On Optimal Finite-length Binary Codes of Four Codewords for Binary Symmetric Channels

Yanyan Dong and Shenghao Yang
The Chinese University of Hong Kong, Shenzhen

***Abstract*—Finite-length binary codes of four codewords are studied for binary symmetric channels (BSCs) with the maximum likelihood decoding. For any block-length, optimal *linear* codes of four codewords have been explicitly characterized, but whether linear codes are optimal or not is unknown in general. In this paper, we show that for any block-length, there exists an optimal code of four codewords that is either linear or in a subset of nonlinear codes, called *Class-I* codes. Based on the analysis of Class-I codes, we derive sufficient conditions such that linear codes are optimal. For block-length less than or equal to 8, our analytical results show that linear codes are optimal. For block-length up to 290, numerical evaluations show that linear codes are optimal.**

I. Introduction

A binary code of block length n and codebook size 2^k is called an (n,k) code, which is said to be *linear* if it is a subspace of $\{0,1\}^n$. Linear codes have been extensively studied in coding theory. For binary symmetric channels (BSCs), asymptotically capacity achieving linear codes with low encoding/decoding complexity have been designed, for example polar codes [1]. However, whether linear codes are optimal or not among all (n,k) codes for BSCs in terms of the maximum likelihood (ML) decoding is a long-standing open problem, dated back to Slepian's 1956 paper [2]. Except for codes that are perfect or quasi-perfect, very little is known about optimal codes for BSC. Note that it is also hard to search optimal codes by computers when n is slightly large [3].

In this paper, we study binary $(n,2)$ codes for fixed n. Optimal linear $(n,2)$ codes have been explicitly characterized for each block length n [4], [5], but whether linear $(n,2)$ codes are optimal or not among all $(n,2)$ codes in terms of the ML decoding is unknown in general [6]. In this paper, we derive a general approach for comparing the ML decoding performance of two $(n,2)$ codes with certain small difference. Based on this approach, we verify that linear $(n,2)$ codes are optimal for a range of n.

In particular, we show that for any block-length n, there exists an optimal $(n,2)$ code that is either linear or in a subset of nonlinear codes, called *Class-I* codes. Based on the analysis of Class-I codes, we derive sufficient conditions such that linear codes are optimal. For $n \leq 8$, our analytical results show that linear codes are optimal. For n up to 290, numerical evaluations show that linear codes are optimal. Moreover, most ML decoding comparison results obtained in this paper are *universal* in the sense that they do not depend on the crossover probability of the BSC.

In the remainder of this paper, we first formulate the problem and introduce our main results. In §III, we outline a general approach for comparing the ML decoding performance of two codes, for which two special cases are used in this paper: two codes with only one column difference (see §IV) and two codes with only one codeword different in two bits (see §VI). §V is dedicated to the analysis of Class-I codes, based on the results in §IV. Omitted proofs can be found in the full version [7].

II. Problem Formulation and Main Results

A. Formulation of (n,k) Codes

A (n,k) binary node $\mathcal{C}$ is a subset of $\{0,1\}^n$ of size 2^k, and is said to be *linear* if it is a subspace of $\{0,1\}^n$. Using the codewords of $\mathcal{C}$ as rows, we can form a $2^k \times n$ binary matrix C, which is used interchangeably with $\mathcal{C}$. For $i = 1, \ldots, 2^k$, let $\mathbf{c}_i$ be the ith row of C, i.e., a codeword of $\mathcal{C}$.

For $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$, let $w(\mathbf{x})$ be the *Hamming weight* of $\mathbf{x}$ and let $\mathbf{x} \oplus \mathbf{y}$ be the bit-wise exclusive OR of $\mathbf{x}$ and $\mathbf{y}$, so that $w(\mathbf{x} \oplus \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$. Let

$$d_C(\mathbf{y}) = \min_{j \in \{1,\ldots 2^k\}} w(\mathbf{c}_j \oplus \mathbf{y}).$$

Consider the communication over a memoryless binary symmetric channel (BSC) with crossover probability ϵ ($0 < \epsilon < \frac{1}{2}$). For a channel input $\mathbf{x} \in \{0,1\}^n$, the channel output is $\mathbf{y} \in \{0,1\}^n$ with probability

$$p(\mathbf{y}|\mathbf{x}) = (1-\epsilon)^{n-w(\mathbf{x}\oplus\mathbf{y})}\epsilon^{w(\mathbf{x}\oplus\mathbf{y})}.$$

Suppose an (n,k) code C is used for this BSC. The maximum-likelihood (ML) decoding rule decodes an output $\mathbf{y}$ to a code word $\mathbf{c}_i$ if $w(\mathbf{c}_i \oplus \mathbf{y}) = d_C(\mathbf{y})$, where a tie is resolved arbitrarily. Define

$$\alpha_d(C) = |\{\mathbf{y} \in \{0,1\}^n : d_C(\mathbf{y})) = d\}|,$$

which is the number of outputs $\mathbf{y}$ that is decoded to a codeword of distance d. Note that the value $\alpha_d(C)$ does not depend on ϵ. The (average) correct decoding probability of C is

$$\lambda_C = \frac{1}{|C|}\sum_{d=0}^{n} \alpha_d(C)(1-\epsilon)^{n-d}\epsilon^d. \tag{1}$$

We say an (n,k) code C is *better or no worse* than another (n,k) code C' if $\lambda_C \geq \lambda_{C'}$. We say an (n,k) code C is *optimal* if it is better than any other (n,k) codes. If valid for all ϵ, a property of a code is said to be *universal*.

