, $A(q,q/2-3\sqrt{q}/2)\geq(q^2-q)/2$, where $A(n,d)$ is the maximum number of binary codewords of length n and minimum distance d. One might want to compare this with the Plotkin upper bound $A(4t, 2t) \leq 8t$, which is attained whenever a Hadamard matrix of order $4t$ exists.

Thinking probabilistically, one could see a codeword in $C(q, n)$ as a 'random subset' of $\mathbf{F}_q$ or, equally, as the output of an experiment of randomly and independently selecting elements of $\mathbf{F}_q$, the probability of choosing a particular one being $1/2 + O(1/\sqrt{q})$, the implied constant depending only on n. Any two such experiments are 'almost independent' , in the sense that the probability of a given element of $\mathbf{F}_q$ to be selected by each of the two such fixed experiments is in the range $1/4 + O(1/\sqrt{q})$. If we consider $C(q, 2)$, we see that in fact we get an explicit example of $(q^2 - q)/2$ 'almost independent' random subsets of $\mathbf{F}_q$, while for fixed n and big q the number of codewords in $C(q, n)$ grows asymptotically like q^n/n. One can further improve by adding the complementary codewords. All these facts might be useful in statistics.

In the fourth section of the paper we shall prove the 'quasi-random' character [4] of the distribution of the 0's and 1's in the codewords of the constructed binary codes, by making use of exponential sums estimates coming from classical results on Artin $L-$series. Also, we shall prove that the codes $C(q, 2)$, although nonlinear, are distance invariant.

In the last section we will consider first the problem of extending the codes $C(q, 2)$ and $C(q, n)$ by adding the complementary codewords. Then we will establish a connection with the binary codes belonging to two known classes, namely that of the duals of the double error correcting BCH codes, and that of the duals of the Melas codes.

2. The basic construction

Let q be an odd prime power. We may choose $\mathbf{j}$ in $\mathbf{F}_{q^2}$ with $\mathbf{F}_{q^2} = \mathbf{F}_q(\mathbf{j})$ and a minimal equation over $\mathbf{F}_q$ of the form $\mathbf{j}^2 = s$, where $s \in \mathbf{F}_q^* - (\mathbf{F}_q^*)^2$. Let $\chi : \mathbf{F}_{q^2}^* \to \{-1, 1\}$ be the quadratic (Legendre) character. Obviously, the restriction of χ to $\mathbf{F}_q^*$ is trivial, every element of $\mathbf{F}_q$ being a square in $\mathbf{F}_{q^2}$. To every element $x \in \mathbf{F}_{q^2} - \mathbf{F}_q$ we associate a $0-1$ vector V_x indexed by the elements of $\mathbf{F}_q$: namely we will define

$$V_x(t) := \frac{1}{2}\left(1 + \chi(x + t)\right) \tag{1}$$

That is, $V_x(t)$ is 1 if $x + t$ is a square and 0 elsewhere. We have defined, in fact, a binary code of length q, which we will denote by $C(q, 2)$. Natural questions arise consequently. How many distinct codewords do appear in this way ? What can we say about their weights ? How can we estimate the Hamming distance between two codewords ? We will show how all the above questions can be pretty fairly answered provided we use the relation (2) below expressing the Hamming distance

$d(V_x, V_y)$ between the codewords V_x and V_y as a character sum. First, let us note the (obvious) fact that

$$d(V_x, V_y) = \frac{1}{2} \sum_{t \in \mathbf{F}_q} |\chi(x+t) - \chi(y+t)|$$

As $|a-b| = 1 - ab$ for every $a, b \in \{-1, 1\}$, one easily finds out that

$$d(x,y) = \frac{1}{2} \left(q - \sum_{t \in \mathbf{F}_q} \chi[(x+t)(y+t)] \right) \tag{2}$$

We need an explicit condition under which $d(V_x, V_y) = 0$. This will be provided by the next proposition.

