EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 17, No. 4, 2024, 4225-4237
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

Weights of Codewords in Random Binary Linear Codes and Their Correlation Properties

İbrahim Özen

Department of Matheamatics, Faculty of Science, Marmara University, Istanbul, Turkey.

Abstract. A classical problem in coding theory addresses moments of the weight spectrum distribution. The results in this work are on the weight moments of individual codewords rather than the weight spectrum. The expectations of single and pairwise products of weights of nonzero words in a random binary linear code are given. We show that the covariance between the weights of any pair of distinct nonzero words is zero. Our main theorem has an application to sequence correlations problem. We prove that the sums of out of phase self correlations, as well as sums of cross-correlations, of nonzero words in a random binary linear code are equal to zero.

2020 Mathematics Subject Classifications: 94B05, 94A55, 94B65

Key Words and Phrases: Random binary linear codes, Weight moments of codewords, Binary sequences, Auto-correlations, Cross-correlations

1. Introduction

A binary linear code C is a subspace of $\mathbb{F}_2^n$. We call n the length of the code. Dimension of the code is its dimension as a linear space. A vector in the code is called a codeword. Let $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ and $\mathbf{e} = (e_1, e_2, \ldots, e_n)$ be two binary vectors. The Hamming distance between them is the number of coordinates that they differ and is defined by

$$d_H(\mathbf{c}, \mathbf{e}) = |\{i : c_i \neq e_i\}|.$$

For every vector $\mathbf{c}$, the number of nonzero positions in $\mathbf{c}$ is called its weight and we denote it by $\|\mathbf{c}\|$. The smallest weight of the nonzero codewords is called the minimum weight of the code. Since C is a linear space, the minimum weight of the code coincides with the minimum of the distances between distinct codewords of the code. If $C \subset \mathbb{F}_2^n$ is a linear code with dimension k and minimum weight d, we say that C is a binary $[n, k, d]$ code. If we let A_i denote the number of codewords with weight i, then the sequence $(A_0, A_d, A_{d+1}, \ldots, A_n)$ is called the weight distribution or the weight spectrum of the code.

DOI: https://doi.org/10.29020/nybg.ejpam.v17i4.5600

Email address: iozen@marmara.edu.tr (İ. Özen)

Every important feature of a code is encoded in its weight distribution. For example, if the weight distribution of a linear code is known we can obtain estimations of its decoding error probability. Gallager estimated the decoding error probability of a random linear code in [4]. So we have a way to compare this property of a specific code as well as information on what is possible for the same quality. This estimation is based on the expectations of the weight distribution of random linear codes. The expectations of weights are part of the code weight distribution problem, namely the evaluations of the expectations

$$\mathbb{E}\left(A_{i_1}A_{i_2}\cdots A_{i_k}\right)$$

for random linear codes. Second and third joint moments of the weight distribution of a random linear code were explored in [6] and [1] independently. Fourth moments were given in [8]. In [9] and [7], big orders of the moments were studied.

In this work we focus on the weight moments of individual codewords in a random binary linear code. Our results have applications on the expectations of correlation sums of words.

Let us denote the words of an $[n,k]$ random binary linear code by $C = \{\mathbf{c}_0 = \mathbf{0}, \mathbf{c}_1, \ldots, \mathbf{c}_{2^k-1}\}$. The weights of the words will be denoted by $\{w_0 = 0, w_1, \ldots, w_{2^k-1}\}$ respectively. In Theorem 1 we obtain the expectations of the nonzero words' weights in a random $[n,k]$ binary linear code and they are given by

$$\mathbb{E}(w_i) = \frac{n}{2}, \quad \text{for } i \neq 0.$$

Expectations of pairwise products of the weights are obtained in Theorem 2 and they can be stated as

$$\mathbb{E}(w_i w_j) = \begin{cases} \frac{n^2+n}{4}, & \text{for } i = j,\ i,j \geq 1, \\ \frac{n^2}{4}, & \text{for } i \neq j,\ i,j \geq 1. \end{cases}$$

An immediate consequence is that the weights of distinct nonzero words in a random binary linear code are statistically uncorrelated.

$$\begin{aligned} \text{covariance}(w_i, w_j) &= \mathbb{E}(w_i w_j) - \mathbb{E}(w_i)\mathbb{E}(w_j) \\ &= 0, \text{ for all } i,j \geq 1 \text{ and } i \neq j \end{aligned}$$

We apply these results to obtain correlation properties of words in a random code.

