Доклади на Българската академия на науките Comptes rendus de l'Académie bulgare des Sciences

Tome 74, No 10, 2021

MATHEMATICS

Coding theory

ON THE MAXIMAL CARDINALITY OF BINARY TWO-WEIGHT CODES

Ivan Landjev, Assia Rousseva*

(Submitted by Academician V. Drensky on June 7, 2021)

Abstract

In this note we prove a general upper bound on the size of a binary $(n, \{d_1, d_2\})$ -code with $d_2 > 2d_1$. This bound is used to settle recent conjectures on the maximal cardinality of an $(n, \{2, d\})$ -code. The special case of d = 4 is also resolved using a classical shifting technique introduced by Erdős, Ko and Rado.

Key words: two-weight codes, the Erdős–Ko–Rado theorem, non-linear codes, main problem of coding theory

2010 Mathematics Subject Classification: 05A05, 05A20, 94B25, 94B65

1. Introduction. A binary (nonlinear) code $C \subset \mathbb{F}_2^n$ is called a two-distance code, or, with a certain abuse of language, a two-weight code if the possible distances between two different words of the code take on two different values, i.e.

$$\{d(u,v) \mid u, v \in C, u \neq v\} = \{d_1, d_2\},\$$

where $0 < d_1 < d_2 < n$. Here d(u, v) denotes the Hamming distance between uand v, i.e. the number of positions in which the words u and v are different. A binary two-weight code of length n, cardinality M, and with distances d_1 and d_2

The research of the first author was supported by the Bulgarian National Science Fund under Contract KP-06-32/2 - 07.12.2019.

The research of the second author was supported by the Research Fund of Sofia University under Contract No 80-10-88/25.03.2021.

DOI:10.7546/CRABS.2021.10.01

is called an $(n, M, \{d_1, d_2\})$ -code. If the cardinality is not specified we speak of an $(n, \{d_1, d_2\})$ -code. A natural problem is to determine the maximal cardinality, denoted by $A_2(n, \{d_1, d_2\})$, of a binary two-weight code of fixed length n and given distances d_1 and d_2 .

A systematic investigation of this problem for the non-linear case was made by BOYVALENKOV et al. [¹], where along with proving upper bounds and various facts about $A_2(n, \{d_1, d_2\})$, the authors state two conjectures:

(A)
$$A_2(n, \{2, 4\}) = \binom{n}{2} + 1$$
 for all $n \ge 6$, and that

(B)
$$A_2(n, \{2, d\}) = \begin{cases} n & \text{for } 5 \le d \le n-2, \\ n+1 & \text{for } d = n-1. \end{cases}$$

Let us note that $A_2(4, \{2, 4\}) = 8$ and $A_2(5, \{2, 4\}) = 16$. The optimal codes are the even weight codes. In the case n = 6 we split the words of the even weight code into 16 pairs of complementary words. Any choice of a word from each of these pairs gives a $(6, \{2, 4\})$ -code. For $n \ge 7$ the code consisting of the zero word and all words of weight two gives an $\binom{n}{2} + 1, \{2, 4\}$ -code.

This paper is structured as follows. In Section 2, we prove a upper bound on the size of a binary $(n, \{d_1, d_2\})$ -code which improves on the bound from Theorem 6 in [¹]. In Section 3, we settle Conjecture (B) using ideas from Section 2, which deals with the special case of $d_1 = 2$. In Section 4, we prove Conjecture (A) using a classical shifting technique introduced by ERDŐS, KO and RADO in [²].

2. A bound on the size of a code with two distances. In this section we consider binary codes with parameters $(n, \{d_1, d_2\})$ with $d_2 > 2d_1$. Without loss of generality we assume that the zero word $\mathbf{0} = (\underbrace{0, 0, \ldots, 0}_{-})$ is in C. The

following observation is now straightforward:

- if wt(c_1) = wt(c_2) = d_1 , then wt($c_1 * c_2$) = $d_1/2$ and $d(c_1, c_2) = d_1$;
- if wt(c_1) = d_1 , wt(c_2) = d_2 , then wt($c_1 * c_2$) = $d_1/2$ and $d(c_1, c_2) = d_2$;
- if $\operatorname{wt}(\boldsymbol{c}_1) = \operatorname{wt}(\boldsymbol{c}_2) = d_2$, then

either wt($c_1 * c_2$) = $d_2 - d_1/2$, and $d(c_1, c_2) = d_1$, or else wt($c_1 * c_2$) = $d_2/2$, and $d(c_1, c_2) = d_2$.