PROPOSITION 1. For every $x, y \in \mathbf{F}_{q^2} - \mathbf{F}_q$, $d(V_x, V_y) = 0$ if and only if $y = \overline{x}$ or $y = x$.

PROOF. We agree to denote the Frobenius action by by $\overline{z} := a - b\mathbf{j} = z^q$ for every $z = a + b\mathbf{j} \in \mathbf{F}_{q^2} - \mathbf{F}_q$. Then, it is easy to see that for every such z, one has $d(V_z, V_{\overline{z}}) = 0$. We need now to prove the converse. Let us denote by ψ the quadratic character of $\mathbf{F}_q$. It is a well known fact that the relation between ψ and its canonical 'lifting' χ is given by

$$\chi(z) = \psi(Nz) \tag{3}$$

for every $z \in \mathbf{F}_{q^2}^*$, where $Nz = z\overline{z} = z^{1+q}$ is the usual norm map from $\mathbf{F}_{q^2}$ to $\mathbf{F}_q$ Let $x, y \in \mathbf{F}_{q^2} - \mathbf{F}_q$ two distinct elements. Suppose that the relation

$$\chi(x+t) = \chi(y+t)$$

holds for every element t of $\mathbf{F}_q$. Eventually we have to prove that x and y are Frobenius conjugate. By using (3), we can rewrite this as

$$\psi((x+t)(\overline{x}+t)) = \psi((y+t)(\overline{y}+t))$$

or, equivalently,

$$\psi[(x+t)(\overline{x}+t)(y+t)(\overline{y}+t)] = 1$$

for any t in the base field. We now recall the celebrated 'Riemann Hypothesis' for algebraic curves over finite fields, first proved by Hasse [7] for elliptic curves, then, in the general case, by Weil [15]. Thus, the number N of $\mathbf{F}_q-$rational points on a genus 1 curve defined over $\mathbf{F}_q$ satisfies the inequality

$$|N - (q+1)| \leq 2\sqrt{q}$$

Let us return now to our proof. From our assumptions it follows that the polynomial

$$P(X) = (X+x)(X+\overline{x})(X+y)(X+\overline{y})$$

is separable (i.e., it has distinct roots). Moreover, we assumed that $P(t)$ is a square in $\mathbf{F}_q^*$ for every t in $\mathbf{F}_q$. In other words the genus 1 curve defined over $\mathbf{F}_q$ by the equation

$$Y^2 = P(X) \tag{4}$$

has $2q$ finite $\mathbf{F}_q$−rational points. One can view geometrically the equation (4) as a two-sheeted covering of $\mathbf{P}^1$, ramified in four finite places, corresponding to the 4 linear factors of $P(X)$. The place at infinity of $\mathbf{P}^1$ is not ramified, so our curve (4) has two more rational points 'at infinity', adding up to a total of $N = 2q + 2$ $\mathbf{F}_q$−rational points. Now, we only have to apply the above stated Hasse−Weil theorem implying in this special case that $q + 1 \leq 2\sqrt{q}$, or $q = 1$, an obvious contradiction. This concludes the proof. □

COROLLARY 2. $C(q, 2)$ has $(q^2 - q)/2$ codewords. □

Next we will prove that the codewords of $C(q, 2)$ have constant weights.

PROPOSITION 3. The weight of each codeword in $C(q, 2)$ is $(q - 1)/2$.

PROOF. The weight $wt(V_x)$ of V_x can be expressed as

$$wt(V_x) = \frac{1}{2}\left(q + \sum_{t \in \mathbf{F}_q} \chi(x + t)\right) =$$

$$= \frac{1}{2}\left(q + \sum_{t \in \mathbf{F}_q} \psi[(x + t)(\overline{x} + t)]\right)$$

Taking into account the well known exact estimates of the complete character sums with quadratic polynomial argument [8] the result follows at once. □

We will now prove how the Weil estimates for character sums with polynomial argument (see [8], chapter 5, theorem 5.41) imply that the Hamming distance between two distinct codewords of $C(q, 2)$ is, as announced, 'almost' constant .