Let $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ and $\mathbf{e} = (e_1, e_2, \ldots, e_n)$ be two vectors in $\mathbb{F}_2^n$. The periodic auto-correlation and cross-correlation functions on vectors of $\mathbb{F}_2^n$ are defined respectively by

$$\begin{aligned} r_{\mathbf{c},\mathbf{c}}(u) &= r_{\mathbf{c}}(u) = \sum_{i=1}^{n} (-1)^{c_i + c_{i+u}} \text{ and} \\ r_{\mathbf{c},\mathbf{e}}(u) &= \sum_{i=1}^{n} (-1)^{c_i + e_{i+u}}, \end{aligned}$$

where the indices are evaluated modulo n.

Let $\mathbf{c}$ and $\mathbf{e}$ be two distinct nonzero codewords in a random binary linear $[n,k]$ code. We obtain in Theorem 3 that

$$\begin{aligned}\sum_{u=1}^{n-1}\mathbb{E}(r_{\mathbf{c}}(u)) &= 0 \text{ and}\\ \sum_{u=0}^{n-1}\mathbb{E}(r_{\mathbf{c},\mathbf{e}}(u)) &= 0.\end{aligned}$$

2. Characteristic Vectors of Codes and Weights of Nonzero Codewords

Our main theorem is based on the characterization of weights of nonzero words in terms of the characteristic vector of a code, given in [2].

We will review this characterization in the following paragraphs.

Let $k \geq 2$ and i be integers with $1 \leq i \leq 2^k-1$. We denote by $\mathbf{i} = (i_0, i_1, \ldots, i_{k-1}) \in \mathbb{F}_2^k$, the binary expansion of $i = \sum_{j=0}^{k-1} i_j 2^j$. We form the $k \times (2^k-1)$ matrix G_k, whose ith column is $\mathbf{i}^T$.

Let C be a binary linear $[n,k]$ code and let G be a generator matrix of C. We will show below that with high probability a random matrix G has no columns of zeros. We will assume from now on that C is a code with a generator matrix without zero columns. For such C and G, we define the characteristic vector χ_C of C with respect to G as the vector $\chi_C = (h_1, h_2, \ldots, h_{2^k-1})^T$, where h_i is the number of columns in G equal to the ith column $\mathbf{i}^T$ of G_k. Clearly we have $\sum_j h_j = n$ (recall that G has no zero columns).

Now since any nonzero codeword is obtained by a nontrivial linear combination of rows of G, all nonzero codewords of C are obtained as the rows of the matrix product $G_k^T G$. So we identify the nonzero words of C with the rows of the product $G_k^T G$; $C = \{\mathbf{0}\} \cup \{\mathbf{c_i}\}_{i=1}^{2^k-1}$, where $\mathbf{c_i}$ is the ith row of $G_k^T G$.

If we denote the matrix product $G_k^T G_k$ by $\mathcal{M}_k$, we can express the relation between the weights of the nonzero codewords of C and the characteristic vector χ_C as in the following lemma.

Lemma 1. *[2, Lemma 1] Let C be a binary $[n,k]$ linear code, let G be a generator matrix of C and let χ_C be the characteristic vector of C with respect to G. Then the weight of ith row $\mathbf{c}_i$ of $G_k^T G$ equals to the ith entry of the column vector $\mathcal{M}_k \chi_C$*

$$\mathcal{M}_k \chi_C = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{2^k-1} \end{bmatrix},$$

where $\{w_1, w_2, \ldots, w_{2^k-1}\}$ is the set of weights of all nonzero words in C.

Let f be a sequence of real numbers of length 2^k, whose terms are indexed from 0 to 2^k-1 and let us denote by $f(i)$ its ith coordinate. Then we define its Walsh transformation $\hat{f}$ as a sequence of real numbers with the same length, whose ith entry is given by

$$\hat{f}(i) = \sum_{j=0}^{2^k-1} f(j)(-1)^{\langle \mathbf{j},\mathbf{i}\rangle},\ 0 \leq i \leq 2^k-1,$$

where $\langle\ ,\ \rangle$ is the standard inner product on $\mathbb{F}_2^k$, $\mathbf{i}$ and $\mathbf{j}$ are the binary expansions of i and j respectively.