B. Main Results about $(n, 2)$ Codes

In this paper, we focus on $(n, 2)$ codes, which has four codewords. The columns of an $(n, 2)$ code C are of vectors in $\{0, 1\}^4$. We use $\langle i \rangle$ to denote the binary vector associated with an integer $i \geq 0$, where the length of the vector is implied in the context. For example,

$$\langle 1 \rangle = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\top, \quad \langle 2 \rangle = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^\top.$$

We use $\{i\}_C$ to denote the index set of the columns of C equal to $\langle i \rangle$, and let $|i|_C$ be the size of $\{i\}_C$. We may write $|i|_C$ as $|i|$ when the code C is implied in the context. For example, the code

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

has the ith column of type $\langle i \rangle$ and $|i| = 1$ for $i = 1, \ldots, 7$.

The following facts about $(n, 2)$ codes are straightforward (see also [4], [5]). First, codes with all-zero columns are not optimal. Second, flipping all the bits in a column does not change the decoding performance. Third, row and column permutations of C do not affect the decoding performance. Due to these facts, we only need to consider C of 7 types of the columns: $\langle 1 \rangle, \langle 2 \rangle, \ldots, \langle 7 \rangle$ for finding an optimal code.

Theorem 1. *Consider an $(n, 2)$ code C with $w(\mathbf{c}_s \oplus \mathbf{c}_t)$ even for certain $1 \leq s \neq t \leq 4$, and with a column of type $\langle 2^{4-s} \rangle$. Let C' be the code obtained by replacing a column of type $\langle 2^{4-s} \rangle$ of C by $\langle 2^{4-s} + 2^{4-t} \rangle$. Then, $\lambda_{C'} \geq \lambda_C$.*

Proof. See Section IV-B. □

For example, suppose an $(n, 2)$ code C has a column $\langle 1 \rangle$ and $w(\mathbf{c}_3 \oplus \mathbf{c}_4)$ even. The above theorem says, if we replace a column of type $\langle 1 \rangle$ of C by $\langle 3 \rangle$, the ML decoding performance is better.

Theorem 2. *Consider an $(n, 2)$ code C with first two columns of the types $\langle 1 \rangle$ (resp. $\langle 2 \rangle$, $\langle 4 \rangle$) and $\langle 7 \rangle$. Let C' be the code obtained by replacing the first two columns of C with $\langle 3 \rangle$ and $\langle 5 \rangle$ (resp. $\langle 3 \rangle$ and $\langle 6 \rangle$, $\langle 5 \rangle$ and $\langle 6 \rangle$). Then $\lambda_{C'} \geq \lambda_C$.*

Proof. See Section VI. □

Using the above two theorems, we can reduce the searching range for an optimal $(n, 2)$ code. Note that a linear $(n, 2)$ code (subject to row interchanging) has $|3| + |5| + |6| = n$.

Definition 1. An $(n, 2)$ code C is of *Class-I* if $|1|$ is odd, $|3|, |5|, |6|$ are of the same parity, and $|1| + |3| + |5| + |6| = n$.

Theorem 3. *An optimal $(n, 2)$ code exists in the set formed by all the linear codes and Class-I codes.*

We have the following properties of Class-I codes.

Theorem 4. *Let C be a Class-I $(n, 2)$ code with $|1| = 1$. Let C' be the code obtained by replacing the $\langle 1 \rangle$ column of C by $\langle s \rangle$, where $s = \arg\min_{i=3,5,6} |i|$. Then $\lambda_{C'} \geq \lambda_C$.*

Proof. See Section V-B □

In the above theorem, code C' is linear.

Theorem 5. *Let C be a Class-I $(n, 2)$ code with $\min_{i=3,5,6} |i| = 0$ or 1. Let C' be the code obtained by replacing one $\langle 1 \rangle$ column of C by $\langle s \rangle$, where $s \in \{3, 5, 6\}$ with $|s| = 0$ or 1. Then $\lambda_{C'} \geq \lambda_C$.*

The above analysis enables us to obtain the following sufficient condition about the optimality of linear codes.

Theorem 6. *Fix a block length n. If for any Class-I $(n, 2)$ code C, there exists an $(n, 2)$ code C' such that $|1|_{C'} < |1|_C$, $|1|_{C'} + |3|_{C'} + |5|_{C'} + |6|_{C'} = n$ and $\lambda_C \leq \lambda_{C'}$, then linear $(n, 2)$ codes are optimal.*

Corollary 7. *For $n \leq 8$, linear $(n, 2)$ codes are optimal.*

For a general block length n, if we can verify the condition of Theorem 6, then there exists an optimal $(n, 2)$ code that is linear. For each Class-I $(n, 2)$ code C, we can compare it with the code C' obtained by replacing one $\langle 1 \rangle$ column of C by $\langle s \rangle$ with $s = \arg\min_{i=3,5,6} |i|$. (See the formula for comparing λ_C and $\lambda_{C'}$ in §V.) Using computer evaluation, we have verified that for $n \leq 290$, linear codes are optimal.

