Finding Good Binary Linear Block Codes based on Hadamard Matrix and Existing Popular Codes

Driss Khebbou¹, Reda Benkhoya², Idriss Chana³, Hussain Ben-azaaz⁴
Ecole Nationale Supérieure d’Arts et Métiers, Moulay Ismail University, Meknès, Morocco¹, ⁴
Faculty of Sciences, Ibn Tofail University, Kénitra, Morocco²
Ecole Supérieure de Technologie, Moulay Ismail University, Meknès, Morocco³

Abstract—Because of their algebraic structure and simple hardware implementation, linear codes as class of error-correcting codes, are used in a multitude of situations such as Compact disk, backland bar code, satellite and wireless communication, storage systems, ISBN numbers and so more. Nevertheless, the design of linear codes with high minimum Hamming distance to a given dimension and length of the code, remains an open challenge in coding theory. In this work, we propose a code construction method for constructing good binary linear codes from popular ones, while using the Hadamard matrix. The proposed method takes advantage of the MacWilliams identity for computing the weight distribution, to overcome the problem of computing the minimum Hamming distance for larger dimensions.

Keywords—Binary linear codes; code construction; minimum hamming distance; error-correcting codes; weight distribution; coding theory; hadamard matrix

I. INTRODUCTION

The basic digital communication chain includes a source, a communication channel, and a receiver. The message is sent from the source to the receiver through a channel. Unless there is an ideal channel, interference will corrupt the message and cause errors, which can be controlled by an error-correcting code. Thus, inner code redundancy is added to the original message downstream of the source. In fact, this redundancy upstream of the receiver is used to correct potential errors without retransmission.

In his fundamental article [1], Shannon showed via his channel coding theorem, the existence of error-correcting codes (ECC), theoretically allowing to transmit data in a channel with a small probability of error, whatever the noise level in the channel. However, the theorem does not specify how to create these codes. Thus the issue of implementing good error-correcting codes remains open in the field of information theory [2]. Great effort has been constantly devoted to constructing error-correcting codes to totally or almost achieve the channel capacity, following Shannon's work. In this way, Arikan developed the first codes (polar codes) with proven capacity, explicit construction, and low coding and decoding complexity [3], with the implementation of their multi-kernel designs [4]. This paper’s inspiration comes from the coding process of polar code.

It is difficult to construct explicitly good codes with the best properties. Therefore, working with the already existing codes, with good properties, could be one of construction alternatives [5]. Thus to determine if the code would be good enough, Markus Grassl made a bounds database [6] for the minimum distance of linear block codes over GF(q), with q ≤ 9, for given length and dimension, including construction details. Hence, if its parameters allow the current bounds to be achieved, the code is called 'good'.

One of the most recent methods to construct good binary linear block codes is presented in [7]. It consists in constructing linear codes from the Hadamard matrix and Bose–Chaudhuri–Hocquenghem (BCH) codes [8]. However, this method suffers from the problem of computing the minimal Hamming distance for higher code dimensions and it is used only for BCH codes. In this paper, a new method to produce good binary linear bloc codes based on the Hadamard matrix and some popular error-correcting codes often used in coding theory [9], [10] is presented. it allows to design many good binary linear block codes with considerable error-correcting capability. This method extends the approach presented in [7] for larger dimensions by exploiting the MacWilliams identity to overcome the problem of computing the minimal distance on the one hand, and to confirm the technique for codes other than BCH codes [8] on the other hand.

The remainder of this paper is structured as follows. In the next section, we detail some of the concepts required in this work, such as linear block codes, dual code of linear block code, MacWilliams identity, and Hadamard matrices. We present a new method of searching good binary linear codes in the third section. In the fourth section, we improve the proposed method by the set of good binary linear block codes found. Finally, we give an interpretation of the results before concluding the paper.