From $\chi_C = (h_1, h_2, \ldots, h_{2^k-1})^T$ we obtain a sequence of integers of length 2^k by $\overline{\chi}_C = (h_0 = 0, h_1, h_2, \ldots, h_{2^k-1})^T$. Then the Walsh transform of $\overline{\chi}_C$ is the vector whose ith entry is given by

$$\hat{\overline{\chi}}_C(i) = \sum_{j=0}^{2^k-1} \overline{\chi}_C(j)(-1)^{\langle \mathbf{j},\mathbf{i}\rangle} = \sum_{j=0}^{2^k-1} h_j(-1)^{\langle \mathbf{j},\mathbf{i}\rangle}.$$

We will also need the matrix

$$\overline{\mathcal{M}}_k = \left[\begin{array}{c|ccc} 0 & 0 & \ldots & 0 \\ \hline \vdots & & \mathcal{M}_k & \\ 0 & & & \end{array}\right].$$

The Walsh transform of the characteristic vector is expressed in terms of the Hadamard matrix obtained by Sylvester construction. These matrices are constructed recursively with the help of the Hadamard product as follows:

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},\ H_i = H_1 \otimes H_{i-1}.$$

If we index the rows and columns of H_k by $0 \leq i \leq 2^k-1$ and $0 \leq j \leq 2^k-1$ respectively, then the (i,j) entry of H_k is given by [5, Page 11]

$$H_k[i,j] = (-1)^{\langle \mathbf{i},\mathbf{j}\rangle}. \tag{1}$$

The lemma below is taken from [2], we will prove it for completeness.

Lemma 2. *[2, Page 267] Let C be a binary $[n,k]$ linear code, let G be a generator matrix of C, and let χ_C be the characteristic vector of C with respect to G. Then we have*

$$H_k\overline{\chi}_C \quad = \quad (J - 2\overline{\mathcal{M}}_k)\overline{\chi}_C = \begin{bmatrix} n \\ n-2w_1 \\ n-2w_2 \\ \vdots \\ n-2w_{2^k-1} \end{bmatrix} = \hat{\overline{\chi}}_C,$$

where J is the all one matrix, $\{w_1, w_2, \ldots, w_{2^k-1}\}$ is the set of weights of all nonzero words in C.

Proof. We will index all entries of the matrices and vectors starting from 0. All assertions for the terms indexed by zero are trivial. Let i be an integer with $1 \leq i \leq 2^k - 1$. Then by (1) the term indexed by i in the product $H_k\overline{\chi}_C$ equals

$$\sum_{j=0}^{2^k-1}(-1)^{\langle \mathbf{i},\mathbf{j}\rangle}\overline{\chi}_C(j) = \sum_{j=0}^{2^k-1}(-1)^{\langle \mathbf{i},\mathbf{j}\rangle}h_j, \tag{2}$$

with $\mathbf{i}$ and $\mathbf{j}$ being the binary representations of i and j respectively.

The same term in the product $(J - 2\overline{\mathcal{M}}_k)\overline{\chi}_C$ is $\sum_{j=0}^{2^k-1}(1 - 2\langle s_i, s_j\rangle)h_j$, where s_l's are the column vectors of G_k. One can easily confirm that we have

$$1 - 2\langle s_i, s_j\rangle = 1 - 2\langle \mathbf{i}, \mathbf{j}\rangle = (-1)^{\langle \mathbf{i},\mathbf{j}\rangle}.$$

So we have $\sum_{j=0}^{2^k-1}(1 - 2\langle s_i, s_j\rangle)h_j = \sum_{j=0}^{2^k-1}(-1)^{\langle \mathbf{i},\mathbf{j}\rangle}h_j$ and by (2), the first equality of the lemma follows. The second equality follows from Lemma 1.

If we compare the *ith* coordinates of $H_k\overline{\chi}_C$ and $\hat{\overline{\chi}}_C$ we obtain the last equality of the lemma

$$(H_k\overline{\chi}_C)(i) = \sum_{j=0}^{2^k-1} h_j(-1)^{\langle \mathbf{i},\mathbf{j}\rangle} = \hat{\overline{\chi}}_C(i).$$

In the next section we will make use of the identity in the following lemma. The assertion of the first line is trivial and the second line is an easy consequence of [3, Lemma 3, Page 58].