III. An Approach of Comparing two $(n, 2)$ Codes

We first define some notations. Let C be an $(n, 2)$ code with the jth codeword/row $\mathbf{c}_j$, $j = 1, \ldots, 4$. Denote

$$d_j(\mathbf{y}) = w(\mathbf{c}_j \oplus \mathbf{y}). \tag{2}$$

For $\mathbf{y} \in \{0, 1\}^k$ and $i = 1, \ldots, k$, denote $\mathbf{y}_i$ as the ith entry of $\mathbf{y}$. For example, $\langle 2 \rangle_3 = 1$. Denote $\mathbf{y}_\mathcal{A}$ the sub-vector of $\mathbf{y}$ formed by the entries indexed by $\mathcal{A}$. For $i = 0, 1, \ldots, 15$, define $\omega_i(\mathbf{y}) = w(\mathbf{y}_{\{i\}_C})$ for $\mathbf{y} \in \{0, 1\}^n$. When $\mathbf{y}$ is clear from the context, we write $\omega_i = \omega_i(\mathbf{y})$. For a vector $\mathbf{y} \in \{0, 1\}^n$, we can rewrite d_j defined in (2) as

$$d_j(\mathbf{y}) = \sum_{i=0}^{15} \left[|i|/2 + (-1)^{\langle i \rangle_j} (\omega_i - |i|/2) \right]. \tag{3}$$

We also write $d_j = d_j(\mathbf{y})$ when $\mathbf{y}$ is clear from the context. For example, for C of columns of types only $\langle 1 \rangle, \langle 2 \rangle, \ldots, \langle 7 \rangle$,

$$d_1(\mathbf{y}) = \omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 + \omega_6 + \omega_7, \tag{4}$$

$$d_2(\mathbf{y}) = \omega_1 + \omega_2 + \omega_3 + \overline{\omega_4} + \overline{\omega_5} + \overline{\omega_6} + \overline{\omega_7}, \tag{5}$$

$$d_3(\mathbf{y}) = \omega_1 + \overline{\omega_2} + \overline{\omega_3} + \omega_4 + \omega_5 + \overline{\omega_6} + \overline{\omega_7}, \tag{6}$$

$$d_4(\mathbf{y}) = \overline{\omega_1} + \omega_2 + \overline{\omega_3} + \omega_4 + \overline{\omega_5} + \omega_6 + \overline{\omega_7}, \tag{7}$$

where $\overline{\omega}_i = |i|_C - \omega_i$.

We compare C with another $(n, 2)$ code C' obtained by modifying C as follows. Let $\mathcal{O}$ be a nonempty, proper subset of $\{1, 2, 3, 4\}$ and let $\mathcal{P}$ be its complement, which is also nonempty. Let C' be the code obtained by flipping the first t bits of $\mathbf{c}_i$ for each $i \in \mathcal{P}$. Denote by $\mathbf{c}'_i$ the ith codeword/row of C', $i = 1, \ldots, 4$. For $\mathbf{y} \in \{0, 1\}^n$, let $f_t(\mathbf{y})$ be the vector obtained by flipping the first t bits of $\mathbf{y}$. We see that $\mathbf{c}'_i = \mathbf{c}_i$ for $i \in \mathcal{O}$ and $\mathbf{c}'_i = f_t(\mathbf{c}_i)$ for $i \in \mathcal{P}$.

Denote by s_τ, $\tau = 1, 2, \ldots, t$ the τth column of C. For $\mathbf{y} \in \{0,1\}^n$, let

$$d'_i(\mathbf{y}) = d_i(\mathbf{y}) + \sum_{\tau=1}^{t} (-1)^{(s_\tau)_i} (\overline{\mathbf{y}_\tau} - \mathbf{y}_\tau). \tag{8}$$

For a nonempty subset $\mathcal{S} \subset \{1, \ldots, 4\}$, let

$$d_{\mathcal{S}}(\mathbf{y}) = \min_{i \in \mathcal{S}} d_i(\mathbf{y}) \quad \text{and} \quad d'_{\mathcal{S}}(\mathbf{y}) = \min_{i \in \mathcal{S}} d'_i(\mathbf{y}).$$

We have

$$d_C(\mathbf{y}) = \min\{d_{\mathcal{O}}(\mathbf{y}), d_{\mathcal{P}}(\mathbf{y})\}, \tag{9}$$

$$d_{C'}(\mathbf{y}) = \min\{d_{\mathcal{O}}(\mathbf{y}), d'_{\mathcal{P}}(\mathbf{y})\}, \tag{10}$$

$$\begin{aligned} d_{C'}(f_t(\mathbf{y})) &= \min\{d_{\mathcal{O}}(f_t(\mathbf{y})), d'_{\mathcal{P}}(f_t(\mathbf{y}))\} \\ &= \min\{d'_{\mathcal{O}}(\mathbf{y}), d_{\mathcal{P}}(\mathbf{y})\}. \end{aligned} \tag{11}$$

Our approach to compare the ML decoding performance of C and C' is based on a pair of partitions $\{\mathcal{Y}_i, i = 1, \ldots, i_0\}$ and $\{\mathcal{Y}'_i, i = 1, \ldots, i_0\}$ of $\{0,1\}^n$, where i_0 indicates the number of subsets in each partition. This pair of partitions satisfy the following properties: 1) for each i, $|\mathcal{Y}_i| = |\mathcal{Y}'_i|$, and 2) for each i, there exists an one-to-one and onto mapping $f_i : \mathcal{Y}_i \to \mathcal{Y}'_i$ such that one of the following conditions hold:

1) for all $\mathbf{y} \in \mathcal{Y}_i$, $d_C(\mathbf{y}) = d_{C'}(f_i(\mathbf{y}))$;
2) for all $\mathbf{y} \in \mathcal{Y}_i$, $d_C(\mathbf{y}) < d_{C'}(f_i(\mathbf{y}))$;
3) for all $\mathbf{y} \in \mathcal{Y}_i$, $d_C(\mathbf{y}) > d_{C'}(f_i(\mathbf{y}))$.