PROPOSITION 4. The Hamming distance between two distinct codewords V_x and V_y of $C(q, 2)$ lies in the interval

$$\left[\frac{q}{2} - \frac{3}{2}\sqrt{q}\ ,\ \frac{q}{2} + \frac{3}{2}\sqrt{q}\right]$$

PROOF. One can write

$$d(V_x, V_y) = \frac{1}{2}\left(q - \sum_{t \in \mathbf{F}_q} \psi(P(t))\right)$$

where $P(X) = (X + x)(X + \overline{x})(X + y)(X + \overline{y})$ is a polynomial in $\mathbf{F}_q[X]$ which factors over $\mathbf{F}_q$ as a product of two distinct monic irreducible polynomials. The number of its distinct roots is $d = 4$ and, by Weil's theorem we get

$$\left| d(V_x, V_y) - \frac{q}{2} \right| \leq \frac{3}{2}\sqrt{q}$$

This concludes the proof. $\square$

NOTE. We certainly can define, in fact, a $0 - 1$ vector V_x for any element $x \in \mathbf{F}_{q^2}$. Provided we agree that $\chi(0) := 1$ (fact which we tacitly assume in the next section), it becomes clear that for any $x \in \mathbf{F}_q$ the associated vector V_x is the constant vector whose all components are 1. We avoided to do this as we planned to provide an example of a constant weight code. However, defining a V_x for every x will prove to be fruitful in the next paragraph, when we shall generalize the codes $C(q, 2)$.

3. Higher dimensional analogues

We now try to define higher dimensional analogues $C(q, n)$ of the codes $C(q, 2)$. The idea is as follows: instead of working with a quadratic extension of finite fields we shall choose to adapt the previous construction to an extension of arbitrary degree $\mathbf{F}_{q^n}/\mathbf{F}_q$. Thus, we will be able to construct for every $n \geq 2$ and each odd prime power q a nonlinear code $C(q, n)$. Unfortunately, if $n > 2$, $C(q, n)$ will prove to be only an 'almost' constant weight code. Let χ be now the quadratic character of $\mathbf{F}_{q^n}$ $(n > 2)$ and x be an element of $\mathbf{F}_{q^n}$. One may use the same relation (1) in order to define a $0 - 1$ vector V_x indexed by the elements of $\mathbf{F}_q$. The Hamming distance between two such vectors has exactly the same formal expression (2). We easily check that $d(x, \overline{x}) = 0$ where $\overline{x} = x^q$ represents the Frobenius action. Thus the vectors V_x are the same along any Frobenius orbit. The basic problem is whether we have any other identifications. Notice that a relation similar to (3) holds here, the only difference being that the norm is given now by $N(z) = z^{1+q+q^2+\ldots+q^{n-1}}$ for every z in $\mathbf{F}_{q^n}$.

Let $x \in \mathbf{F}_{q^n}$. Then we have the obvious polynomial identity:

$$N(X + x) = P(X)^{n/e} \tag{5}$$

where $P(X)$ is the minimal polynomial of $-x$ over $\mathbf{F}_q$, e is its degree, and

$$N(X + x) = (X + x)(X + x^q)(X + x^{q^2})\ldots(X + x^{q^{n-1}})$$

is the characteristic polynomial of $-x$ over $\mathbf{F}_q$.

Now, if $x, y \in \mathbf{F}_{q^n}$, $P(X), Q(X) \in \mathbf{F}_q[X]$ are the minimal polynomials over $\mathbf{F}_q$ of $-x, -y$, respectively, with the corresponding degrees e and g, say, then one can write down the Hamming distance $d(V_x, V_y)$, by using (5), as follows:

$$d(V_x, V_y) = \frac{1}{2}\left(q - \sum_{t \in \mathbf{F}_q} \psi[P(t)^{n/e} Q(t)^{n/g}]\right) \tag{6}$$

Here ψ has the same meaning as before: it represents the quadratic character of $\mathbf{F}_q$, whose lifting to $\mathbf{F}_{q^n}$ is χ.