II. NOTATION AND PRELIMINARIES

In digital transmission, binary error-correcting codes denoted as [n, k, d_min], can be employed to limit the incidence of word errors. Converting a k-bit word to an n-bit codeword (n > k), is the coding process. This conversion creates a code C with 2^k n -bit codewords chosen from a set of 2^n codewords. It has three main parameters: the length of codeword n, the dimension of coded block message k and the minimum Hamming distance between codewords d_min. This minimum distance ensures that a codeword will not be transformed, due to noise, into another codeword, and it allows to get the error correction capability.

www.ijacsa.thesai.org
A. Linear Block Codes Theory

A binary linear code is a sub-vector space over \( \mathbb{F}_2 \) with dimension \( k \). The code is a set of \( 2^k \) codewords, each one is a linear combination of the \( k \) basis vectors, that form a \( k \times n \) generator matrix, \( G \in \mathbb{F}_2^{k \times n} \). In other words, the codeword space \( \mathcal{V} \) of the code can be obtained as follow:

\[
\mathcal{V} = \{ c = u \ast G | u \in \mathbb{F}_2^k \}
\]  

(1)

Where \( u = (u_0, u_1, ..., u_k) \) is called the message to be sent, and \( c = (c_0, c_1, ..., c_n) \) is the codeword produced after encoding the message \( u \).

The one-to-one correspondence between messages and codewords is a fundamental force of block codes; thus, a message is successfully retrieved if the decoder identifies its codewords is an equivalent codeword. So, the minimum Hamming distance parameter of a code allows defining a difference limit between two valid codewords. It is the outcome of:

\[
d_{\text{min}}(C) = \min\{d(c, c') : c, c' \in \mathcal{C} \text{ and } c \neq c'\}
\]  

(2)

In the case of binary linear block codes, the minimum Hamming distance is equivalent to the smallest non-zero weight of a codeword of \( \mathcal{C} \), so that the weight of a codeword \( c \) is the number of its non-zero symbols. It is defined as:

\[
w(c_i) = \begin{cases} 1 & \text{if } c_i \neq 0 \\ 0 & \text{if } c_i = 0 \end{cases}
\]

\[
w(c) = \sum_{i=1}^{n} w(c_i)
\]  

(3)

Another way to define a linear code is to use a matrix \( H \in \mathbb{F}_2^{n \times (n-k)} \) called parity-check matrix, which yields:

\[
\mathcal{C} = \{(c_1, c_2, ..., c_n) | (c_1, c_2, ..., c_n) \ast H^T = 0\}
\]  

(4)

So, for each linear block code \( C(n,k,d_{\text{min}}) \) defined by its generator matrix whose rows structure a basis of a linear vector subspace, another linear block code exists. It is called dual code \( C^\perp \), known by length \( n \), dimension \( (n-k) \), and the vector space consisting of all orthogonal vectors (codewords) with the linear code \( C \) vectors. This means that two n-tuples \( x \) and \( y \) are orthogonal if their inner product is zero:

\[
(x,y) = \sum_{i=1}^{n} (x_i \ast y_i) = 0
\]  

(5)

If \( G = [I_k \mid P] \) is the generator matrix of a linear code \( C(n,k,d_{\text{min}}) \) in the systematic form, then the generator matrix of its dual code is called parity-check matrix, such as:

\[
H = [P^\perp \mid I_{n-k}]
\]  

(6)

B. Weight Distribution and MacWilliams Identity

As mentioned above, the minimum distance is the lower weight \( w(c) \) as defined in (3), of a nonzero codeword among all of the \( 2^k \) codewords in linear code. The importance of this parameter lays in the error correction capacity of the code through \( d_{\text{min}} = 2t + 1 \), where \( t \) denotes the number of errors that the code is capable of correcting. However, the minimum distance does not give an idea about the other codewords’ weight.

Acquiring knowledge of a code’s weight distribution is essential and allows the computation of its analytical performance [11]. The weight distribution of an error-correcting code is a vector of size \( n \) whose \( i^{th} \) element indicates the number of codewords having the weight \( (i-1) \). Otherwise, the weight distribution can be expressed in polynomial form as follows:

\[
W(z) = w_0 + w_1 z + \cdots + w_n z^{n-1}
\]  

(7)

where \( w_i \) is the number of codewords with weight \( i \) obtained by (3).

Although the weight distribution does not inherently identify a code, it provides useful information that has both practical and theoretical significance. MacWilliams equation [12], a series of linear relations between the weight distributions of a code and its dual, is one of the most fundamental conclusion in weight distributions.

