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 memoryless binary symmetric channels (BSCs) with the maximum likelihood decoding. For any block-length, best *linear* codes of four codewords have been explicitly characterized, but whether linear codes are better than nonlinear codes 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 296, 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 memoryless 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. The best 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 296, 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.

II. Problem Formulation and Main Results

A. Formulation of (n, k) Codes

An (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 $\langle i \rangle^k$ to denote the binary vector of length k associated with an integer $i \geq 0$. When the length of the vector is implied in the context the superscript is omitted. For example,

$$\langle 1 \rangle^4 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\top, \quad \langle 2 \rangle^4 = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^\top.$$

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 $(7, 2)$ 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.

Corollary 2. *Consider an $(n, 2)$ code C with $\sum_{i=1}^6 |i|_C = n$. There exists a code C' with $\lambda_{C'} \geq \lambda_C$ and $|1|_{C'} + |3|_{C'} + |5|_{C'} + |6|_{C'} = n$.*

Proof. In this proof, we write $|i|_C$ as $|i|$. Suppose at least two of $|1|, |2|, |4|$ are positive, since otherwise, the proof is done. We argue the case that $|1|$ and $|2|$ are positive. Other cases can be converted to this case by interchanging rows. Write

$$w(\mathbf{c}_2 \oplus \mathbf{c}_3) = |2| + |3| + |4| + |5|$$
$$w(\mathbf{c}_3 \oplus \mathbf{c}_4) = |1| + |2| + |5| + |6|$$
$$w(\mathbf{c}_2 \oplus \mathbf{c}_4) = |1| + |3| + |4| + |6|.$$

We claim that one of the above three weights must be even. For example, assume $w(\mathbf{c}_2 \oplus \mathbf{c}_3)$ is odd. Then $|3| + |4|$ and $|2| + |5|$ are of different parity, so that one of $w(\mathbf{c}_3 \oplus \mathbf{c}_4)$ and $w(\mathbf{c}_2 \oplus \mathbf{c}_4)$ must be even.

Suppose $w(\mathbf{c}_2 \oplus \mathbf{c}_3)$ is even. As $|2|$ is positive, Theorem 1 implies a better code with $|2|$ smaller. Repeating the above argument, there exists a better code C' where at most one of $|1|_{C'}, |2|_{C'}, |4|_{C'}$ is positive and $\sum_{i=1}^6 |i|_{C'} = n$. The corollary is proved by properly interchanging rows of C'. □

Corollary 3. *Consider a non-Class-I, nonlinear $(n, 2)$ code C with $|1|_C + |3|_C + |5|_C + |6|_C = n$. There exists an either linear or Class-I code C' with $\lambda_{C'} \geq \lambda_C$ and $|1|_{C'} < |1|_C$.*

Proof. In this proof, we write $|i|_C$ as $|i|$. Since C is nonlinear, $|1| > 0$. We claim that at least one of the following three weights are even:

$$w(\mathbf{c}_1 \oplus \mathbf{c}_4) = |1| + |3| + |5| \tag{2}$$
$$w(\mathbf{c}_3 \oplus \mathbf{c}_4) = |1| + |5| + |6| \tag{3}$$
$$w(\mathbf{c}_2 \oplus \mathbf{c}_4) = |1| + |3| + |6|, . \tag{4}$$

When $|1|$ is odd, $|3|$, $|5|$ and $|6|$ are not of the same parity since C is not of Class-I, which implies at least one of (2), (3), (4) is even. By Theorem 1, there is a better code C_1 with $|1|_{C_1} = |1| - 1$ even and $|1|_{C_1} + |3|_{C_1} + |5|_{C_1} + |6|_{C_1} = n$.

When $|1|$ is even, if (2), (3), (4) are all odd, then $|3| + |5|$, $|5| + |6|$ and $|3| + |6|$ are all odd, which is not possible for any integers $|3|$, $|5|$, $|6|$. Then at least one of (2), (3), (4) is even, Theorem 1 implies a better code C_1 with $|1|_{C_1} = |1| - 1$ odd and $|1|_{C_1} + |3|_{C_1} + |5|_{C_1} + |6|_{C_1} = n$.

For both case, a better code with $|1|$ strictly smaller always exists if C is non-Class-I, nonlinear. By repeating the similar argument on C_1, we eventually obtain a better code C' which either has $|1|_{C'} = 0$, i.e., linear or is of Class-I so that (2), (3), (4) are all odd. □

Theorem 4. *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 5. *An optimal $(n, 2)$ code exists in the set formed by all the linear codes and Class-I codes.*

Proof. As column flipping does not change the ML decoding performance, we consider a code C_1 with $\sum_{i=1}^7 |i| = n$ obtained by column flipping of C. We then discuss C_1 in two cases.

If $0 < |7| \leq |1| + |2| + |4|$ in C_1, by Theorem 4, there exists a code C_2 with $\lambda_{C_2} \geq \lambda_{C_1}$ and $\sum_{i=1}^6 |i| = n$ obtained by replacing, one-by-one, pairs of columns of types $\langle 7 \rangle$ and $\langle 2^s \rangle$ $(s = 0, 1, 2)$. Following Corollary 2, there exists code C_3, no worse than C_2, where $|1| + |3| + |5| + |6| = n$. Then by Corollary 3, there exists an either linear or Class-I code C_4 such that $\lambda_{C_4} \geq \lambda_{C_3}$.