PROPOSITION 5. V_x is a vector with all the components 1 whenever n/e is even, where e represents the degree of the minimal polynomial of x over $\mathbf{F}_q$.

PROOF. The weight of V_x will be given by

$$wt(V_x) = \frac{1}{2}\left(q + \sum_{t \in \mathbf{F}_q} \psi[N(x+t)]\right) =$$

$$= \frac{1}{2}\left(q + \sum_{t \in \mathbf{F}_q} \psi[P(t)^{n/e}]\right) \tag{7}$$

where $P(X) \in \mathbf{F}_q[X]$ is the minimal polynomial (of degree e) of $-x$ over $\mathbf{F}_q$. Thus, whenever n/e is even, the corresponding V_x is is the constant 1 vector. An alternative but more elementary solution runs as follows. As $n/e = [\mathbf{F}_{q^n} : \mathbf{F}_q(x)]$, we see that whenever n/e is even all the elements having the form $x+t$ for some t in $\mathbf{F}_q$ belong to a field $\mathbf{F}_q(x)$ for which $\mathbf{F}_{q^n}$ is an extension of even degree, and consequently they are squares in $\mathbf{F}_{q^n}$. □

The following question pops up naturally: are there any other situations (besides the ones described above) in which two such binary vectors V_x and V_y coincide ?

Indeed , let us suppose that x and y represent two different Frobenius orbits, and that n/e and n/g are not both even. Then $-x$, $-y$ are also in distinct Frobenius orbits, their minimal polynomials, $P(X)$ and $Q(X)$ respectively are distinct, and consequently the polynomial

$$H(X) = P(X)^{n/e} Q(X)^{n/g}$$

has $e+g$ distinct roots. Also it is easy to see that $H(X)$ is not, in this case, a square of some other polynomial. All we need to is to apply now the Weil estimates. By using them we see that

$$\left| \sum_{t \in \mathbf{F}_q} \psi(P(t)^{n/e} Q(t)^{n/g}) \right| \leq (e+g-1)\sqrt{q} \tag{8}$$

Because obviously $e, g \leq n$, we find, from (8):

$$\left| \sum_{t \in \mathbf{F}_q} \psi(P(t)^{n/e} Q(t)^{n/g}) \right| \leq (2n-1)\sqrt{q}$$

It is now clear that the $0-1$ sequences corresponding to the Frobenius orbits through x and y are distinct provided that $q > (2n-1)^2$. More generally, the $0-1$ vectors associated to distinct Frobenius orbits of cardinalities e and g, respectively (certainly e and g are divisors of n), at least one of the numbers $n/e,\ n/g$ being odd, are distinct as long as $q > (e+g-1)^2$. Under the condition $q > (2n-1)^2$, the set of all $0-1$ words having the form V_x for some $x \in \mathbf{F}_{q^n}$ and which are not constant 1 vectors will form a nonlinear code which we will denote by $C(q,n)$. These represent the obvious generalization of the codes $C(q,2)$ introduced in the previous section. We are naturally led to the following theorem.

