On the weight distribution of perfect binary codes

Marco Pavone
Dipartimento di Ingegneria
Università degli Studi di Palermo
Palermo 90128 ITALIA

Abstract

In this paper we give a new proof of the closed-form formula for the weight distribution of a perfect binary single-error-correcting code.

2010 AMS MSC: 11T71, 94B25, 94B05
Keywords: Perfect codes, Binary codes, Hamming codes, weight distribution
E-mail: marco.pavone@unipa.it
ORCID: 0000-0002-8674-5841
Funding: Università di Palermo (FFR)

1 Introduction

A binary code C of length n is any subset of the n-dimensional vector space $\mathrm{GF}(2)^n$ over $\mathrm{GF}(2)$. The *(Hamming) distance* $\mathrm{d}(x, y)$ between two vectors x, y in $\mathrm{GF}(2)^n$ is the number of coordinates in which x and y differ. The *(Hamming) weight* $\mathrm{wt}(x)$ of a vector x is the number of non-zero coordinates of x. Equivalently, the weight of x is equal to the distance $\mathrm{d}(x, \mathbf{0})$ between x and the all-zero vector $\mathbf{0} = (0, \ldots, 0)$ of length n.

The *covering radius* of C is the smallest value of the non-negative integer r such that the balls of radius r, with respect to the Hamming distance, centered at the codewords of C, cover all of $\mathrm{GF}(2)^n$. Equivalently, the covering radius of C can be defined as

$$r = \max_{y \in \mathrm{GF}(2)^n} \min_{x \in C} \mathrm{d}(x, y). \tag{1}$$

The *packing radius* of C is the largest value of the non-negative integer s such that the balls of radius s, centered at the codewords of C, are mutually disjoint. A binary code of length n is said to be *perfect* if the covering radius is equal to the packing radius, that is, if there exists a (necessarily unique) non-negative integer r such that every x in $\mathrm{GF}(2)^n$ is within distance r from exactly one codeword of C. In this case, C is said to be an r-perfect code.

It is well known that perfect binary codes of length n exist only for $r = 0$, $r = n$, $r = (n-1)/2$ with n odd, $r = 1$ with $n = 2^m - 1$, $m \geq 2$, and $r = 3$ with $n = 23$ (see e.g. [3]). The first three types of codes are called trivial, whereas the last case corresponds to the binary Golay code (see e.g. [7]). In the remaining case, the perfect codes with $r = 1$ and $n = 2^m - 1$, $m \geq 2$, are called perfect binary single-error-correcting codes (or 1-perfect binary codes). For the sake of brevity, it is customary to refer to these codes simply as *perfect binary codes*.

The linear perfect binary codes are unique, and are the well-known binary Hamming codes (see e.g. [7]). The first examples of non-linear perfect binary codes of length $n = 2^m - 1$ were constructed by Vasil'ev in 1962 for any $m \geq 4$ [13]. Since then, they have been intensively studied, and the classification problem for such codes is far from being solved. Perfect binary codes are one of the most important

topics in the theory of error-correcting codes, and are important also for combinatorics, graph theory, group theory and cryptography.

Given a code C of length n, and an integer $0 \leq i \leq n$, let us denote by A_i the number of codewords of Hamming weight i in C. The ordered sequence $(A_0, A_1, \ldots, A_n)$ is called the *weight distribution* of the code C. The weight distribution, which is an important research topic in coding theory, contains crucial information about the error-correcting capability of the code and the probability of error detection and correction, and is often related to interesting problems in number theory and design theory.

In Section 2 we give a new, elementary, proof of the closed-form expression for the weight distribution of a perfect binary code. Unlike in the original proof by Etzion and Vardy [3], we do not rely on the weight distributions of the translates of the code. In Section 3 we observe that the recursive relation for the weight distribution of a q-ary Hamming code of length $n = (q^m - 1)/(q - 1)$ over $\mathrm{GF}(q)$, with q a prime power (see e.g. [10, Problem 4.8, p. 121]), is valid more generally for any 1-perfect q-ary code of length $n = (q^m - 1)/(q - 1)$.