Lemma 3. *For any integer* $0 \leq i \leq 2^k - 1$*, we have*

$$\sum_{j=0}^{2^k-1}(-1)^{\langle \mathbf{j},\mathbf{i}\rangle} = \begin{cases} 2^k, & \text{for } i = 0, \\ 0, & \text{for } i \neq 0. \end{cases}$$

3. First and Second Moments of the Codeword Weights

A random binary linear code is the row-space of a random $k \times n$ matrix, whose entries are identically and independently distributed in $\mathbb{F}_2$. A well known property of a random $k \times n$ matrix over a finite field is that when $k/n < 1$, the matrix has rank k with probability tending to 1, as n tends to infinity. So we will assume that a random binary $k \times n$ matrix whose terms are distributed identically and independently in $\mathbb{F}_2$ has rank k and its row space C is an $[n, k]$ linear code. The words of such a code will be denoted by $C = \{\mathbf{c}_0 = \mathbf{0}, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{2^k-1}\}$, with the same relation to $G_k^T G$ as above. The weights of the nonzero words will be denoted by $\{w_1, w_2, \ldots, w_{2^k-1}\}$ respectively.

Another fact about this ensemble of matrices is that, the probability that a random $k \times n$ matrix has zero columns tends to 0 as n tends to infinity with $k/n > 0$. This can easily be observed by

$$\lim_{n\to\infty} \quad 1 - \frac{(2^k - 1)^n}{2^{kn}} = 0,$$

with the condition that $R = k/n \neq 0$.

By Lemma 2, the expectations of weights of the nonzero codewords in a random code can be calculated by

$$\begin{aligned}\mathbb{E}(n-2w_i) &= \mathbb{E}\left(\hat{\chi}_C(i)\right) \\ &= \sum_{j=0}^{2^k-1} \mathbb{E}(h_j)(-1)^{\langle \mathbf{j},\mathbf{i}\rangle}.\end{aligned}$$

For a positive integer n, we denote by $[n]$ the set $\{0,1,\ldots,n-1\}$. We can calculate the expectations of the pairwise products of nonzero words' weights as follows.

$$\begin{aligned}\mathbb{E}[(n-2w_i)(n-2w_j)] &= \mathbb{E}\left(\hat{\chi}_C(i)\hat{\chi}_C(j)\right) \\ &= \sum_{s,m\in[2^k]} \mathbb{E}(h_s h_m)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle+\langle \mathbf{m},\mathbf{j}\rangle}.\end{aligned}$$

Let us denote by $P(h_s)$, the probability that a random binary $k\times n$ matrix has exactly h_s columns that are equal to $\mathbf{s}^T$, for $0 \leq s \leq 2^k-1$. Similarly, let $P(h_s,h_m)$ denote the probability that a random matrix has exactly h_s columns equal to $\mathbf{s}^T$ and h_m columns equal to $\mathbf{m}^T$, for $0 \leq s,m \leq 2^k-1$ and $s \neq m$.

Proposition 1. *For a random binary $k\times n$ matrix, we have*

$$P(h_s) = \binom{n}{h_s}\frac{(2^k-1)^{(n-h_s)}}{2^{kn}}$$

and

$$P(h_s,h_m) = \binom{n}{h_s+h_m}\binom{h_s+h_m}{h_m}\frac{(2^k-2)^{(n-(h_s+h_m))}}{2^{kn}}.$$

Proof. The number of matrices with exactly h_s columns equal to $\mathbf{s}^T$ is given by

$$\binom{n}{h_s}(2^k-1)^{(n-h_s)}.$$

The binomial coefficient gives the number of different choices for placing $\mathbf{s}^T$ and the factor $(2^k-1)^{(n-h_s)}$ counts the number of ways to fill the other columns with the remaining (2^k-1) vectors arbitrarily. The number of $k\times n$ binary matrices is 2^{kn}. So we have

$$P(h_s) = \binom{n}{h_s}\frac{(2^k-1)^{(n-h_s)}}{2^{kn}}.$$

Now let $0 \leq s,m \leq 2^k-1$ and $s \neq m$. For the number of $k\times n$ binary matrices with exactly h_s columns equal to $\mathbf{s}^T$ and h_m columns equal to $\mathbf{m}^T$, we fix (h_s+h_m) positions for $\mathbf{s}^T$ and $\mathbf{m}^T$. Hence we have the binomial coefficient $\binom{n}{h_s+h_m}$. Then we determine the h_m positions among the (h_s+h_m) columns for the vector $\mathbf{m}^T$. The number of choices

for that is $\binom{h_s+h_m}{h_m}$. The choice for columns $\mathbf{s}^T$ is determined by that of $\mathbf{m}^T$ uniquely. The columns other than these $(h_s + h_m)$ positions can be filled by the remaining $(2^k - 2)$ vectors arbitrarily. So we have

$$P(h_s, h_m) = \binom{n}{h_s + h_m}\binom{h_s + h_m}{h_m}\frac{(2^k-2)^{(n-(h_s+h_m))}}{2^{kn}}.$$