Here we denote, as usual, by $c_1 * c_2$ the star of the vectors c_1 and c_2 , i.e. if $c_1 = (\alpha_1, \ldots, \alpha_n), c_2 = (\beta_1, \ldots, \beta_n)$, then

$$\boldsymbol{c}_1 \ast \boldsymbol{c}_2 = (\alpha_1 \beta_1, \alpha_2 \beta_2, \dots, \alpha_n \beta_n).$$

I. Landjev, A. Rousseva

Lemma 1. Let C be an $(n, \{d_1, d_2\})$ -code with $\mathbf{0} \in C$, $d_2 > 2d_1$, and let $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \in C$ be words of weight d_2 . If

$$d(\boldsymbol{c}_1, \boldsymbol{c}_2) = d(\boldsymbol{c}_1, \boldsymbol{c}_3) = d_1,$$

then $d(c_2, c_3) = d_1$.

Proof. Assume for a contradiction that $d(\mathbf{c}_2, \mathbf{c}_3) = d_2$. Denote by α the number of positions in which all three words \mathbf{c}_1 , \mathbf{c}_2 , and \mathbf{c}_3 have 1's. Then by $\operatorname{wt}(\mathbf{c}_1 * \mathbf{c}_3) = d_2 - d_1/2$ we get $d_2 - d_1/2 - \alpha \leq d_1/2$, whence $\alpha > d_1$. On the other hand, $\operatorname{wt}(\mathbf{c}_2 * \mathbf{c}_3) = d_1/2$ implies $d_1/2 - \alpha \geq 0$, a contradiction.

This lemma implies that the graph with vertices – the words of weight d_2 , and neighbourhood between two vertices iff the corresponding words are at distance d_1 , is a union of (possibly trivial) cliques.

Let us denote by A the (M-1)-by-n matrix having as rows the non-zero words of C. By Lemma 1, the words can be ordered in such way that

$$AA^{T} = \begin{pmatrix} \frac{d_{1}}{2}J + \frac{d_{1}}{2}I & \frac{d_{1}}{2}J & \frac{d_{1}}{2}J & \dots & \frac{d_{1}}{2}J \\ \frac{d_{1}}{2}J & (d_{2} - \frac{d_{1}}{2})J + \frac{d_{1}}{2}I & \frac{d_{2}}{2}J & \dots & \frac{d_{2}}{2}J \\ \frac{d_{1}}{2}J & \frac{d_{2}}{2}J & (d_{2} - \frac{d_{1}}{2})J + \frac{d_{1}}{2}I & \dots & \frac{d_{2}}{2}J \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{d_{1}}{2}J & \frac{d_{2}}{2}J & \frac{d_{2}}{2}J & \dots & (d_{2} - \frac{d_{1}}{2})J + \frac{d_{1}}{2}I \end{pmatrix},$$

where the diagonal matrices are of size $k_0 \times k_0, k_1 \times k_1, \ldots, k_s \times k_s$, respectively, where $k_0 + k_1 + \cdots + k_s = M - 1$.

Using standard techniques for computing determinants, one can verify that the determinant of the matrix B is not zero.

Theorem 2. Let C be a binary $(n, M, \{d_1, d_2\})$ -code. Then $M \leq n + 1$.

Proof. By the above argument det $AA^T \neq 0$ and hence the matrix AA^T is of full rank over \mathbb{Q} . Now using the Sylvester inequality, we get

$$M-1 = \operatorname{rank} AA^T \le \operatorname{rank} A \le n$$

which proves the theorem.

Corollary 3. If $d_2 > 2d_1$, we have $A_2(n, \{d_1, d_2\}) \le n + 1$.