Let \( C \) be a \( (n,k,d) \) linear code over \( \mathbb{F}_q^n \) with enumerator polynomial \( W(z) = \sum_{i=0}^{n} w_i z^i \), and let \( W^\perp(z) \) be the enumerator polynomial of the dual code \( C^\perp \). Then:

\[
W^\perp(z) = q^{-k}(1 + (q - 1)z)^n W(\frac{1-z}{1+(q-1)z})
\]  

(8)

C. Hadamard Matrix

The Hadamard matrix \( H_m \) is a square matrix of order \( m \), with \( m \) being a power of 2, and entries in \( \{-1, +1\} \) as:

\[
H_m H_m^T = m I_m
\]  

(9)

Sylvester presented the first examples of these matrices in 1867 [13], before naming them Hadamard matrices in 1893 [14], after Hadamard who generalized them for orders other than \( 2^m \). Many employments for these matrices have been found in telecommunications and signal processing. In fact, the use of Hadamard matrices to construct efficient error-correcting codes is one of the reasons that increased interest in discovering new Hadamard matrix constructions.

In a binary case, we can replace \( \{-1, +1\} \) of \( H_m \) by \( \{1, 0\} \) then \( H_m \) is obtained by the following technique:

\[
H_1 = [0] \equiv [1]
\]

\[
H_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}
\]

\[
H_m = H_2 \otimes H_{m/2}
\]  

(10)

where \( \otimes \) denotes the Kronecker product.

The orthogonality of the Hadamard matrix (9) guarantees that each permutation of rows or columns yields another Hadamard matrix [15].

III. NEW METHOD TO FIND GOOD BINARY LINEAR CODES

In [7], a method based on the outcome of the Kronecker product, between the Hadamard matrix and the redundant part of a generator matrix of a Bose, Ray-Chaudhuri et Hocquenghem (BCH) code is presented, to construct good binary linear codes. It allows us, from a \( (n,k,d_{\text{min}}) \) BCH code and a Hadamard matrix of order \( m \), to build good binary linear codes having a given dimension \( k' < 20 \) and length \( n' = m \ast n \). However, for higher dimensions, this approach has a problem to calculate the minimum Hamming distance, it is one of the open problems [16] in the field of information theory for large dimensions.
So for dimensions \( k' > 20 \), the method presented in [7] remains restricted according to the performance of a simple computer to calculate the minimum distance for codes with dimensions greater than 20. In this work, we practically took advantage of the dual properties of linear block codes and MacWilliams identity as it can be seen in figure 1 and outlined in the steps bellow, in order to fix this issue and validate the process by constructing good codes with high dimensions.

The technique consists of treating the minimum Hamming distance computation problem of the larger dimensions by searching good binary linear codes via their dual codes, with small dimensions, and calculating the weight distribution obtained using the identity of MacWilliams identity as described in (8). By definition, the minimum Hamming distance of a linear code corresponds to the smallest weight of its codewords, so it is obvious to extract the minimum distance of a linear code from its weight distribution, it corresponds to the index of the first non-null element of the weight distribution of a linear code (first element excluded, because it corresponds to the zero's codeword).

The details of the method we propose to improve the dimensions of the constructed good binary linear codes are developed in the following steps. Let’s use:

- \( P \): \( n * (n − k) \) matrix extracted from a generator matrix of the popular used code in the systematic form.
- \( LB \): Lower bound is the best-known minimum distance found in all pre-existing works.
- \( H_m \): Hadamard matrix of order \( m \).
- \( k' \): Dimension of the desired code to be built.
- \( H \): Parity check matrix constructed as described in (6).
- \( C^\perp \): Dual code constructed from the parity check matrix \( H \).
- \( \text{parityCheckMatrix}() \): Function to transform a generator matrix to parity check matrix.
- \( A_{filtered} \) the matrix \( A \) after the elimination of unnecessary rows (rows whose weight is less than \( LB \)).

**Inputs:** \( P, k', LB \)

**Outputs:** List of \( (n', k', d') \) binary linear codes.

Fig. 1. New Method to Find Good Binary Linear Codes using MacWilliams Identity to Reduce the Complexity of the Minimum Distance Computation.
is 3 and it contains 16 codewords of weight 3.