2 Weight distributions of 1-perfect binary codes

If C is a binary code of length n, and y is a vector in $\mathrm{GF}(2)^n$, then

$$C + y = \{x + y \,|\, x \in C\}$$

is also a code of length n, which is called *translate* of C by y (of course, if C is linear, then $C + y$ is the coset of C containing y). The study of the translates of C is an important tool for investigating the properties of a code C. For instance, since

$$\mathrm{d}(x, y) = \mathrm{d}(x + y, \mathbf{0}) = \mathrm{wt}(x + y), \tag{2}$$

it follows from (1) that the covering radius of C is the maximum of the smallest weight in any translate of C. This property shows also how one can use translates for decoding. Suppose, for instance, that at least one transmission error occurs, and some received word y is not in C. Then, by (2), x is a codeword in C at smallest distance from y if and only if $x + y$ is a vector of least weight in the translate $C + y$ (called the coset leader of $C + y$).

Another immediate well-known property of translates, in the case of a 1-error perfect binary code C of length n, is that, if $e_1, \ldots, e_n$ are the vectors of the canonical basis of $\mathrm{GF}(2)^n$, then $\mathrm{GF}(2)^n$ can be partitioned as the disjoint union of $C, C+e_1, \ldots, C+e_n$. In particular, the number $|C|$ of codewords in C must satisfy the equality $|C|(n+1) = 2^n$. Also, this partition of the space $\mathrm{GF}(2)^n$ was used by Solov'eva [12] and, independently, by Phelps [9], to construct perfect binary codes of length $2n + 1$. This was one of the first constructions of non-linear perfect binary codes since Vasil'ev [13].

For a 1-perfect binary code C of length n, Etzion and Vardy [3, Proposition 4.1] found a closed-form expression for the weight distribution of C, starting from the well-known doubly-recursive relation

$$(n - i + 1)A_{i-1} + A_i + (i + 1)A_{i+1} = \binom{n}{i} \tag{3}$$

(see e.g. [7, p. 129]), where A_i denotes the number of codewords of weight i in C. In the special case of a binary Hamming code, the closed-form expression had already been given by Shapiro and Slotnick in [11, Remark 2, p. 28].

Note that the equation (3) has only two possible solutions, depending on whether C contains the zero vector ($A_0 = 1, A_1 = 0$) or not ($A_0 = 0, A_1 = 1$). If $e_1, \ldots, e_n$ are the vectors of the canonical basis of $\mathrm{GF}(2)^n$, then, as we noted above, $\mathrm{GF}(2)^n$ can be partitioned as the disjoint union of the (1-error perfect) codes $C, C+e_1, \ldots, C+e_n$. If C contains the zero vector, then the codes $C+e_1, \ldots, C+e_n$ do not contain it, hence they all share the same weight distribution. Therefore, if B_i denotes the common number of words of weight i in any of the codes $C+e_1, \ldots, C+e_n$, one obtains that $A_i + nB_i = \binom{n}{i}$, which, together with the relation $(n-i+1)A_{i-1} + A_i + (i+1)A_{i+1} = (n-i+1)B_{i-1} + B_i + (i+1)B_{i+1}$ (see (3) above),

produces by induction an explicit expression for A_i and B_i [3]. Note that B_i is precisely the value taken up by A_i in the case where C does not contain the zero vector.

We now give a new proof of the closed-form expression for A_i. The idea is very simple, and relies on the observation that by writing the equality (3) twice, for two consecutive values of the index i, one gets by addition a linear recurrence relation between $A_{i+1}+A_{i+2}$ and $A_{i-1}+A_i$, which, by induction, produces an explicit expression for $A_{i-1}+A_i$. Finally, again by induction, one obtains the closed-form expression for $A_i = (A_{i-1}+A_i) - A_{i-1}$. Besides this general scheme, the rest of the proof is just a very easy exercise, which involves only elementary properties of the binomial coefficients. Moreover, our proof considers only the codewords of the given perfect binary code, without resorting to the weight distributions of the translates of the code.

2.1 Theorem: [3, Proposition 4.1] *Let C be a perfect binary code of length $n = 2^m - 1$, with $m \geq 3$. For each $0 \leq i \leq n$, if A_i is the number of codewords of weight i in C, then*

$$A_i = \begin{cases} \dfrac{1}{n+1}\binom{n}{i} + (-1)^{i+\lfloor i/2 \rfloor}\, \dfrac{n}{n+1}\binom{\frac{n-1}{2}}{\lfloor i/2 \rfloor} & \text{if } \mathbf{0} \in C \\ \dfrac{1}{n+1}\binom{n}{i} - (-1)^{i+\lfloor i/2 \rfloor}\, \dfrac{1}{n+1}\binom{\frac{n-1}{2}}{\lfloor i/2 \rfloor} & \text{if } \mathbf{0} \notin C, \end{cases} \tag{4}$$

where $\lfloor \cdot \rfloor$ is the floor function, and $\mathbf{0}$ denotes the zero vector $(0, \ldots, 0)$ in $\mathrm{GF}(2)^n$.