Let us note that this theorem improves significantly on the bound given in Theorem 6 from [1], which for the case q = 2 gives $A_2(n, \{d_1, d_2\}) \leq 2n + 1$. Equality in Corollary 3 can be achieved, for instance, if $d_1 = 2$, $d_2 = n - 1$. As we shall prove in the next section, there exist pairs (d_1, d_2) for which this bound can be improved.

3. On $(n, \{2, d\})$ -codes. Using the idea from the previous section, we can tackle the second part of Conjecture 1 from $[^1]$. The authors conjecture there that $A_2(n, \{2, d\}) = n$ for $n \ge 6, 5 \le d \le n - 1$, and $A_2(n, \{2, n - 1\}) = n + 1$.

The construction of an $(n, M = n, \{2, d\})$ -code, as well as of a code of cardinality n + 1 for d = n - 1 is given in [¹]. The upper bound is easily verified in

C. R. Acad. Bulg. Sci., 74, No 10, 2021

the cases of d odd, as demonstrated in [¹]. For d = n - 1 this bound follows from Theorem 2. Below we consider the most interesting case where d is even.

Let us assume that $n \ge 8$ and $6 \le d \le n-2$, d even. Furthermore, let C be an $(n, M, \{2, d\})$ -code with d and n satisfying the above restrictions.

Without loss of generality $\mathbf{0} \in C$. All the remaining words are of weight 2 or d. We denote by a the number of words in C that are of Hamming weight 2. We have $1 \leq a \leq M-2$ since there exist two words at distance 2 and not all distances between different words are equal to 2. As before, denote by A the (M-1)-by-n matrix that has as rows the non-zero words of C. Then up to a row and column permutation A has the following form:

(1)
$$A = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & & \ddots & & & \ddots & & \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \\ \hline 0 & 1 & 1 & \dots & 1 & & & \\ \vdots & & \ddots & & & & & & \\ 0 & 1 & 1 & \dots & 1 & & & \\ \hline 1 & 0 & 0 & \dots & 0 & & B & & \\ \vdots & & \ddots & & & & & & \\ 1 & 0 & 0 & \dots & 0 & & & B & \\ \end{bmatrix}$$

The matrix B is the (M - a - 1)-by-(n - a - 1) matrix formed by the bottom M - a - 1 rows (corresponding to the words of weight d) and the rightmost n - a - 1 columns. We denote by C_i , i = 0, 1, the set of all words of C that are of weight d and have i in the first coordinate.

By Lemma 1 the graph with vertices – the words of weight d, and edges – the pairs of words of weight d that are at distance 2, is a union of (possibly trivial) cliques.

1) Let us first assume that $a > \frac{d}{2}$. Assume that both C_0 and C_1 are nonempty. For $c_0 \in C_0$ and $c_1 \in C_1$, we have

$$d(c_0, c_1) \ge a + 1 + (d - 1) - (d - a) = 2a > d,$$

a contradiction. Thus we have either $C_0 = \emptyset$, or $C_1 = \emptyset$.

(a) Assume $C_1 = \emptyset$. Since every two words from C_0 are obviously at distance 2, we have

$$BB^T = (d - a - 1)J + I,$$

and it is easily checked that det $BB^T \neq 0$. Now we have

$$M - a - 1 = \operatorname{rank} BB^T \le \operatorname{rank} B \le n - a - 1,$$

whence $M \leq n$.

I. Landjev, A. Rousseva

(b) Assume $C_0 = \emptyset$. Now we have

$BB^T =$	$\begin{pmatrix} d-1 & d \\ d-2 & d \end{pmatrix}$	d-2 d-1	d - 2 d - 2							
	\vdots d-2 d	\vdots \cdot $d-2$ \ldots	$d = \frac{1}{2}$		$\frac{d}{2}J$			$\frac{d}{2}J$		
		$\frac{d}{2}$,	J	$\begin{array}{c} d-1\\ \vdots\\ d-2 \end{array}$	····	$\begin{array}{c} d-2\\ \vdots\\ d-1 \end{array}$		$\frac{d}{2}J$		
		$\frac{d}{2}$,	J		$\frac{d}{2}J$		$d-1$ \vdots $d-2$	···· ··.	$\begin{array}{c} d-2\\ \vdots\\ d-1 \end{array}$	
		:			:			÷		·)

It can be proved again that $\det BB^T \neq 0$, and we can repeat the above argument:

 $M - a - 1 = \operatorname{rank} BB^T \le \operatorname{rank} B \le n - a - 1.$

2) Now we consider the case where $1 \le a \le \frac{d}{2}$. The structure of C is again the same as in (1). We keep the notation form 1), i.e. C_0 is the set of all words of weight d that start with 0, and C_1 is the set of all words of weight d that start with 1.