Theorem 1. *Let C be a random binary linear $[n, k]$ code with words and weights denoted as above. Then the expectations of the weights of nonzero codewords are given by*

$$\mathbb{E}(w_i) = \frac{n}{2} \text{ for } i \geq 1. \tag{3}$$

Proof. By Lemma 2, we have

$$\mathbb{E}(n - 2w_i) = \mathbb{E}\left(\hat{\chi}_C(i)\right) = \sum_{s\in[2^k]} \mathbb{E}(h_s)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle}.$$

By Proposition 1 we proceed as follows:

$$\begin{aligned}
\mathbb{E}(n - 2w_i) &= \sum_{s\in[2^k]} \mathbb{E}(h_s)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle} \\
&= \sum_{s\in[2^k]} \sum_{h_s\in[n+1]} P(h_s)h_s(-1)^{\langle \mathbf{s},\mathbf{i}\rangle} \\
&= \sum_{s\in[2^k]} \sum_{h_s\in[n+1]} \frac{\binom{n}{h_s}(2^k-1)^{n-h_s}}{2^{kn}}h_s(-1)^{\langle \mathbf{s},\mathbf{i}\rangle} \\
&= \sum_{s\in[2^k]} \sum_{h_s\in[n+1]} \frac{\binom{n}{h_s}(2^k-1)^{h_s}}{2^{kn}}(n-h_s)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle}
\end{aligned}$$

In the last equality we performed a change of variables and put $(n - h_s)$ for h_s, by the symmetry of the binomial coefficients. Let $K_1 = 2^k - 1$. In the following equation we have derivative with respect to K_1.

$$\begin{aligned}
\mathbb{E}(n - 2w_i) &= \sum_{s\in[2^k]} \left(n - \frac{K_1}{2^{kn}}\left[(K_1+1)^n\right]'\right)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle} \\
&= \left(n - \frac{K_1}{2^{kn}}\left[n(2^k)^{n-1}\right]\right)\sum_{s\in[2^k]} (-1)^{\langle \mathbf{s},\mathbf{i}\rangle} \\
&= \left(\frac{n}{2^k}\right)\sum_{s\in[2^k]} (-1)^{\langle \mathbf{s},\mathbf{i}\rangle} \\
&= 0 \quad \text{for} \quad i \geq 1.
\end{aligned}$$

The last equality is obtained by Lemma 3. So the expectations of the nonzero weights follow

$$\mathbb{E}(w_i) = \frac{n}{2} \text{ for } i \geq 1.$$

We will explore the expectations of the products of nonzero weights in the next theorem.

Theorem 2. *Expectations of the pairwise products of nonzero weights in a random binary linear $[n, k]$ code are given by*

$$\mathbb{E}(w_i w_j) = \begin{cases} \frac{n^2+n}{4}, & \text{for } i = j,\ i, j \geq 1, \\ \frac{n^2}{4}, & \text{for } i \neq j,\ i, j \geq 1. \end{cases} \tag{4}$$

Proof. By Lemma 2, the expectations of the products of nonzero weights are given by

$$\begin{aligned} \mathbb{E}((n-2w_i)(n-2w_j)) &= \mathbb{E}\left(\hat{\chi}_C(i)\hat{\chi}_C(j)\right) \\ &= \sum_{s,m\in[2^k]} \mathbb{E}(h_s h_m)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle + \langle \mathbf{m},\mathbf{j}\rangle} \\ &= \sum_{s\in[2^k]} \mathbb{E}(h_s^2)(-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle} \\ &\quad + \sum_{\substack{s,m\in[2^k] \\ s\neq m}} \mathbb{E}(h_s h_m)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle + \langle \mathbf{m},\mathbf{j}\rangle}. \end{aligned}$$

The first sum in the last equality is the contribution of cases where $s = m$ and the second sum is the contribution of the cases where $s \neq m$. Now let K_1 and K_2 denote $2^k - 1$ and $2^k - 2$ respectively.