Proof. Let us first prove that, for any integer i, with $1 \leq i \leq n$,

$$A_{i-1} + A_i = \begin{cases} \dfrac{1}{n+1}\binom{n+1}{i} & \text{if } i \text{ is odd} \\ \dfrac{1}{n+1}\binom{n+1}{i} + (-1)^{i/2}\, \dfrac{n}{n+1}\binom{\frac{n+1}{2}}{i/2} & \text{if } i \text{ is even and } \mathbf{0} \in C \\ \dfrac{1}{n+1}\binom{n+1}{i} - (-1)^{i/2}\, \dfrac{1}{n+1}\binom{\frac{n+1}{2}}{i/2} & \text{if } i \text{ is even and } \mathbf{0} \notin C. \end{cases} \tag{5}$$

If $\mathbf{0} \in C$, then $A_0 = 1$. Therefore $A_1 = 0$ and $A_2 = 0$, else there would exist a vector in $\mathrm{GF}(2)^n$ within distance 1 from at least two codewords of C (see also the equality (3)). If $\mathbf{0} \notin C$, then $A_0 = 0$, thus $A_1 = 1$ again by definition of perfect code, whence $A_2 = \frac{n-1}{2}$ by the equality (3). It follows that, in either case, the formula (5) is satisfied for $i = 1, 2$.

For $1 \leq i \leq n-2$, let us consider the equality (3), and the equality

$$(n-i)A_i + A_{i+1} + (i+2)A_{i+2} = \binom{n}{i+1} \tag{6}$$

obtained from (3) by replacing i with $i+1$. By adding up the equalities (3) and (6), one gets a linear recurrence relation between $A_{i+1} + A_{i+2}$ and $A_{i-1} + A_i$,

$$(n-i+1)(A_{i-1} + A_i) + (i+2)(A_{i+1} + A_{i+2}) = \binom{n+1}{i+1},$$

that is,

$$A_{i+1} + A_{i+2} = \frac{1}{i+2}\left(\binom{n+1}{i+1} - (n-i+1)\,(A_{i-1} + A_i)\right). \tag{7}$$

We can now prove (5) by induction. If $1 \leq i \leq n-2$ is odd, then $i+2$ is also odd, hence, by (7) and (5),

$$\begin{aligned} A_{i+1} + A_{i+2} &= \frac{1}{i+2}\left(\binom{n+1}{i+1} - \frac{n-i+1}{n+1}\binom{n+1}{i}\right) \\ &= \frac{1}{i+2}\left(\binom{n+1}{i+1} - \frac{i+1}{n+1}\binom{n+1}{i+1}\right) \\ &= \frac{1}{i+2}\binom{n+1}{i+1}\frac{n-i}{n+1} \\ &= \frac{1}{n+1}\binom{n+1}{i+2}. \end{aligned}$$

Finally, let us consider the case where $2 \leq i \leq n-3$ is even, say $i = 2m$. Thus $i+2 = 2(m+1)$, hence, if $\mathbf{0} \in C$, then, by (7) and (5),

$$\begin{aligned} A_{i+1} + A_{i+2} &= \frac{1}{i+2}\binom{n+1}{i+1} - \frac{n-i+1}{i+2}\left(\frac{1}{n+1}\binom{n+1}{i} + (-1)^m \frac{n}{n+1}\binom{\frac{n+1}{2}}{m}\right) \\ &= \frac{1}{i+2}\binom{n+1}{i+1} - \frac{i+1}{(i+2)(n+1)}\binom{n+1}{i+1} \\ &\quad + (-1)^{m+1}\frac{\frac{n+1}{2}-m}{m+1}\frac{n}{n+1}\binom{\frac{n+1}{2}}{m} \\ &= \frac{n+1-(i+1)}{(i+2)(n+1)}\binom{n+1}{i+1} + (-1)^{m+1}\frac{n}{n+1}\binom{\frac{n+1}{2}}{m+1} \\ &= \frac{1}{n+1}\binom{n+1}{i+2} + (-1)^{m+1}\frac{n}{n+1}\binom{\frac{n+1}{2}}{m+1}. \end{aligned}$$

The case where $\mathbf{0} \notin C$ is similar, and can be worked out with minor variations. This completes the induction, hence the equality (5) is proved.