Let us note that if $c_0 \in C_0$, and $c_1 \in C_1$, then $d(c_0, c_1) = d$. This is obvious if $a \ge 2$. Assume that a = 1. Set

$$c_0 = (0, 1, \underbrace{1, \dots, 1}_{d-1}, 0, \dots, 0),$$

 $c_1 = (1, 0, \underbrace{1, \dots, 1}_{d-1}, 0, \dots, 0).$

Consider a word $\mathbf{c}' \in C$. Obviously, $\mathbf{c}' = (0, 1, *, *, \dots, *)$, or $\mathbf{c}' = (1, 0, *, *, \dots, *)$. In both cases, we have

$$|d(c_0, c') - d(c_1, c')| = 2$$

which is impossible since this difference can take on only the values 0 and d-2.

Now we compute again det BB^T which turns out to be not 0. Hence using the chain of inequalities

$$M - a - 1 = \operatorname{rank} BB^T \le \operatorname{rank} B \le n - a - 1$$

we get again $M \leq n$. Thus we have proved the following theorem.

Theorem 4. If C is an $(n, M, \{2, d\})$ -code with $n \ge 8$, $6 \le d \le n-2$, d even, then $M \le n$.

C. R. Acad. Bulg. Sci., 74, No 10, 2021

This implies the validity of Conjecture 1(b) from [¹]: Corollary 5. If $d \ge 5$ then

$$A_2(n, \{2, d\}) = \begin{cases} n & \text{for } 5 \le d \le n-2, \\ n+1 & \text{for } d = n-1. \end{cases}$$

4. The shifting technique. In this section, we consider the case $d_1 = 2$, $d_2 = 4$. The following definition goes back to Erdős, Ko and Rado [²] (see also FRANKL [³]) and is introduced here for binary vectors and binary codes.

Let $C \subset \mathbb{F}_2^n$ and let $\boldsymbol{v} \in \mathbb{F}_2^n$. We denote by $\operatorname{supp}(\boldsymbol{v})$ the set of non-zero coordinate positions of \boldsymbol{v} . So $\operatorname{supp}(\boldsymbol{v})$ can be thought of as a subset of $\{1, \ldots, n\}$. The (i, j)-shift of \boldsymbol{v} is defined by

(2)
$$s_{i,j}(\boldsymbol{v}) = \begin{cases} \boldsymbol{v} + \boldsymbol{e}_i + \boldsymbol{e}_j & \text{if } i \notin \operatorname{supp}(\boldsymbol{v}), \ j \in \operatorname{supp}(\boldsymbol{v}), \ \boldsymbol{v} + \boldsymbol{e}_i + \boldsymbol{e}_j \notin C; \\ \boldsymbol{v} & \text{otherwise.} \end{cases}$$

Here e_i is the unit vector with 1 in position *i*. The (i, j)-shift of a binary code *C* is defined by

(3)
$$S_{i,j}(C) = \{s_{i,j}(\boldsymbol{v}) \mid \boldsymbol{v} \in C\}.$$

Our proof is based on the following lemma.

Lemma 6. Let C be a $(n, \{2, 4\})$ -code. Then $S_{i,j}(C)$ is also an $(n, \{2, 4\})$ -code.