THEOREM 6. If $q > (2n-1)^2$, a $0-1$ vector V_x has all the components equal to 1 if and only if $[\mathbf{F}_{q^n} : \mathbf{F}_q(x)]$ is even. The Hamming distances between distinct codewords of $C(q,n)$ are of the form $q/2 + O(\sqrt{q})$. The weight of any non-constant codeword V_x is 'almost' constant, being on the form $q/2 + O(\sqrt{q})$. All the implied constants depend only on n. □

If, for example, n is odd and $q > (2n-1)^2$ then the number of codewords in $C(q,n)$ coincides with the number of all Frobenius orbits of $\mathbf{F}_{q^n}/\mathbf{F}_q$. At the other extreme, let us consider the case of $2-$extensions, that is the case in which n is a power of 2, so let $n = 2^k$ and $q > (2n-1)^2$. Then any two Frobenius orbits which are both non-maximal (i.e., this is the case when both of them have less than 2^k elements) give rise to the same codeword of $C(q,n)$. More generally, under the assumptions of the previous theorem, the number of codewords in $C(q,n)$ equals the number of those Frobenius orbits in $\mathbf{F}_{q^n}/\mathbf{F}_q$ whose 'co-cardinality' n/e is odd.

NOTE. We have seen that under the restrictive condition

$$q > (2n-1)^2 \tag{9}$$

a $0-1$ vector V_x has all the components 1 if and only if $[\mathbf{F}_{q^n} : \mathbf{F}_q(x)]$ is even. The 'if' part doesn't require any condition while the converse holds under the assumption (9). Can we drop (9) completely ? We shall show by an example that this cannot be done in general. Indeed, let us consider a fixed prime power q, while n will be chosen to be odd. If n is big enough, one can find an element x for which the corresponding V_x has all the components equal to 1. Indeed let M be the number of the elements $x \in \mathbf{F}_{q^n}$ for which the quadratic character χ takes the value 1 on each element of the form $x+t$ with t in $\mathbf{F}_q$. There is a classical result on the distribution of quadratic residues in finite fields [12], to the effect that, given $\epsilon_1, \epsilon_2, \ldots \epsilon_n$ in $\{-1, 1\}$, and n distinct field elements $a_1, a_2, \ldots, a_n$, then the number $N(\epsilon_1, \epsilon_2, \ldots \epsilon_n)$ of elements x in $\mathbf{F}_q$ (q odd) having the property that

$$\chi(x + a_i) = \epsilon_i$$

for any $i = 1, 2, ..., n$ is estimated as

$$N(\epsilon_1, \epsilon_2, ...\epsilon_n) = \frac{q}{2^n} + O\left(n\sqrt{q}\right)$$

where the implied constant is absolute. Thus, M is given by a formula of the type

$$M = \frac{q^n}{2^q} + O(q^{n/2+1})$$

For some big enough n, M will be nonzero, and consequently one could find an x for which V_x is a constant 1 vector.

4. Quasi-randomness and distance-regularity

We refer here to the the paper [4] in which the concept of quasi-randomness is discussed in connection with the residue class rings $\mathbf{Z}_n$. There the authors provide a list of ten equivalent definitions for what are called 'quasi-random subsets of $\mathbf{Z}_n$'. Here we shall use their exponential sum characterization. Namely, suppose we are able to define, for every n belonging to an infinite set of positive integers, a certain subset $S_n \subset \mathbf{Z}_n$. We shall say that this produces a sequence quasi-random subsets within the respective residue class rings if for any $j \neq 0$ in $\mathbf{Z}_n$ we have the estimate

$$\sum_{x \in \mathbf{S}_n} \exp\left(2\pi ijx/n\right) = o(n)$$

As a nice example, it is proved [4] by a Gaussian sum argument that the perfect squares within the finite prime fields form quasi-random subsets.

Obviously, the above definition has a formal analogue for finite fields. Thus, if we are able to define, for every q belonging to an infinite set of prime powers, a certain subset $S_q \subset \mathbf{F}_q$, we shall agree to say that the subsets we define are quasi-random within the respective finite fields if, in whatever way we choose nontrivial additive characters ω of the corresponding finite fields, the following estimate holds:

$$\sum_{x \in \mathbf{S}_n} \omega(x) = o(q)$$

Let's now go back to our codes. We can associate to any codeword V_x in $C(q, n)$ a certain subset $S(q; x)$ of $\mathbf{F}_q$ in a very simple way: an element t will be in $S(q; x)$ whenever $x + t$ is a square in $\mathbf{F}_{q^n}$, that is, whenever the codeword V_x has an 1 on the position indexed by the element t. In what follows the parameter n will be considered to be fixed. We shall prove that the subsets defined above are, in the sense we agreed on above, quasi-random. In order to do so we use traditional results on Artin $L-$series in order to estimate exponential sums of the type

$$\sum_{t\in S(q;x)} \omega(t) \tag{10}$$

where ω are nontrivial additive characters of the finite fields in case. Indeed one obviously has the following estimates:

$$\sum_{t\in S(q;x)} \omega(t) = \frac{1}{2}\sum_{t\in \mathbf{F}_q} [1+\chi(x+t)]\,\omega(t) + O(1) =$$

$$= \frac{1}{2}\sum_{t\in \mathbf{F}_q} [1+\psi(P(t))]\,\omega(t) + O(1) = \frac{1}{2}\sum_{t\in \mathbf{F}_q} \psi(P(t))\omega(t) + O(1)$$

As before, we have denoted with $P(X)\in \mathbf{F}_q[X]$ the degree n characteristic polynomial of $-x$ over $\mathbf{F}_q$, while ψ is the quadratic character of $\mathbf{F}_q$. The classical estimate for this type of exponential sums follows as a corollary of well known results on Artin $L-$series [12]. Thus, we find that the absolute value of (10) is bounded from above by $n\sqrt{q}/2 + O(1)$. This concludes the proof of the quasi-random character of the above defined subsets $S(q;x)$. Thus, a codeword in $C(q,n)$ can 'safely' be seen as a 'random subset' of $\mathbf{F}_q$ or, equally, as the output of an experiment of random and independent selection of elements of $\mathbf{F}_q$, the probability of picking up a particular one being $1/2 + O(1/\sqrt{q})$. ¿From theorem 6 we find that these experiments are 'almost independent' in the sense that the probability of a given element of $\mathbf{F}_q$ to be selected by each of the two such fixed experiments is in the range $1/4 + O(1/\sqrt{q})$. The implied constants depend only on n. Thinking at $C(q,2)$ only, we see that in fact we managed to construct an explicit example of $(q^2-q)/2$ such 'almost-independent' random subsets of $\mathbf{F}_q$, each one having $(q-1)/2$ elements. By appropriately modifying of the 'O' constants, the codes $C(q,n)$ will provide, for fixed n and big q, examples of roughly q^n/n such 'random subsets'. This can be further improved, if we consider the extensions of the codes $C(q,n)$ by adding the complements of their codewords (see the next section).

We turn now to the codes $C(q,2)$ in order to prove that they are distance invariant, that is, for any positive integer k, the number of codewords at the distance k from a given codeword V_x depends only on k and not on x (this holds, for example, for every linear code). The proof of this fact is easy. Indeed, let x,y be two elements of $\mathbf{F}_{q^2}-\mathbf{F}_q$ which are not Frobenius conjugate. Then, one can find elements $a,b\in \mathbf{F}_q$ with the property that $ax+b=y$. For any codeword V_z, $z\in \mathbf{F}_{q^2}-\mathbf{F}_q$ at a Hamming distance k from V_x, we make correspond the codeword V_{az+b}, which will follow to be at a Hamming distance k from V_y. To see this, we use a property of the distance d which follows easily from the definition. Namely, for any x,z in $\mathbf{F}_{q^2}-\mathbf{F}_q$ and any a,b in $\mathbf{F}_q$, one has

$$d(V_x,V_z) = d(V_{ax+b},V_{az+b})$$

Easy to see that the correspondence

$$V_z \to V_{az+b}$$

is the desired bijection. This concludes the proof of the distance invariance for the codes $C(q,2)$. Incidentally we have found a permutation group consisting of affine transformations which preserves the Hamming distances between the codewords of $C(q,2)$.