For example, to build a code with dimension $k' = 8$, proceeding to the construction of a code with $k' = 4$. In other words, it would be sufficient to check-in a space of size $2^4$ instead of searching in a space of dimension $2^8$. From [17], the best-known minimum distance (LB) for $n' = 12$ and $k' = 8$ is 3, so $A_{filtered}$ will be obtained by eliminating all rows with a weight less than 3 as defined in (11).

By combining 4 rows of $A_{filtered}$ as a generator matrix of a suspect $(12,4,x)$ code, calculating the weight distribution of the code and applying the MacWilliams identity, codes with the following weight distribution is obtained:

$$[1,0,0,0,16,39,48,48,48,39,16,0,0,1]$$

Which means that the minimum distance of the linear code is 3 and it contains 16 codewords of weight 3.

IV. Experimental Results

Three types of results are presented in this section; the first one is obtained by the new method mentioned in the previous section, the second is an extension of [7] for the Golay and Reed-Muller codes, and the third one is based on the codes of the first result. All programs have been implemented in GAP via the GUAVA package over $\mathbb{F}_2$ and $\mathbb{F}_3$ [18].

A. Results Obtained using the MacWilliams Identity

The method as defined in [7], through a computer calculations with Intel(R) Core(TM) i5-4210U RAM 4 CPU @1.70GHz configuration, does not permit to generate good binary linear codes with a dimension greater than 20. But, for dimensions greater than 20 and using the same computer, the new approach helps us to verify the validity of the concept, for dimensions greater than 20, and it allowed us to find new good binary linear codes. Table 1 describes the set of good binary linear codes with larger dimensions $(k > 20)$, built using BCH codes by applying the presented approach.

In [7], it is focused on the construction of good binary linear codes from the Hadamard matrix and BCH codes. In this work, we tried to apply the approach for other codes with good properties. Table 2 describes the good codes constructed from Golay code (23,12).

Applicability of the technique for Reed-Muller codes produced satisfactory results, as shown in table 3.

| TABLE I. GOOD BINARY LINEAR CODES USING BCH CODES |
|-----------------------------------------------|
| Rate  | Code [n, k, d]   | $d_{magma}$ | Lower bound |
|-------|-----------------|-------------|-------------|
| 0.86  | [30,26,2]       | 2           | 2           |
| 0.9   | [30,27,2]       | 2           | 2           |
| 0.93  | [30,28,2]       | 2           | 2           |
| 0.71  | [32,23,4]       | 4           | 4           |
| 0.75  | [32,24,3]       | 3           | 4           |
| 0.84  | [32,27,2]       | 2           | 2           |
| 0.87  | [32,28,2]       | 2           | 2           |
| 0.69  | [36,25,4]       | 4           | 5           |
| 0.72  | [36,26,4]       | 4           | 4           |
| 0.78  | [37,29,3]       | 3           | 4           |
| 0.76  | [38,29,4]       | 4           | 4           |
| 0.55  | [40,22,7]       | 7           | 8           |
| 0.6   | [40,24,7]       | 7           | 7           |
| 0.62  | [40,25,6]       | 6           | 6           |
| 0.72  | [40,29,4]       | 4           | 5           |
| 0.75  | [40,30,3]       | 3           | 4           |
| 0.78  | [60,47,6]       | 6           | 6           |
| 0.83  | [60,50,3]       | 3           | 4           |
| 0.81  | [60,49,4]       | 4           | 4           |
| 0.84  | [78,66,4]       | 4           | 4           |

| TABLE II. GOOD BINARY LINEAR CODES USING GOLAY CODE |
|-----------------------------------------------|
| Rate  | Code [n, k, d]   | $d_{magma}$ | Lower bound |
|-------|-----------------|-------------|-------------|
| 0.86  | [22,14,4]       | 4           | 4           |
| 0.68  | [22,15,4]       | 3           | 4           |
| 0.9   | [22,17,3]       | 3           | 3           |
| 0.93  | [22,18,2]       | 2           | 2           |
| 0.71  | [22,20,2]       | 2           | 2           |