If $|1|+|2|+|4|<|7|$ in C_1, by Theorem 4, there exists a better code C_2' with $|1|+|2|+|4|=0$. By flipping columns, we can obtain a code C_3' of the same performance of C_2' that has $|1|>0$ and $|1|+|3|+|5|+|6|=n$. Again, by Corollary 3, the proof is completed. □

We have the following properties of Class-I codes.

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

Proof. See Section V-B □

In the above theorem, code C' is linear.

Theorem 7. *Let C be a Class-I $(n,2)$ code with $\min_{i=3,5,6}|i|_C=0$ or 1. Let C' be the code obtained by replacing one $\langle 1\rangle$ column of C by $\langle s\rangle$, where $s\in\{3,5,6\}$ has $|s|_C=0$ or 1. Then $\lambda_{C'}\geq\lambda_C$.*

Proof. See Section V-C □

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

Theorem 8. *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.*

Proof. Assume that all optimal $(n,2)$ codes are nonlinear. By Theorem 5, there must exists an optimal $(n,2)$ code C that is Class-I. By the condition of the theorem, there exists an optimal code C' such that $|1|_{C'}<|1|_C$ and $|1|_{C'}+|3|_{C'}+|5|_{C'}+|6|_{C'}=n$. If $|1|_{C'}=0$, then C' is linear and we get a contradiction to the assumption that all optimal $(n,2)$ codes are nonlinear. If C' is Class-I, we repeat the above argument. If C' is non-Class-I and nonlinear, then by Corollary 3, there exists an optimal code C'' with $|1|_{C''}<|1|_{C'}$ that is either linear or Class-I. If C'' is linear, we get a contradiction to the assumption. If C'' is Class-I, we can repeat the above argument. As $|1|_C$ is finite, the process will eventually stop with an optimal linear code, i.e., a contradiction to the assumption that all optimal $(n,2)$ codes are nonlinear. □

Corollary 9. *For $n\leq 8$, linear $(n,2)$ codes are optimal.*

Proof. Fix $n\leq 8$. For a Class-I $(n,2)$ code with $|1|=1$, by Theorem 6, there exists a better linear code. For a Class-I $(n,2)$ code C with $|1|\geq 3$, we have $|3|+|5|+|6|\leq 5$ which implies $\min\{|3|,|5|,|6|\}\leq 1$. By Theorem 7, we have $\lambda_C\leq\lambda_{C'}$ for the $(n,2)$ code C' obtained by replacing one $\langle 1\rangle$ column of C by $\langle s\rangle$, where $s\in\{3,5,6\}$ with $|s|_C=0$ or 1. By Theorem 8, linear $(n,2)$ codes are optimal for $n\leq 8$. □

For a general block length n, if we can verify the condition of Theorem 8, 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 $n\leq 296$, 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

$$d_j(\mathbf{y})=w(\mathbf{c}_j\oplus\mathbf{y}). \tag{5}$$

For a binary vector $\mathbf{y}$ of k bits, denote $\mathbf{y}_i$ as the ith entry of $\mathbf{y}$. For example, $(\langle 2\rangle^4)_3=([0\ 0\ 1\ 0]^\top)_3=1$. 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 $w_i(\mathbf{y})=w(\mathbf{y}_{\{i\}_C})$ for $\mathbf{y}\in\{0,1\}^n$. When $\mathbf{y}$ is clear from the context, we write $w_i=w_i(\mathbf{y})$. For a vector $\mathbf{y}\in\{0,1\}^n$, we can rewrite d_j defined in (5) as

$$d_j(\mathbf{y})=\sum_{i=0}^{15}\left[|i|/2+(-1)^{\langle i\rangle^4_j}(w_i-|i|/2)\right]. \tag{6}$$

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 $\langle 1\rangle,\langle 2\rangle,\ldots,\langle 7\rangle$,

$$d_1(\mathbf{y})=w_1+w_2+w_3+w_4+w_5+w_6+w_7, \tag{7}$$
$$d_2(\mathbf{y})=w_1+w_2+w_3+\overline{w_4}+\overline{w_5}+\overline{w_6}+\overline{w_7}, \tag{8}$$
$$d_3(\mathbf{y})=w_1+\overline{w_2}+\overline{w_3}+w_4+w_5+\overline{w_6}+\overline{w_7}, \tag{9}$$
$$d_4(\mathbf{y})=\overline{w_1}+w_2+\overline{w_3}+w_4+\overline{w_5}+w_6+\overline{w_7}, \tag{10}$$

where $\overline{w_i}=|i|_C-w_i$.