Proof. Let us consider two words $\boldsymbol{u}, \boldsymbol{v} \in C, \, \boldsymbol{u} \neq \boldsymbol{v}$. We have to show that $d(s_{i,j}(\boldsymbol{u}), s_{i,j}(\boldsymbol{v})) \in \{2, 4\}$. We have four possibilities:

- (1) $s_{i,j}(\boldsymbol{u}) = \boldsymbol{u}, \, s_{i,j}(\boldsymbol{v}) = \boldsymbol{v};$
- (2) $s_{i,j}(\boldsymbol{u}) = \boldsymbol{u}, \, s_{i,j}(\boldsymbol{v}) \neq \boldsymbol{v};$
- (3) $s_{i,j}(\boldsymbol{u}) \neq \boldsymbol{u}, \, s_{i,j}(\boldsymbol{v}) = \boldsymbol{v};$
- (4) $s_{i,j}(\boldsymbol{u}) \neq \boldsymbol{u}, \, s_{i,j}(\boldsymbol{v}) \neq \boldsymbol{v}.$

It is clear that in cases (1) and (4), we have $d(s_{i,j}(\boldsymbol{u}), s_{i,j}(\boldsymbol{v})) = d(\boldsymbol{u}, \boldsymbol{v})$. Cases (2) and (3) are similar and are treated in the same way. Hence we shall consider just case (2). Since $s_{i,j}(\boldsymbol{u}) = \boldsymbol{u}$ one of the following must take place:

- (i) $i \notin \operatorname{supp}(\boldsymbol{u}), j \notin \operatorname{supp}(\boldsymbol{u});$
- (ii) $i \in \operatorname{supp}(\boldsymbol{u}), j \in \operatorname{supp}(\boldsymbol{u});$
- (iii) $i \in \operatorname{supp}(\boldsymbol{u}), j \notin \operatorname{supp}(\boldsymbol{u});$
- (iv) $i \notin \operatorname{supp}(\boldsymbol{u}), j \in \operatorname{supp}(\boldsymbol{u}), \boldsymbol{u} + \boldsymbol{e}_i + \boldsymbol{e}_j \in C.$

I. Landjev, A. Rousseva

The vectors \boldsymbol{v} and $s_{i,j}(\boldsymbol{v})$ have exactly one unit in positions i and j in all cases. So, in cases (i) and (ii) $d(\boldsymbol{u}, \boldsymbol{v}) = d(\boldsymbol{u}, s_{i,j}(\boldsymbol{v}))$. Similarly, in case (iii), we get

$$d(\boldsymbol{u}, s_{i,j}(\boldsymbol{v})) = d(\boldsymbol{u}, \boldsymbol{v}) - 2.$$

If $d(\boldsymbol{u}, \boldsymbol{v}) = 4$, then $d(\boldsymbol{u}, s_{i,j}(\boldsymbol{v})) = 2$. If $d(\boldsymbol{u}, \boldsymbol{v}) = 2$, then $d(\boldsymbol{u}, s_{i,j}(\boldsymbol{v})) = 0$, i.e. $\boldsymbol{u} = s_{i,j}(\boldsymbol{v})$ which contradicts the definition of an (i, j)-shift.

Finally, in case (iv)

$$d(u, s_{i,j}(v)) = d(u, v + e_i + e_j) = d(u + e_i + e_j, v) \in \{2, 4\},$$

since $\boldsymbol{u} + \boldsymbol{e}_i + \boldsymbol{e}_j \in C$.

A code C with the property

$$S_{i,j}(C) = C$$

for all i < j is called stable. Clearly, every code can be transformed to a stable code by performing at most $\binom{n}{2}$ shifts, e.g. the shifts $S_{i,j}$ for all pairs i, j with i < j.

Now we are going to prove our main result that implies the exact value of $A_2(n, \{2, 4\})$.

Theorem 7. Let C be a binary $(n, \{2, 4\})$ -code with $n \ge 6$. Then

$$|C| \le \binom{n}{2} + 1.$$

Proof. Assume for a contradiction that C is an $(n, \{2, 4\})$ -code of cardinality

$$|C| > \binom{n}{2} + 1.$$

Because of Lemma 6 we can assume that C is a stable code. Since the case n = 6 was settled in the introduction, we can assume that $n \ge 7$. Since the Hamming metric is translation invariant we can also assume that the zero word is in C. Hence all words in C are of weight 2 or 4.