Such a pair of partitions exists. For example, when $i_0 = 2^n$, $\mathcal{Y}_i = \mathcal{Y}'_i = \{\langle i \rangle\}$ for $i = 0, 1, \ldots, i_0 - 1$ form a pair of partitions satisfying the desired properties. But this example does not help to simplify the problem. For the two special cases we will discuss, there exists such a pair of partitions with $i_0 = 5$.

In the following discussion, for a binary variable $x \in \{0,1\}$, we write $\bar{x} = 1 - x$. We write $\min\{a, b\}$ as $a \wedge b$. For a function $g : \{0,1\}^n \to \mathbb{R}$, we write $\{\mathbf{y} \in \{0,1\}^n : g(\mathbf{y}) \geq 0\}$ as $\{g \geq 0\}$ to simplify the notations.

IV. Change of one Column

In this section, we study how the ML decoding performance is affected after changing one column of an $(n, 2)$ code.

A. General Results

Consider an $(n, 2)$ code C with the first column $\langle s \rangle$. Let C' be the code formed by changing the first column of C to $\langle s' \rangle$, $s' \neq s$. Following the notations in §III, $\mathcal{O}$ is the set of index j such that $\langle s \rangle_j = \langle s' \rangle_j$, and $\mathcal{P}$ is the set of index j such that $\langle s \rangle_j \neq \langle s' \rangle_j$. As $s \neq s'$, both $\mathcal{O}$ and $\mathcal{P}$ are nonempty. In this case, d'_i defined in (8) becomes

$$d'_i(\mathbf{y}) = d_i(\mathbf{y}) + (-1)^{\langle s \rangle_i} (\overline{\mathbf{y}_1} - \mathbf{y}_1). \tag{12}$$

Consider an example with $s = 1$ and $s' = 3$. Now $\mathcal{O} = \{1, 2, 4\}$ and $\mathcal{P} = \{3\}$. Substituting $\langle 1 \rangle = [0\ 0\ 0\ 1]^\top$ into (12),

$$\begin{aligned} d'_1(\mathbf{y}) &= d_1(\mathbf{y}) - \mathbf{y}_1 + \overline{\mathbf{y}_1}, \\ d'_2(\mathbf{y}) &= d_2(\mathbf{y}) - \mathbf{y}_1 + \overline{\mathbf{y}_1}, \\ d'_3(\mathbf{y}) &= d_3(\mathbf{y}) - \mathbf{y}_1 + \overline{\mathbf{y}_1}, \\ d'_4(\mathbf{y}) &= d_4(\mathbf{y}) + \mathbf{y}_1 - \overline{\mathbf{y}_1}. \end{aligned}$$

and hence

$$d_{\mathcal{O}}(\mathbf{y}) = d_1 \wedge d_2 \wedge d_4 \tag{13}$$

$$d_{\mathcal{P}}(\mathbf{y}) = d_3 \tag{14}$$

$$d'_{\mathcal{O}}(\mathbf{y}) = [(d_1 \wedge d_2) - \mathbf{y}_1 + \overline{\mathbf{y}_1}] \wedge (d_4 + \mathbf{y}_1 - \overline{\mathbf{y}_1}) \tag{15}$$

$$d'_{\mathcal{P}}(\mathbf{y}) = d_3 - \mathbf{y}_1 + \overline{\mathbf{y}_1}. \tag{16}$$

We are ready to form the partitions. Define the following 5 subsets of $\{0,1\}^n$:

$$\begin{aligned} \mathcal{Y}_1 &= \{d_{\mathcal{O}} \leq d_{\mathcal{P}} < d'_{\mathcal{P}}\} \cup \{d_{\mathcal{O}} \leq d'_{\mathcal{P}} \leq d_{\mathcal{P}}, d'_{\mathcal{O}} \leq d'_{\mathcal{P}}\}, \\ \mathcal{Y}_2 &= \{d_{\mathcal{P}} \leq d'_{\mathcal{P}}, d_{\mathcal{P}} < d_{\mathcal{O}}\} \cup \{d'_{\mathcal{P}} < d_{\mathcal{P}} \leq d_{\mathcal{O}}, d_{\mathcal{P}} \leq d'_{\mathcal{O}}\}, \\ \mathcal{Y}_3 &= \{d'_{\mathcal{P}} = d'_{\mathcal{O}} < d_{\mathcal{P}} = d_{\mathcal{O}}\}, \\ \mathcal{Y}_4 &= \{d_{\mathcal{P}} = d'_{\mathcal{P}} = d_{\mathcal{O}} < d'_{\mathcal{O}}\}, \\ \mathcal{Y}_5 &= \{d'_{\mathcal{P}} = d_{\mathcal{O}} < d'_{\mathcal{O}} = d_{\mathcal{P}}\}. \end{aligned} \tag{17, 18}$$

For $i = 2, 4, 5$, define

$$\mathcal{Y}'_i = \{f_1(\mathbf{y}) : \mathbf{y} \in \mathcal{Y}_i\},$$

where function f_1 (defined in §III) flips the first bit of a binary vector. The next lemma shows that both $\{\mathcal{Y}_i, i = 1, \ldots, 5\}$ and $\{\mathcal{Y}_1, \mathcal{Y}'_2, \mathcal{Y}_3, \mathcal{Y}'_4, \mathcal{Y}'_5\}$ are partitions of $\{0,1\}^n$ and satisfy the desired properties we described in §III.