$$\begin{aligned} \mathbb{E}((n-2w_i)(n-2w_j)) = &\frac{1}{2^{kn}} \sum_{\substack{s\in[2^k] \\ h_s\in[n+1]}} \binom{n}{h_s} K_1^{(n-h_s)} h_s^2 (-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle} \\ &+ \sum_{\substack{s,m\in[2^k] \\ s\neq m \\ h_s,h_m\in[n+1]}} \frac{\binom{n}{h_m\ h_s} K_2^{(n-h_s-h_m)} h_s h_m (-1)^{\langle \mathbf{s},\mathbf{i}\rangle + \langle \mathbf{m},\mathbf{j}\rangle}}{2^{kn}}, \end{aligned}$$

where

$$\binom{n}{h_m\ h_s} = \frac{n!}{(h_m)!(h_s)!(n-(h_m+h_s))!}$$

is the multinomial coefficient. We continue with a change of variables $T = h_s + h_m$ in the second sum and we get

$$\mathbb{E}((n-2w_i)(n-2w_j)) = \frac{1}{2^{kn}} \sum_{\substack{s\in[2^k] \\ h_s\in[n+1]}} \binom{n}{h_s} K_1^{(n-h_s)} h_s^2 (-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle}$$

$$
\begin{aligned}
&+ \sum_{\substack{s,m\in[2^k]\\ s\neq m\\ T\in[n+1]\\ h_m\in[T+1]}} \frac{\binom{n}{T}\binom{T}{h_m}K_2^{(n-T)}(T-h_m)h_m(-1)^{\langle \mathbf{s},\mathbf{i}\rangle+\langle \mathbf{m},\mathbf{j}\rangle}}{2^{kn}} \\
=&\frac{1}{2^{kn}} \sum_{\substack{s\in[2^k]\\ h_s\in[n+1]}} \binom{n}{h_s} K_1^{(n-h_s)}(h_s^2-h_s)(-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle} \\
&+ \frac{1}{2^{kn}} \sum_{\substack{s\in[2^k]\\ h_s\in[n+1]}} \binom{n}{h_s} K_1^{(n-h_s)}h_s(-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle} \\
&- \sum_{\substack{s,m\in[2^k]\\ s\neq m\\ T\in[n+1]\\ h_m\in[T+1]}} \frac{\binom{n}{T}\binom{T}{h_m}K_2^{(n-T)}(h_m^2-h_m)(-1)^{\langle \mathbf{s},\mathbf{i}\rangle+\langle \mathbf{m},\mathbf{j}\rangle}}{2^{kn}} \\
&+ \sum_{\substack{s,m\in[2^k]\\ s\neq m\\ T\in[n+1]\\ h_m\in[T+1]}} \frac{\binom{n}{T}\binom{T}{h_m}K_2^{(n-T)}(T-1)h_m(-1)^{\langle \mathbf{s},\mathbf{i}\rangle+\langle \mathbf{m},\mathbf{j}\rangle}}{2^{kn}}
\end{aligned} \tag{5}
$$

After summation over h_s, h_m and T we obtain the following, where the derivatives are taken with respect to the variables u and v. After derivatives we substitute $u=1$ and $v=2$.

$$
\begin{aligned}
\mathbb{E}((n-2w_i)(n-2w_j)) &= \frac{1}{2^{kn}} \sum_{s\in[2^k]} [(K_1+u)^n]''(-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle} \\
&\quad + \frac{1}{2^{kn}} \sum_{s\in[2^k]} [(K_1+u)^n]'(-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle} \\
&\quad + \frac{1}{2^{kn}} \sum_{\substack{s,m\in[2^k]\\ s\neq m}} [(K_2+v)^n]''(-1)^{\langle \mathbf{s},\mathbf{i}\rangle+\langle \mathbf{m},\mathbf{j}\rangle)} \\
&= \frac{n(n-1)}{2^{2k}} \sum_{s\in[2^k]} (-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle}
\end{aligned}
$$

$$+\frac{n}{2^k}\sum_{s\in[2^k]}(-1)^{\langle \mathbf{s},(\mathbf{i}+\mathbf{j})\rangle}$$
$$+\frac{n(n-1)}{2^{2k}}\sum_{\substack{s,m\in[2^k]\\ s\neq m}}(-1)^{\langle \mathbf{s},\mathbf{i}\rangle+\langle \mathbf{m},\mathbf{j}\rangle}$$