| TABLE III. GOOD BINARY LINEAR CODES USING REED-MULLER CODES |
|-----------------------------------------------|
| Rate  | Code [n, k, d]   | $d_{magma}$ | Lower bound |
|-------|-----------------|-------------|-------------|
| 0.5   | [16,8,5]        | 5           | 5           |
| 0.56  | [16,9,3]        | 3           | 4           |
| 0.75  | [16,12,2]       | 2           | 2           |
| 0.81  | [16,13,2]       | 2           | 2           |
| 0.63  | [22,14,4]       | 4           | 4           |
| 0.59  | [22,13,4]       | 4           | 5           |
| 0.81  | [22,18,2]       | 2           | 2           |
B. Good Extended and Punctured Binary Linear Codes

Extending and puncturing are two methods of code construction [19], which maintain the code dimension $k$ while varying its length $n$. In the case of extending code, parity bits are added, which can contribute to increase a minimum distance. Whereas puncturing removes parity bits, which can lead to decrease a minimum distance. Let $C_{\text{ext}}(n + 1, k)$ a binary linear code who is the extended code of the linear $C(n, k)$. The extension is completed by adding a new coordinate (parity check bit) to each codeword of $C$ so that the codeword length goes up. Put differently, each codeword $v_{\text{ext}} = (v_1, v_2, ..., v_n, v_{n+1})$ of the extended code $C_{\text{ext}}$ is generated by attaching a coordinate to the codeword $v = (v_1, v_2, ..., v_n)$ from $C$, in order that $v_{n+1} = \sum_{i=1}^{n} v_i$, where sum is modulo 2 addition in binary case.

In this reflection, new good codes were defined by applying the extending and puncturing to the good codes mentioned in Tables 1, 2, and 3, as well as to the codes contained in related previous work [7]. Table 4 shows all the good extended and punctured binary linear codes found.

C. Interpretation

Lately, error-correcting code designers have been concerned with finding a high code rate which is defined as the ratio of the number of information symbols $k$ to the length of codeword $n$, to take maximum advantage of the capacity of the channel.

In this work, the focus is on error-correcting codes with a rate greater than 0.5. Most of the constructed codes have a minimum Hamming distance equal to the lower bound, allowing us to identify them as well as good binary linear error-correcting codes. In some of the results above, for given $n'$ and $k'$, most of the codes are found with the same wanted minimum distance (LB) existing in [6], and the chosen one is the one with the smallest number of codewords with minimum weight in the weight distribution. However, it should be mentioned that just a few codes with the lower limit have been reported in the literature for the codes that did not achieve the LB, and that the research discovered multiple different codes with the lower limit (LB – 1).

In comparison to the results obtained in [7], the technique provided in this paper allows us to construct good binary linear codes with larger dimensions and good properties. Unlike previous research, instead of shedding light on BCH codes only, the strategy yields positive outcomes for a variety of different codes, such as Golay and Reed-Muller codes.

All of the good codes discovered in this and previous research have been validated in software, designed to solve algebra problems MAGMA [20], [21], which supports several coding theories.

The exponential explosion of possible combinations, from $A_{\text{filtered}}$, of suspect codes for higher dimensions continues to be a problem of finding good codes observed during this work. This issue will continue to be a source of reflection in the future. The main objective of this simulation is to demonstrate that the proposed methodology is applicable to larger dimensions as well as codes other than used codes.

V. Conclusion

In this paper, an extension of the method of constructing good linear codes from BCH codes and Hadamard matrices, stated in the literature to higher dimensions and for other popular codes. In this way, a set of good binary linear block codes were discovered by exploring the duality of linear codes and MacWilliams identity on the one hand, and by extending and puncturing the discovered results on the other. The majority of the found codes match the bound of the existing codes in the literature. The search issue for good error-correcting code search problem is very large for most standard search techniques. In this case, and to overcome the problem of the exponential explosion of the number of combinations, genetic algorithms can be an efficient way to find good solutions in a relatively short time, and it can be a research direction for future work.

References

[1] C. E. Shannon, “A Mathematical Theory of Communication,” Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[2] D. Joyner and J.-L. Kim, Selected Unsolved Problems in Coding Theory. Birkhäuser Basel, 2011. doi: 10.1007/978-0-8176-8256-9.