5. Further comments

The following question arises naturally : what about if we try to enlarge the codes $C(q,2)$ by adding the complements of their codewords ? Obviously, for $x \in \mathbf{F}_{q^2} - \mathbf{F}_q$ and $t \in \mathbf{F}_q$, the 't–th' component of the complement $\overline{V_x}$ of the codeword V_x will be given by

$$\overline{V_x}(t) = 1 - V_x(t) = \frac{1}{2}(1 - \chi(x+t)) \tag{11}$$

Using the same type of approach as in the proof of proposition 1, we find that a codeword V_x never equals a complement $\overline{V_y}$. In this manner we find out an extended binary code $C^e(q,2)$, having $q^2 - q$ codewords, half of them having the weight $(q-1)/2$ and half the weight $(q+1)/2$. The Hamming distance between two codewords of $C^e(q,2)$ will be in the same range $[q/2 - 3\sqrt{q}/2, \quad q/2 + 3\sqrt{q}/2]$. Indeed, it is enough to prove that the Hamming distance between a V_x and a $\overline{V_y}$ is in this range. Indeed, by using (1) and (11), this distance can be expressed in a way similar to (2):

$$d(V_x, \overline{V_y}) = \frac{1}{2}\left(q + \sum_{t \in \mathbf{F}_q} \chi[(x+t)(y+t)]\right)$$

and we already know that the absolute value of the inner sum was found to be smaller than $3\sqrt{q}/2$. One may extend the codes $C(q,n)$ in a similar way. Using the same approach as that in section 3, we find that if $q > (2n-1)^2$, a codeword V_x of $C(q,n)$ never equals a complement $\overline{V_y}$. Under the same condition one finds then that the weights and distances for the codewords of $C^e(q,n)$ are within the same range as those of the codes $C(q,n)$. The details are left to the reader.

One may notice some similarities between the codes constructed above and the codes belonging to two other classes, that is the classes of codes dual to the double error correcting BCH codes $\mathbf{B}_m(q)$ and Melas codes $\mathbf{M}_m(q)$, respectively. The codes $\mathbf{B}_m(q)$ and $\mathbf{M}_m(q)$ are defined starting from a finite field $\mathbf{F}_q$, $q = 2^m$, $m > 2$. Each one has q^2 codewords of length $q-1$, we have one codeword in the dual $\mathbf{B}_m(q)$ or in the dual of $\mathbf{M}_m(q)$ associated to each pair (λ, μ) of elements of $\mathbf{F}_q$.

It $t \in \mathbf{F}_q^*$, the '$t-$th' component of that codeword belonging to the dual of $\mathbf{B}_m$ which corresponds to the pair (λ, μ) is given by $Tr(\lambda t + \mu t^3)$, while the '$t-$th' component of the codeword belonging to the dual of $\mathbf{M}_m$ which corresponds to the pair (λ, μ) will be $Tr(\lambda t + \mu t^{-1})$, where Tr is the absolute trace function from $\mathbf{F}_q$ to $\mathbf{F}_2$.