Lemma 8. *For the $(n, 2)$ codes C and C' formulated above, both $\{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4, \mathcal{Y}_5\}$ and $\{\mathcal{Y}_1, \mathcal{Y}'_2, \mathcal{Y}_3, \mathcal{Y}'_4, \mathcal{Y}'_5\}$ are partitions of $\{0,1\}^n$. Moreover,*

1) *For $\mathbf{y} \in \mathcal{Y}_1$, $d_C(\mathbf{y}) = d_{C'}(\mathbf{y}) = d_{\mathcal{O}}$;*
2) *For $\mathbf{y} \in \mathcal{Y}_2$, $d_C(\mathbf{y}) = d_{C'}(\mathbf{y}') = d_{\mathcal{P}}$ where $\mathbf{y}' \triangleq f_1(\mathbf{y}) \in \mathcal{Y}'_2$;*
3) *For $\mathbf{y} \in \mathcal{Y}_3$, $d_C(\mathbf{y}) = d_{\mathcal{P}} = d_{C'}(\mathbf{y}) + 1 = d'_{\mathcal{P}} + 1$;*
4) *For $\mathbf{y} \in \mathcal{Y}_4$, $d_C(\mathbf{y}) = d_{\mathcal{O}} = d_{C'}(\mathbf{y}') = d_{\mathcal{P}}$ where $\mathbf{y}' \triangleq f_1(\mathbf{y}) \in \mathcal{Y}'_4$;*
5) *For $\mathbf{y} \in \mathcal{Y}_5$, $d_C(\mathbf{y}) + 1 = d_{\mathcal{O}} + 1 = d_{C'}(\mathbf{y}') = d_{\mathcal{P}}$ where $\mathbf{y}' \triangleq f_1(\mathbf{y}) \in \mathcal{Y}'_5$.*

Now we move on to compare $\lambda_{\mathcal{C}}$ and $\lambda_{\mathcal{C}'}$ as defined in (1). Define for $i = 1, \ldots, 5$ and $d = 0, 1, \ldots, n$,

$$\alpha_d^i(C) = |\{\mathbf{y} \in \mathcal{Y}_i : d_C(\mathbf{y}) = d\}|. \tag{19}$$

As $\{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4, \mathcal{Y}_5\}$ is a partition of $\{0,1\}^n$, we have

$$\alpha_d(C) = \sum_{i=1}^{5} \alpha_d^i(C).$$

By the definition of $\mathcal{Y}_3$ and $\mathcal{Y}_5$, $\alpha_0^3(C) = 0$ and $\alpha_n^5(C) = 0$.

Theorem 9. *For two $(n, 2)$ codes C and C' with only one column different, $\lambda_{\mathcal{C}'} \geq \lambda_{\mathcal{C}}$ if and only if*

$$\sum_{d=1}^{n} [\alpha_d^3(C) - \alpha_{d-1}^5(C)] \left(\frac{\epsilon}{1-\epsilon}\right)^{d-1} \geq 0.$$

Corollary 10. *For two $(n, 2)$ codes C and C' with only one column different, $\lambda_{\mathcal{C}'} \geq \lambda_{\mathcal{C}}$ if for $d = 1, \ldots, n$,*

$$\sum_{i=1}^{d} \alpha_i^3(C) \geq \sum_{i=0}^{d-1} \alpha_i^5(C).$$

If we can compare C and C' based on Corollary 10, their relation is *universal* in the sense that it does not depend on ϵ.

B. Proof of Theorem 1

Now we give a proof of Theorem 1.

As interchanging rows/columns does not change the performance of C, we only consider the following case when proving the theorem: C has the first column $\langle 1\rangle$ and $w(\mathbf{c}_3\oplus\mathbf{c}_4)$ is even. Let C' be the code obtained by replacing the first column of C by $\langle 3\rangle$. Substituting $s=1$ and $s'=3$ to the discussion in §IV-A, we have $\mathcal{O}=\{1,2,4\}$ and $\mathcal{P}=\{3\}$, and hence

$$\mathcal{Y}_5=\{d'_3=d_{\{1,2,4\}}<d_3=d'_{\{1,2,4\}}\}.$$

Assume $\mathcal{Y}_5$ is nonempty and fix $\mathbf{y}\in\mathcal{Y}_5$. As $d'_3(\mathbf{y})=d_3(\mathbf{y})-\mathbf{y}_1+\overline{\mathbf{y}_1}$, we have $\mathbf{y}_1=1$. Further, due to

$$\begin{aligned}d_{\{1,2,4\}}(\mathbf{y})&=d_1\wedge d_2\wedge d_4,\\ d'_{\{1,2,4\}}(\mathbf{y})&=(d_1-1)\wedge(d_2-1)\wedge(d_4+1),\end{aligned}$$

we have $d_{\{1,2,4\}}=d_4$ and hence $d_3=d_4+1$. By (3),

$$\begin{aligned}&d_3(\mathbf{y})+d_4(\mathbf{y})\\ &=\sum_i\Big[|i|/2+(-1)^{\langle i\rangle_3}(\omega_i-|i|/2)\Big]+\\ &\quad\sum_i\Big[|i|/2+(-1)^{\langle i\rangle_4}(\omega_i-|i|/2)\Big]\\ &=\sum_{i:\langle i\rangle_3\neq\langle i\rangle_4}|i|+2\sum_{i:\langle i\rangle_3=\langle i\rangle_4}\Big[|i|/2+(-1)^{\langle i\rangle_3}(\omega_i-|i|/2)\Big]\\ &=w(\mathbf{c}_3\oplus\mathbf{c}_4)+2\sum_{i:\langle i\rangle_3=\langle i\rangle_4}\Big[|i|/2+(-1)^{\langle i\rangle_3}(\omega_i-|i|/2)\Big].\end{aligned}$$

As $w(\mathbf{c}_3\oplus\mathbf{c}_4)$ is even, we see that d_3+d_4 is even, which is a contradiction to $d_3=d_4+1$. Therefore, $\mathcal{Y}_5=\emptyset$ and hence by Corollary 10, $\lambda_{C'}\geq\lambda_C$.