[3] E. Arikan, “Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels,” IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3051–3073, Jul. 2009, doi: 10.1109/TIT.2009.2021379.

[4] F. Gabry, V. Bioglio, I. Land, and J.-C. Belfiore, “Multi-kernel construction of polar codes,” in 2017 IEEE International Conference on Communications Workshops (ICC Workshops), May 2017, pp. 761–765. doi: 10.1109/ICC.2017.7962750.

[5] M. Tomlinson, C. J. Tjah, M. A. Ambroze, M. Ahmed, and M. Jibril, “Good Binary Linear Codes,” in Error-Correction Coding and Decoding: Bounds, Codes, Decoders, Analysis and Applications, M. Tomlinson, C. J. Tjah, M. A. Ambroze, M. Ahmed, and M. Jibril, Eds. Cham: Springer International Publishing, 2017, pp. 101–136. doi: 10.1007/978-3-319-51103-0_5.

[6] G. Markus, “‘Bounds on the minimum distance of linear codes and quantum codes.” Online available at http://www.codeables.de. Accessed on 2021-08-04.

[7] D. Khebbou, R. Benkhoya, and I. Chana, “Construction of Some Good Binary Linear Codes Using Hadamard Matrix and BCH Codes,” in Proceedings of Sixth International Congress on Information and Communication Technology, Singapore, 2022, pp. 523–532. doi: 10.1007/978-981-16-2377-6_49.

[8] R. C. Bose and D. K. Ray-Chaudhuri, “On a class of error correcting binary group codes,” Information and Control, vol. 3, no. 1, pp. 68–79, Mar. 1960, doi: 10.1016/S0019-9958(60)90287-4.

[9] I. Reed, “A class of multiple-error-correcting codes and the decoding scheme,” Transactions of the IRE Professional Group on Information...
[10] Marcel Golay, “Notes on Digital Coding.” 1949.
[11] L. Bolcar, “Weights of Linear Codes and their Dual,” Academic Festival, Apr. 2020, [Online]. Available: https://digitalcommons.sacredheart.edu/acadfest/2020/all/102.
[12] J. MacWilliams, “A theorem on the distribution of weights in a systematic code,” The Bell System Technical Journal, vol. 42, no. 1, pp. 79–94, Jan. 1963, doi: 10.1002/j.1538-7305.1963.tb04003.x.
[13] J. J. Sylvester, “LX. Thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to Newton’s rule, ornamental tile-work, and the theory of numbers,” The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 34, no. 232, pp. 461–475, Dec. 1867, doi: 10.1080/1478646708539914.
[14] J. Hadamard, Resolution D’une Question Relative Aux Determinants - in Bulletin des Sciences Mathematiques, Septembre 1893, 1st Edition. See Description, 1893.
[15] A. LaClair, “A Survey on Hadamard Matrices,” Chancellor’s Honors Program Projects, May 2016, [Online]. Available: https://trace.tennessee.edu/utk_chanhonoproj/1971.
[16] M. Askali, A. Azouaoui, S. Nouh, and M. Belkasmi, “On the Computing of the Minimum Distance of Linear Block Codes by Heuristic Methods,” International Journal of Communications, Network and System Sciences, vol. 5, no. 11, Art. no. 11, Nov. 2012, doi: 10.4236/ijcns.2012.511081.
[17] Grassl, Markus, “Bounds on the minimum distance of linear codes and quantum codes.” http://www.codetables.de.
[18] “The GAP Group,” 2015. http://www.gap-system.org.
[19] V. Pless and W. C. Huffman, Eds., “Basic concepts of linear codes,” in Fundamentals of Error-Correcting Codes, Cambridge: Cambridge University Press, 2003, pp. 1–52. doi: 10.1017/CBO9780511807077.002.
[20] W. Bosma, J. Cannon, and C. Playoust, “The Magma Algebra System I: The User Language,” Journal of Symbolic Computation, vol. 24, no. 3, pp. 235–265, Sep. 1997, doi: 10.1006/jsco.1996.0125.
[21] J. Cannon and W. Bosma, “Handbook of Magma Functions,” Jan. 2008.