In the above two classes of (linear, this time) codes one finds again an 'almost constant' character for the weights of the codewords: if we disregard the zero word, the weights of all the other codewords are within a range of the form $q/2 + O(\sqrt{q})$. The weight distribution for them is completely known (for a complete description based on algebraic geometric methods related to families of elliptic curves over finite fields one may see [11]). For example, if m is odd, the weights of the nonzero codewords in the dual of $\mathbf{B}_m$ are $(q + \sqrt{2q})/2$, $q/2$ and $(q - \sqrt{2q})/2$, the frequency of each weight being $(q-1)(q-\sqrt{2q})/4$, $q(q-1)/2 + q - 1$ and $(q-1)(q+\sqrt{2q})/4$, respectively. For the duals of the Melas codes the weight distributions present a similar character. For more details we refer to [11]. The codes $C(q,2)$ and $C^e(q,2)$ defined above can be looked at as possible 'nonlinear companions' for the duals of $\mathbf{B}_m$ and $\mathbf{M}_m$, their weight and distance distributions presenting the similar feature of being within a range of the form $q/2 + O(\sqrt{q})$, while the number of codewords is asymptotically of the form $O(q^2)$: $C(q,2)$ has $(q^2-q)/2$ codewords (presenting the additional feature of having a constant weight) while $C^e(q,2)$ has $q^2 - q$ codewords (with only two possible weights, one unit apart, $(q-1)/2$ and $(q+1)/2$). Assuming that $q > (2n-1)^2$, we see that the weight and distance distributions of the codes $C(q,n)$ and $C^e(q,n)$ present a similar behavior, while from the point of view of the number of codewords the situation is asymptotically better, this being of the form $O(q^n)$ (although the implied constants depend on n). Indeed, by using theorem 6 above, it is clear that the number of codewords in $C(q,n)$ will be given by

$$|C(q,n)| = \sum_{d|n,\, \frac{n}{d} \equiv 1 (mod 2)} N_d$$

where N_d represents the number of monic irreducible polynomials over $\mathbf{F}_q$ of degree d. For a fixed n and big q, that is asymptotically q^n/n. By using the well known formula for N_d we find

$$|C(q,n)| = \sum_{d|n,\, \frac{n}{d} \equiv 1 (mod 2)} \left[\frac{1}{d} \sum_{s|d} \mu(s) q^{\frac{d}{s}} \right]$$

where μ represents here the Möbius function. In the special case in which n is an odd prime we find, for example that

$$|C(q,n)| = N_1 + N_n = q + \frac{q^n - q}{n}$$

References

[1] Blake, I.F., Mullin, R.C., An introduction to algebraic and combinatorial coding theory, Academic Press, 1976

[2] Bombieri, E., Counting points on curves over finite fields (d'apres Stepanov), Sem.Bourbaki 1972/1973, Exp.430,LNM, vol 383 (1974), 234-241

[3] Chung, F.R.K., Diameters and eigenvalues, Journal of the American Math.Soc., vol.2, no.2 (1989), 187-196

[4] Chung, F.R.K., Graham, R.L., Quasi-Random Subsets of $\mathbf{Z}_n$, J.Combinatorial Theory, Series A, 61 (1992), 64-86

[5] Davenport, H., On primitive roots in finite fields, Quart.J.Math.,(2), 8 (1937), 308-312

[6] Deuring,M., Lectures on the theory of algebraic functions of one variable, LNM vol 314, Springer-Verlag, 1973

[7] Hasse, H., Theorie der relativ-zyklischen algebraichen Functionenkorper, insbesondere bei endlichen Konstantenkorper, J.Reine Angew.Math., 172, 37-54, 1934

[8] Lidl,R., Niederreiter,H., Finite fields, Encyclopedia of Mathematics and its applications, vol.20, Addison-Wesley, Reading, Mass., 1983

[9] MacWilliams, F.J., Sloane, N.J.A., The Theory of Error-Correcting Codes, North Holland, 1977

[10] Schmidt,W., Equations over finite fields, an elementary approach, LNM, vol.536, Springer 1976

[11] Schoof, R., Families of curves and weight distribution of codes, Bull.A.M.S., vol.32, 2 (1995),171-183

[12] Stepanov, S.A., Arithmetic of algebraic curves, New York, 1994

[13] Stichtenoth, H, Algebraic function fields and codes, Springer-Verlag,1993

[14] Tsfasman, M.,Vladut, S., Algebraic-geometric codes, Math. and its Appl., Kluwer Acad. Publishers, Dordrecht, The Netherlands, 1991

[15] Weil, A., Sur les courbes algebriques et les varietes qui s'en deduisent, Actualites Sci. Ind., no.1041, Hermann, Paris, 1948