V. Analysis of Class-I Codes

In this section, we consider a Class-I $(n,2)$ code C with the first column $\langle 1\rangle$. Let C' be the code obtained by replacing the first column of C to $\langle 3\rangle$. The ML decoding performance of C and C' can be compared using the approach introduced in §IV-A.

A. Characterizations of $\mathcal{Y}_3$ and $\mathcal{Y}_5$

Guided by Theorem 9 and Corollary 10, we first study $\mathcal{Y}_3$ and $\mathcal{Y}_5$ defined in (17) and (18). The following lemma can be proved using (13) – (16).

Lemma 11. *For a Class-I $(n,2)$code C with the first column $\langle 1\rangle$ and C' obtained by replacing the first column of C to $\langle 3\rangle$,*

$$\begin{aligned}\mathcal{Y}_3&=\{\mathbf{y}_1=1,d_4\geq d_1\wedge d_2=d_3\},\\ \mathcal{Y}_5&=\{\mathbf{y}_1=1,d_1\wedge d_2\geq d_4+2=d_3+1\}.\end{aligned}$$

1) Characterization of α_i^3: For $\mathbf{y}\in\mathcal{Y}_3$, by Lemma 8, $d_C(\mathbf{y})=d_3$. By (4) – (7) and Lemma 11, we have the following necessary and sufficient condition for $\mathbf{y}\in\mathcal{Y}_3$ with $d_C(\mathbf{y})=i$: $\mathbf{y}_1=1$ and

$$\begin{aligned}&w_1+\overline{w_3}=i-w_5-\overline{w_6},\\ &w_1-\overline{w_1}\leq\overline{w_5}+w_6-w_5-\overline{w_6},\\ &w_3-\overline{w_3}=w_5+\overline{w_6}-(w_5+w_6)\wedge(\overline{w_5}+\overline{w_6}).\end{aligned}$$

We discuss two cases according to $w_5+w_6<\overline{w_5}+\overline{w_6}$ or not.

Define $\mathcal{Y}_3^A(i)$ as the collection of $\mathbf{y}$ satisfying $\mathbf{y}_1=1$ and

$$\begin{aligned}w_5+w_6&<(|5|+|6|)/2,&(20)\\ w_1+w_5&=i-(|3|+|6|)/2,&(21)\\ w_1+w_5-w_6&\leq(|1|+|5|-|6|)/2,&(22)\\ w_3+w_6&=(|3|+|6|)/2.&(23)\end{aligned}$$

We have

$$|\mathcal{Y}_3^A(i)|=\sum_{\substack{w_1\geq1,w_3,w_5,w_6:\\ (20),(21),(22),(23)}}\binom{|1|-1}{w_1-1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}.$$

Define $\mathcal{Y}_3^B(i)$ as the collection of $\mathbf{y}$ satisfying $\mathbf{y}_1=1$ and

$$\begin{aligned}w_5+w_6&\geq(|5|+|6|)/2,&(24)\\ w_1+\overline{w_6}&=i-(|3|+|5|)/2,&(25)\\ w_1+w_5-w_6&\leq(|1|+|5|-|6|)/2,&(26)\\ w_3-w_5&=(|3|-|5|)/2.&(27)\end{aligned}$$

We have

$$|\mathcal{Y}_3^B(i)|=\sum_{\substack{w_1\geq1,w_3,w_5,w_6:\\ (24),(25),(26),(27)}}\binom{|1|-1}{w_1-1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}.\quad(28)$$

We see that $\alpha_i^3=|\mathcal{Y}_3^A(i)|+|\mathcal{Y}_3^B(i)|$.

2) Characterization of α_i^5: For $\mathbf{y}\in\mathcal{Y}_5$, by Lemma 8, $d_C(\mathbf{y})=d_3-1$. By (4) – (7) and Lemma 11, we have the following necessary and sufficient condition for $\mathbf{y}\in\mathcal{Y}_5$ with $d_C(\mathbf{y})=i$: $\mathbf{y}_1=1$ and

$$\begin{aligned}&w_1+\overline{w_3}=i+1-w_5-\overline{w_6},\\ &w_1-\overline{w_1}=\overline{w_5}+w_6-w_5-\overline{w_6}+1,\\ &w_3-\overline{w_3}\geq w_5+\overline{w_6}-(w_5+w_6)\wedge(\overline{w_5}+\overline{w_6})+1,\end{aligned}$$

which can be further simplified as $\mathbf{y}_1=1$ and

$$\begin{aligned}w_3&=(n+|3|-1)/2-i,&(29)\\ w_1+w_5-w_6&=(|1|+|5|-|6|+1)/2,&(30)\\ w_3+w_6&\geq(|3|+|6|)/2+1,&(31)\\ w_3-w_5&\geq(|3|-|5|)/2+1.&(32)\end{aligned}$$

Hence

$$\alpha_i^5=\sum_{\substack{w_1\geq1,w_3,w_5,w_6:\\ (29),(30),(31),(32)}}\binom{|1|-1}{w_1-1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}.$$

B. Class-I Codes with $|1|=1$

Following the discuss in the last subsection, we consider the special case with $|1|=1$, and prove Theorem 4 for the case $|3|=\min\{|3|,|5|,|6|\}$. For other cases, we can perform row interchanging and column bit flipping to convert the problem to this case.