Let us now consider the formula (4). For $i = 0, 1$, the equality (4) is trivial. For $2 \leq i \leq n$, we can proceed by induction. Again, we can assume that $\mathbf{0} \in C$, the other case being similar. Alternatively, if we denote by B_i the right-hand side of (4) in the case where $\mathbf{0} \notin C$ (and we still denote by A_i the right-hand side of (4) in the case where $\mathbf{0} \in C$), then, as we noted earlier, $A_i + nB_i = \binom{n}{i}$, hence the formula for B_i can be immediately derived from that for A_i.

If i is odd, say $i = 2m+1$, then $\lfloor i/2 \rfloor = m$, $i-1 = 2m$ and, by (5) and (4),

$$\begin{aligned} A_i &= (A_{i-1} + A_i) - A_{i-1} \\ &= \frac{1}{n+1}\binom{n+1}{i} - \left(\frac{1}{n+1}\binom{n}{i-1} + (-1)^{3m}\frac{n}{n+1}\binom{\frac{n-1}{2}}{m}\right) \\ &= \frac{1}{n+1}\binom{n}{i} + (-1)^{3m+1}\frac{n}{n+1}\binom{\frac{n-1}{2}}{m}, \end{aligned}$$

that is, (4) holds.

Finally, if i is even, say $i = 2m$, then $\lfloor i/2 \rfloor = m$, $i-1 = 2m-1$, $\lfloor (i-1)/2 \rfloor = m-1$, and, by (5),

and (4),

$$\begin{aligned} A_i &= (A_{i-1} + A_i) - A_{i-1} \\ &= \frac{1}{n+1}\binom{n+1}{i} + (-1)^m \frac{n}{n+1}\binom{\frac{n+1}{2}}{m} \\ &\quad -\left(\frac{1}{n+1}\binom{n}{i-1} + (-1)^{3m} \frac{n}{n+1}\binom{\frac{n-1}{2}}{m-1}\right) \\ &= \frac{1}{n+1}\binom{n}{i} + (-1)^{3m} \frac{n}{n+1}\binom{\frac{n-1}{2}}{m}, \end{aligned}$$

that is, (4) holds. The proof is now complete. □

2.2 Remark: In the special case where C is the $(2^m - 1, 2^m - m - 1, 3)$-Hamming code of length $n = 2^m - 1$, $m \geq 3$, there exists a one-to-one correspondence between the codewords of weight k in C, $1 \leq k \leq n$, and the family $\mathcal{B}_k^*$ consisting of the subsets of $\mathrm{GF}(2)^m \setminus \{\mathbf{0}\}$ of size k whose elements sum up to $\mathbf{0}$. Moreover, $\mathcal{D}_k^* = (\mathrm{GF}(2)^m \setminus \{\mathbf{0}\}, \mathcal{B}_k^*)$ is a 2-(n, k, λ) design, and the problem of the weight distribution of C corresponds to the computation of the numbers of blocks of the *additive* designs $\mathcal{D}_k^*$ ([5]; see also [8, 4] for the case where the point-set of the design is $\mathrm{GF}(p)^m \setminus \{\mathbf{0}\}$, with p an odd prime. See [1, 2] for the general setting of additive designs).

From yet another point of view, the blocks of the design $\mathcal{D}_k^*$ can be seen in the context of the well-known *subset sum problem* over finite fields, in that they are the (unordered) solutions of the equation $x_1 + \cdots + x_k = 0$, where $x_1, \ldots, x_k \in \mathrm{GF}(p)^m \setminus \{\mathbf{0}\}$, $p = 2$, and $x_i \neq x_j$ for all $i \neq j$. The number of different solutions of the equation, in the general case of a prime number p, was first given in a celebrated result by Li and Wan [6, Theorem 1.2], which, in the special case where $p = 2$, reduces to the above formula (4) by Etzion and Vardy [3, Proposition 4.1].

3 Weight distributions of 1-perfect q-ary codes

A simple recursive formula for the weight distribution can be found also in the general case where C is a 1-perfect q-ary code, as a generalization of the above relation (3), which was valid for perfect binary codes. A q-ary code C of length n is any subset of the space $\mathrm{GF}(q)^n$, where q is a prime power. If for some $r \geq 0$ every x in $\mathrm{GF}(q)^n$ is within distance r from exactly one codeword of C, then the code C is called r-perfect. It is well known (see e.g. [7]) that nontrivial perfect q-ary codes must have length $n = (q^m - 1)/(q - 1)$, for some integer $m \geq 2$.