By Lemma 3 the result is

$$\mathbb{E}((n-2w_i)(n-2w_j))=\begin{cases} n, & \text{for } i=j,\ i,j\geq 1,\\ 0, & \text{for } i\neq j,\ i,j\geq 1.\end{cases}$$

Finally we get the expectations of products of nonzero codeword weights as

$$\mathbb{E}(w_iw_j)=\begin{cases} \frac{n^2+n}{4}, & \text{for } i=j,\ i,j\geq 1,\\ \frac{n^2}{4}, & \text{for } i\neq j,\ i,j\geq 1.\end{cases}$$

Two random variables X and Y are called statistically uncorrelated if the covariance between them is 0,

$$\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)=0.$$

We observe that this is the case for the weights of distinct nonzero codewords in a random code.

Corollary 1. *Weights of distinct nonzero words of a random binary linear code are uncorrelated*

$$\operatorname{cov}(w_i,w_j)=0,\ \text{for all } i,j\geq 1 \text{ and } i\neq j.$$

4. Correlation Properties of Nonzero Words in a Random Binary Liner Code

Let $\mathbf{c}=(c_1,c_2,\ldots,c_n)$ and $\mathbf{e}=(e_1,e_2,\ldots,e_n)$ be two vectors in $\mathbb{F}_2^n$. We define the periodic auto-correlation and cross-correlation functions on vectors of $\mathbb{F}_2^n$ as follows

$$r_{\mathbf{c},\mathbf{c}}(u) = r_{\mathbf{c}}(u)=\sum_{i=1}^{n}(-1)^{c_i+c_{i+u}} \text{ and}$$
$$r_{\mathbf{c},\mathbf{e}}(u) = \sum_{i=1}^{n}(-1)^{c_i+e_{i+u}}$$

respectively. The indices are evaluated modulo n.

Given $\mathbf{e}\in\mathbb{F}_2^n$ let us denote by $\mathbf{e}^u$, the vector obtained by shifting each coordinate of e by u to the left

$$\mathbf{e}^u=(e_{1+u},e_{2+u},\ldots,e_n,e_1,\ldots,e_u).$$

The correlation functions $r_{\mathbf{c}}$ and $r_{\mathbf{c},\mathbf{e}}$ evaluates the following: If in a coordinate the vectors agree, then this coordinates contributes a $(+1)$ and if the the entries in the same coordinate are different this coordinate contributes a (-1). So we can rewrite the correlation functions as follows

$$r_{\mathbf{c}}(u) = \sum_{i=1}^{n}(-1)^{c_i+c_{i+u}} = n - 2d_H(\mathrm{c}, \mathrm{c}^u) \text{ and} \tag{6}$$

$$r_{\mathbf{c},\mathbf{e}}(u) = \sum_{i=1}^{n}(-1)^{c_i+e_{i+u}} = n - 2d_H(\mathrm{c}, \mathrm{e}^u), \tag{7}$$

where d_H is the Hamming distance. The following formulation of Hamming distance on $\mathbf{F}_2^n$ will be useful

$$d_H(\mathbf{c}, \mathbf{e}) = \|\mathbf{c}\| + \|\mathbf{e}\| - 2\langle \mathbf{c}, \mathbf{e}\rangle, \tag{8}$$

where $\|\cdot\|$ denotes the weight of the vector and the inner product is evaluated over $\mathbb{R}$.

The following theorem gives the expectations of sums of auto-correlations and sums of cross-correlations between the nonzero words in a random binary linear code.

Theorem 3. *Let C be a random binary linear $[n, k]$ code and let $\mathbf{c}$ and $\mathbf{e}$ be two distinct nonzero codewords. Then we have*

$$\sum_{u=1}^{n-1}\mathbb{E}(r_{\mathbf{c}}(u)) = 0 \text{ and}$$

$$\sum_{u=0}^{n-1}\mathbb{E}(r_{\mathbf{c},\mathbf{e}}(u)) = 0.$$

Proof. By (6) and (8) we have

$$\begin{aligned} r_{\mathbf{c}}(u) &= n - 2d_H(\mathbf{c}, \mathbf{c}^u) \\ &= n - 2(\|\mathbf{c}\| + \|\mathbf{c}^u\| - 2\langle \mathbf{c}, \mathbf{c}^u\rangle). \end{aligned}$$