When $|1|=1$, $w_1=1$. Using the characterization in the last subsection, we have

$$\sum_{i=0}^{d-1}\alpha_i^5=\sum_{\mathcal{W}_5}\binom{|3|}{\omega_3}\binom{|5|}{\frac{|5|-|6|}{2}+\omega_6}\binom{|6|}{\omega_6}\quad(33)$$

where

$$\mathcal{W}_5 = \left\{ \begin{matrix} (31), \\ (30)+(32), \\ \omega_3 \geq \frac{n+|3|+1}{2} - d \end{matrix} \right\} = \left\{ \begin{matrix} \omega_3+\omega_6 \geq \frac{|3|+|6|}{2}+1, \\ \omega_3-\omega_6 \geq \frac{|3|-|6|}{2}+1, \\ \omega_3 \geq \frac{n+|3|+1}{2} - d \end{matrix} \right\}.$$

Similarly,

$$\begin{aligned} \sum_{i=1}^{d} \alpha_i^3 &\geq \sum_{i=1}^{d} |\mathcal{Y}_3^B(i)| \\ &= \sum_{\mathcal{W}_3} \binom{|3|}{\omega_3} \binom{|5|}{\frac{|5|-|3|}{2}+\omega_3} \binom{|6|}{\omega_6} \\ &= \sum_{\mathcal{W}_3'} \binom{|3|}{\frac{|3|-|6|}{2}+\omega_6'} \binom{|5|}{\frac{|5|-|6|}{2}+\omega_6'} \binom{|6|}{\frac{|6|-|3|}{2}+\omega_3'} \quad (34) \end{aligned}$$

where

$$\mathcal{W}_3 = \left\{ \begin{matrix} (24)+(27), \\ (26)+(27), \\ \omega_6 \geq \frac{n+|6|+1}{2} - d \end{matrix} \right\} = \left\{ \begin{matrix} \omega_3+\omega_6 \geq \frac{|3|+|6|}{2} \\ \omega_3-\omega_6 \leq \frac{|3|-|6|-1}{2} \\ \omega_6 \geq \frac{n+|6|+1}{2} - d \end{matrix} \right\},$$

$$\mathcal{W}_3' = \left\{ \begin{matrix} \omega_3'+\omega_6' \geq \frac{|3|+|6|}{2} \\ \omega_3'-\omega_6' \geq \frac{|3|-|6|}{2} \\ \omega_3' \geq \frac{n+|3|+1}{2} - d \end{matrix} \right\},$$

and (34) is obtained by change of variables $\omega_3' - \frac{|3|}{2} = \omega_6 - \frac{|6|}{2}$ and $\omega_6' - \frac{|6|}{2} = \omega_3 - \frac{|3|}{2}$.

Note that $\mathcal{W}_5 \subset \mathcal{W}_3'$ and for $(\omega_3, \omega_6) \in \mathcal{W}_5$, due to $|3| \leq |6|$, we have

$$\binom{|3|}{\omega_3} \binom{|6|}{\omega_6} \leq \binom{|3|}{\frac{|3|-|6|}{2}+\omega_6} \binom{|6|}{\frac{|6|-|3|}{2}+\omega_3}.$$

Comparing (33) and (34), we obtain $\sum_{i=1}^{d} \alpha_i^3 \geq \sum_{i=0}^{d-1} \alpha_i^5$ for any $d = 1, \ldots, n$. By Corollary 10, $\lambda_{\mathcal{C}'} \geq \lambda_{\mathcal{C}}$, proving Theorem 4.

VI. Proof of Theorem 2

In this section, we outline the proof of Theorem 2, which use another case of the general approach in §III. Let C be an $(n, 2)$ code with the first two columns $\langle 1 \rangle$ and $\langle 7 \rangle$. Let C' be the code obtained by flipping the first two bits of $\mathbf{c}_3$, so that the first two column of C' are $\langle 3 \rangle$ and $\langle 5 \rangle$. (Other cases of Theorem 2 can be converted to this case by interchanging rows.)

Following the notations in §III, $\mathcal{O} = \{1, 2, 4\}$, $\mathcal{P} = \{3\}$, and

$$d_i'(\mathbf{y}) = d_i(\mathbf{y}) + (-1)^{\langle 1 \rangle_i} (\overline{\mathbf{y}_1} - \mathbf{y}_1) + (-1)^{\langle 7 \rangle_i} (\overline{\mathbf{y}_2} - \mathbf{y}_2).$$

When $\mathbf{y}_1 = \mathbf{y}_2$, we have

$$d_{\mathcal{P}}'(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}). \quad (35)$$