Denote

$$C_i = \{v = (v_1, \dots, v_n) \in C \mid v_n = i\}, \ i = 0, 1.$$

We shall use induction on the length of C. Therefore we can assume that $|C_0| \leq \binom{n-1}{2}$ which in turn implies that $|C_1| > n-1$.

Assume that $e_i + e_n \in C$, $i \neq n$. Since C is stable it contains also all vectors $e_i + e_j$ for all $j \in \{1, \ldots, n\} \setminus \{i\}$. This implies that all words in C of weight 4 have 1 in position *i*. Otherwise, such a word of weight 4 is at distance 6 from at least one of $e_i + e_j$. This uses the fact that $n \geq 7$. This observation implies immediately that there are at most four words of weight 2 in C_1 . If there are

C. R. Acad. Bulg. Sci., 74, No 10, 2021

exactly four words of weight 2 in C_1 , then C_1 cannot contain a word of weight 4 and hence $|C_1| \leq n-1$. This implies

$$|C| = |C_0| + |C_1| \le 1 + \binom{n-1}{2} + \binom{n-1}{1} = 1 + \binom{n}{2}.$$

Now let there exist exactly three words of weight 2 in C_1 : $e_{i_j} + e_n$, j = 1, 2, 3. Then the only possible word of weight 4 in C_1 is $e_{i_1} + e_{i_2} + e_{i_3} + e_n$ and

$$|C| = |C_0| + |C_1| \le 1 + \binom{n-1}{2} + (1+3) < \binom{n}{2} + 1.$$

Now assume C_1 has two words of weight 4, \boldsymbol{u} and \boldsymbol{v} say. We consider the case where $d(\boldsymbol{u}, \boldsymbol{v}) = 4$, i.e.

$$u = e_{i_1} + e_{i_2} + e_{i_3} + e_n, \quad v = e_{i_1} + e_{j_2} + e_{j_3} + e_n,$$

where i_1, i_2, i_3, j_2, j_3 are all different. Let $k \in \{1, \ldots, n-1\} \setminus \{i_1, i_2, i_3, j_2, j_3\}$ $(n \geq 7)$. Since C is stable we have that $\boldsymbol{w} = \boldsymbol{e}_{i_1} + \boldsymbol{e}_{i_2} + \boldsymbol{e}_{i_3} + \boldsymbol{e}_k \in C$. Now $d(\boldsymbol{w}, \boldsymbol{v}) = 6$, a contradiction. Thus we have proved that if $\boldsymbol{u}, \boldsymbol{v} \in C_1$ and they are both of weight 4, then $d(\boldsymbol{u}, \boldsymbol{v}) = 2$. Now by the Erdős–Ko–Rado theorem the number of the words of weight 4 in C_1 is at most $\binom{n-1-2}{3-2} = n-3$. Hence

$$|C| = |C_0| + |C_1| \le \binom{n-1}{2} + 1 + (n-3) \le \binom{n}{2} + 1.$$

Corollary 8. $A_2(n, \{2, 4\}) = \binom{n}{2} + 1$ for all $n \ge 6$.

REFERENCES

- BOYVALENKOV P., K. DELCHEV, D. V. ZINOVIEV, V. A. ZINOVIEV (2021) On twoweight codes, Discrete Math., 344(5), https://doi.org/10.1016/j.disc.2021. 112318.
- [²] ERDŐS P., C. KO, R. RADO (1961), Intersection theorems for systems of finite sets, Quart. J. Math. Oxford Ser. (2), **12**, 313–320.
- [³] FRANKL P. (1987) The shifting technique in extremal set theory. In: Surveys in Combinatorics, London Math. Soc. Lecture Note Ser., **123**, Cambridge, Cambridge University Press, 81–110.

Institute of Mathematics and Informatics	*Faculty of Mathematics and Informatics
Bulgarian Academy of Sciences	Sofia University "St. Kliment Ohridski"
Akad. G. Bonchev St, Bl. 8	5, J. Bourchier Blvd
1113 Sofia, Bulgaria	1164 Sofia, Bulgaria
e-mail: ivan@math.bas.bg	$e ext{-mail:} assia@fmi.uni-sofia.bg$

I. Landjev, A. Rousseva