Summation over u and taking expectations of both sides gives

$$\begin{aligned} \sum_{u=1}^{n-1}\mathbb{E}(r_{\mathbf{c}}(u)) &= \sum_{u=1}^{n-1}[n - 2(\mathbb{E}(\|\mathbf{c}\|) + \mathbb{E}(\|\mathbf{c}^u\|) - 2\mathbb{E}(\langle \mathbf{c}, \mathbf{c}^u\rangle))] \\ &= -n(n-1) + 4\mathbb{E}\left(\sum_{u=1}^{n-1}\langle \mathbf{c}, \mathbf{c}^u\rangle\right). \end{aligned}$$

In the last equation we made use of (3). Now we will utilize the following simple fact: For any $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ and $\mathbf{e} = (e_1, e_2, \ldots, e_n)$ in $\mathbb{F}_2^n$ we have

$$\|\mathbf{c}\|\|\mathbf{e}\| = (c_1 + c_2 + \ldots + c_n)(e_1 + e_2 + \ldots + e_n)$$

$$= \sum_{u=0}^{n-1} \langle \mathbf{c}, \mathbf{e}^u \rangle. \tag{9}$$

With this fact, (3) and (4), we have

$$\begin{aligned} \sum_{u=1}^{n-1} \mathbb{E}(r_{\mathbf{c}}(u)) &= -n(n-1) + 4\left(\mathbb{E}(\|\mathbf{c}\|\|\mathbf{c}\|) - \mathbb{E}(\|\mathbf{c}\|)\right) \\ &= -n(n-1) + 4\left(\frac{n^2+n}{4} - \frac{n}{2}\right) \\ &= 0. \end{aligned} \tag{10}$$

Now let $\mathbf{c}$ and $\mathbf{e}$ be two distinct nonzero codewords in C. For the cross-correlations, following the same steps as above we get

$$\begin{aligned} \sum_{u=0}^{n-1} \mathbb{E}(r_{\mathbf{c},\mathbf{e}}(u))) &= \sum_{u=0}^{n-1} \mathbb{E}[n - 2d_H(\mathbf{c}, \mathbf{e}^u)] \\ &= \sum_{u=0}^{n-1} [n - 2(\mathbb{E}(\|\mathbf{c}\|) + \mathbb{E}(\|\mathbf{e}^u\|) - 2\mathbb{E}(\langle \mathbf{c}, \mathbf{e}^u \rangle))] \\ &= -n^2 + 4\mathbb{E}\left(\sum_{u=0}^{n-1} \langle \mathbf{c}, \mathbf{e}^u \rangle\right) \\ &= -n^2 + 4\mathbb{E}(\|\mathbf{c}\|\|\mathbf{e}\|). \end{aligned}$$

Finally by (4) we get the result

$$\sum_{u=0}^{n-1} \mathbb{E}(r_{\mathbf{c},\mathbf{e}}(u)) = 0.$$

5. Conclusion

We obtained the expectations of single and pairwise products of weights of codewords in random binary linear codes. We showed that the weights of two nonzero words are statistically uncorrelated. We used the expectations of the products of nonzero weights to show that the expectations of the sums of out of phase auto-correlations and the expectations of the sums of cross-correlations of nonzero words are zero.

References

[1] V Blinovsky, U Erez, and S Litsyn. Weight distribution moments of random linear/coset codes. *Design. Code. Cryptogr.*, 57:127–138, 2010.

[2] I Bouyukliev, S Bouyuklieva, T Maruta, and P Piperkov. Characteristic vector and weight distribution of a linear code. *Cryptogr. Commun.*, 13:263–282, 2021.

[3] C Carlet. *Boolean Functions for Cryptography and Coding Theory.* Cambridge University Press, Cambridge, 2021.

[4] R G Gallager. *Low Density Parity Check Codes.* MIT Press, Cambridge, 1963.

[5] K J Horadam. *Hadamard Matrices and Their Applications.* Princeton University Press, 2012.

[6] A A Klyachko and I. Özen. Correlations between the ranks of submatrices and weights of random codes. *Finite Fields T. App.*, 15(4):497–516, 2009.

[7] N Linial and J Mosheiff. On the weight distribution of random binary linear codes. *Random Struct. Alg.*, 56:5–36, 2020.

[8] I Özen. Fourth moments of the code-weights. *Commun. Algebra*, 51(4):1761–1771, 2023.

[9] A Samorodnitsky. Weight distribution of random linear codes and krawtchouk polynomials. *Random Struct. Alg.*, 65:261–274, 2024.