Let C be a 1-perfect q-ary code of length $n = (q^m - 1)/(q - 1)$, and let $1 \leq i \leq n - 1$. The number of vectors of weight i in $\mathrm{GF}(q)^n$ is precisely $\binom{n}{i}(q-1)^i$. Any such vector is either a codeword in C (of weight i), or is at distance 1 from exactly one codeword of C, whose weight can be either $i - 1$, i or $i + 1$. Now each codeword in C of weight $i + 1$ is at distance 1 from $i + 1$ vectors of weight i in $\mathrm{GF}(q)^n$, each of which is obtained by changing into 0 one of the $i + 1$ nonzero coordinates of the codeword. Each codeword in C of weight i is at distance 1 from $(q - 2)i$ vectors of weight i in $\mathrm{GF}(q)^n$, each of which is obtained by changing one of the i nonzero coordinates of the codeword into a different nonzero value in $\mathrm{GF}(q)$. Finally, each codeword in C of weight $i - 1$ is at distance 1 from $(q - 1)(n - i + 1)$ vectors of weight i in $\mathrm{GF}(q)^n$, each of which is obtained by changing one of the $n - (i - 1)$ zero coordinates of the codeword into a nonzero value in $\mathrm{GF}(q)$. It follows that

$$(q-1)(n-i+1)A_{i-1} + ((q-2)i+1)A_i + (i+1)A_{i+1} = \binom{n}{i}(q-1)^i. \tag{8}$$

Note that, for $q = 2$, the formula simply reduces to the above relation (3). In the special case of the q-ary Hamming code $H(m, q)$ over $\mathrm{GF}(q)$, which is a single-error-correcting perfect linear code of length

$n = (q^m - 1)/(q - 1)$, with $q \neq 2$ a prime power and $m \geq 2$, the above relation (8) can be found in [10, Problem 4.8, p. 121].

It is natural to ask whether a closed-form expression for A_i can be found from (8) by arguing as in the proof of the above Theorem 2.1, that is, by writing (8) for the successive value of the index i, in order to get, by addition, a recursive relation between $A_{i+1} + A_{i+2}$ and $A_{i-1} + A_i$. Unfortunately, in this case the argument fails. One may ask, however, if the relation (8) can be used in some other way to derive the desired closed-form formula for the weight distribution of the code.

References

[1] A. Caggegi, G. Falcone, M. Pavone, On the additivity of block designs, *J. Algebr. Comb.* **45**, 271–294 (2017).

[2] A. Caggegi, G. Falcone, M. Pavone, Additivity of affine designs, *J. Algebr. Comb.* **53**, 755–770 (2021).

[3] T. Etzion, A. Vardy, Perfect Binary Codes: Constructions, Properties, and Enumeration, *IEEE Trans. Inf. Theory* **40** (3), 754–763 (1994).

[4] G. Falcone, M. Pavone, Permutations of zero-sumsets in a finite vector space, *Forum Math.* **33** (2), 349–359 (2021).

[5] G. Falcone, M. Pavone, Binary Hamming codes and Boolean designs, *Des. Codes Cryptogr.* (2021). DOI: 10.1007/s10623-021-00853-z.

[6] J. Li, D. Wan, On the subset sum problem over finite fields, *Finite Fields Appl.* **14** (4), 911–929 (2008).

[7] F. J. MacWilliams, N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, New York (1977).

[8] M. Pavone, Subset sums and block designs in a finite vector space (submitted).

[9] K. T. Phelps, A combinatorial construction of perfect codes, *SIAM J. Algebraic and Discrete Methods* **4**, 398–403 (1983).

[10] R. M. Roth, *Introduction to coding theory*, Cambridge University Press, Cambridge (2006).

[11] H. S. Shapiro, D. L. Slotnick, On the Mathematical Theory of Error-Correcting Codes, *IBM J. Res. Dev.* **3** (1), 25–34 (1959).

[12] F. I. Solov'eva, Binary nongroup codes (in Russian), *Methody Discret. Analiz.* **37**, 65–76 (1981).

[13] J. L. Vasil'ev, On nongroup close-packed codes, *Probl. Kibernet.* **8**, 375–378 (in Russian) (1962).