When $\mathbf{y}_1 \neq \mathbf{y}_2$, we have $d_1'(\mathbf{y}) = d_1(\mathbf{y})$, $d_4'(\mathbf{y}) = d_4(\mathbf{y})$, and

$$d_2'(\mathbf{y}) - d_2(\mathbf{y}) = d_3'(\mathbf{y}) - d_3(\mathbf{y}) = \pm 2, \quad (36)$$

and hence

$$\begin{aligned} &(d_{\mathcal{O}}'(\mathbf{y}) - d_{\mathcal{O}}(\mathbf{y}))(d_{\mathcal{P}}'(\mathbf{y}) - d_{\mathcal{P}}(\mathbf{y})) \\ &= (d_{\{1,2,4\}}'(\mathbf{y}) - d_{\{1,2,4\}}(\mathbf{y}))(d_3'(\mathbf{y}) - d_3(\mathbf{y})) \geq 0. \quad (37) \end{aligned}$$

Define the following subsets of $\{0, 1\}^n$:

$$\begin{aligned} \mathcal{Y}_1 &= \{\mathbf{y}_1 = \mathbf{y}_2\}, \\ \mathcal{Y}_2 &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} \leq d_{\mathcal{P}} \wedge d_{\mathcal{P}}'\}, \\ \mathcal{Y}_3 &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d_{\mathcal{P}}', d_{\mathcal{P}} \leq d_{\mathcal{O}} \wedge d_{\mathcal{O}}'\}, \\ \mathcal{Y}_4 &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{P}}' < d_{\mathcal{O}} \wedge d_{\mathcal{O}}' < d_{\mathcal{P}}\}, \\ \mathcal{Y}_5 &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}}' \leq d_{\mathcal{P}} \wedge d_{\mathcal{P}}' < d_{\mathcal{O}}, d_{\mathcal{O}}' < d_{\mathcal{P}}\}. \end{aligned}$$

Recall the function f_2 defined in §III that flips the first two bits of a binary vector. For $i = 3, 4$, let

$$\mathcal{Y}_i' = \{f_2(\mathbf{y}) : \mathbf{y} \in \mathcal{Y}_i\}.$$

We justify that $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4, \mathcal{Y}_5$ form a partition of $\{0, 1\}^n$ and $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3', \mathcal{Y}_4', \mathcal{Y}_5$ form a partition of $\{0, 1\}^n$. We first show that

$$\mathcal{Y}_4 \cup \mathcal{Y}_5 = \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d_{\mathcal{P}}', d_{\mathcal{P}} > d_{\mathcal{O}} \wedge d_{\mathcal{O}}'\}, \quad (38)$$

then we obtain $\bigcup_{i=1}^{5} \mathcal{Y}_i = \{0, 1\}^n$. Moreover, $\mathcal{Y}_1, \ldots, \mathcal{Y}_5$ are all disjoint by checking the definition. Thus $\mathcal{Y}_1, \ldots, \mathcal{Y}_5$ form a partition of $\{0, 1\}^n$.

We further show that

$$\mathcal{Y}_3' \cup \mathcal{Y}_4' \subseteq \mathcal{Y}_3 \cup \mathcal{Y}_4. \quad (39)$$

Since f_2 is an one-to-one mapping, we get $\mathcal{Y}_3' \cup \mathcal{Y}_4' = \mathcal{Y}_3 \cup \mathcal{Y}_4$. Therefore, $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3', \mathcal{Y}_4', \mathcal{Y}_5$ form a partition of $\{0, 1\}^n$.

Moreover, we prove the following claims:

1) For $\mathbf{y} \in \mathcal{Y}_1$, $d_C(\mathbf{y}) = d_{C'}(\mathbf{y})$;
2) For $\mathbf{y} \in \mathcal{Y}_2$, $d_C(\mathbf{y}) = d_{C'}(\mathbf{y}) = d_{\mathcal{O}}$;
3) For $\mathbf{y} \in \mathcal{Y}_3$, $d_C(\mathbf{y}) = d_{C'}(f_2(\mathbf{y})) = d_{\mathcal{P}}$;
4) For $\mathbf{y} \in \mathcal{Y}_4$, $d_C(\mathbf{y}) = d_{\mathcal{O}} \wedge d_{\mathcal{P}} \geq d_{C'}(f_2(\mathbf{y})) = d_{\mathcal{O}}'$;
5) For $\mathbf{y} \in \mathcal{Y}_5$, $d_C(\mathbf{y}) = d_{\mathcal{O}} \wedge d_{\mathcal{P}} \geq d_{C'}(\mathbf{y}) = d_{\mathcal{P}}'$.

Following the similar argument as in §IV-A, we can show that $\lambda_{C'} \geq \lambda_C$.

VII. Concluding Remarks

It is attractive to prove in general whether linear $(n, 2)$ codes are optimal or not. One further research direction is to extend the technique for comparing the decoding performance of two codes to codes of more than four codewords.

References

[1] 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, 2009.

[2] D. Slepian, "A class of binary signaling alphabets," *Bell System Technical Journal*, vol. 35, no. 1, pp. 203–234, 1956.

[3] W. W. Peterson and E. J. Weldon Jr., *Error-Correcting Codes*. MIT Press, 1972.

[4] J. Cordaro and T. Wagner, "Optimum (n, 2) codes for small values of channel error probability (corresp.)," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 349–350, 1967.

[5] P.-N. Chen, H.-Y. Lin, and S. M. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7346–7378, 2013.

[6] H.-Y. Lin, S. M. Moser, and P.-N. Chen, "Correction to weak flip codes and their optimality on the binary erasure channel." [Online]. Available: http://shannon.cm.nctu.edu.tw/html/paper/LMC2018C.pdf

[7] Y. Dong and S. Yang, "On optimal finite-length binary codes of four codewords for binary symmetric channels (full version)." [Online]. Available: arXiv