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

$$d_i'(\mathbf{y})=d_i(\mathbf{y})+\sum_{\tau=1}^{t}(-1)^{(s_\tau)_i}(\overline{\mathbf{y}_\tau}-\mathbf{y}_\tau). \tag{11}$$

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

$$d_C(\mathbf{y})=\min\{d_{\mathcal{O}}(\mathbf{y}),d_{\mathcal{P}}(\mathbf{y})\}, \tag{12}$$
$$d_{C'}(\mathbf{y})=\min\{d_{\mathcal{O}}(\mathbf{y}),d'_{\mathcal{P}}(\mathbf{y})\}, \tag{13}$$
$$\begin{aligned}d_{C'}(f_t(\mathbf{y}))&=\min\{d_{\mathcal{O}}(f_t(\mathbf{y})),d'_{\mathcal{P}}(f_t(\mathbf{y}))\}\\&=\min\{d'_{\mathcal{O}}(\mathbf{y}),d_{\mathcal{P}}(\mathbf{y})\}.\end{aligned} \tag{14}$$

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 $i_0 = 2^n$, $\mathcal{Y}_i = \mathcal{Y}'_i = \{\langle i \rangle^n\}$ 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 (11) becomes

$$d'_i(\mathbf{y}) = d_i(\mathbf{y}) + (-1)^{\langle s \rangle_i}(\overline{\mathbf{y}_1} - \mathbf{y}_1). \tag{15}$$

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 (15),

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

and hence

$$d_{\mathcal{O}}(\mathbf{y}) = d_1 \wedge d_2 \wedge d_4 \tag{16}$$

$$d_{\mathcal{P}}(\mathbf{y}) = d_3 \tag{17}$$

$$d'_{\mathcal{O}}(\mathbf{y}) = [(d_1 \wedge d_2) - \mathbf{y}_1 + \overline{\mathbf{y}_1}] \wedge (d_4 + \mathbf{y}_1 - \overline{\mathbf{y}_1}) \tag{18}$$

$$d'_{\mathcal{P}}(\mathbf{y}) = d_3 - \mathbf{y}_1 + \overline{\mathbf{y}_1}. \tag{19}$$

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

$$\begin{aligned} \mathcal{Y}_1 &= \{d_{\mathcal{O}} \leq d_{\mathcal{P}} < d'_{\mathcal{P}}\} \cup \{d_{\mathcal{O}} \leq d'_{\mathcal{P}} \leq d_{\mathcal{P}}, d'_{\mathcal{O}} \leq d'_{\mathcal{P}}\}, \\ \mathcal{Y}_2 &= \{d_{\mathcal{P}} \leq d'_{\mathcal{P}}, d_{\mathcal{P}} < d_{\mathcal{O}}\} \cup \{d'_{\mathcal{P}} < d_{\mathcal{P}} \leq d_{\mathcal{O}}, d_{\mathcal{P}} \leq d'_{\mathcal{O}}\}, \\ \mathcal{Y}_3 &= \{d'_{\mathcal{P}} = d'_{\mathcal{O}} < d_{\mathcal{P}} = d_{\mathcal{O}}\}, \qquad (20) \\ \mathcal{Y}_4 &= \{d_{\mathcal{P}} = d'_{\mathcal{P}} = d_{\mathcal{O}} < d'_{\mathcal{O}}\}, \\ \mathcal{Y}_5 &= \{d'_{\mathcal{P}} = d_{\mathcal{O}} < d'_{\mathcal{O}} = d_{\mathcal{P}}\}. \qquad (21) \end{aligned}$$

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 described in §III.

Lemma 10. *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 $d = 0, 1, \ldots, n$,

$$\alpha_d^i(C) = |\{\mathbf{y} \in \mathcal{Y}_i : d_C(\mathbf{y}) = d\}|. \tag{22}$$

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 11. *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 12. *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 12, 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 (6),

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

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 12, $\lambda_{C'} \geq \lambda_C$.

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 11 and Corollary 12, we first study $\mathcal{Y}_3$ and $\mathcal{Y}_5$ defined in (20) and (21).

Lemma 13. *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 $\langle 3\rangle$,*

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

Proof. For code C and C' defined above, we have (16) – (19). For $\mathbf{y} \in \mathcal{Y}_3$, $d_3 - \mathbf{y}_1 + \overline{\mathbf{y}_1} < d_3$ implies $\mathbf{y}_1 = 1$, and $d_3 = d_1 \wedge d_2 \wedge d_4$ and $d_3 - 1 = [(d_1 \wedge d_2) - 1] \wedge (d_4 + 1)$ together implies $d_1 \wedge d_2 \leq d_4$ and $d_3 = d_1 \wedge d_2$.

For $\mathbf{y} \in \mathcal{Y}_5$, $d_3 - \mathbf{y}_1 + \overline{\mathbf{y}_1} < d_3$ implies $\mathbf{y}_1 = 1$, and $d_3 - 1 = d_1 \wedge d_2 \wedge d_4$ and $d_3 = [(d_1 \wedge d_2) - 1] \wedge (d_4 + 1)$ together implies $d_4 + 1 \leq (d_1 \wedge d_2) - 1$ and $d_3 - 1 = d_4$. □

1) Characterization of α_i^3: For $\mathbf{y} \in \mathcal{Y}_3$, by Lemma 10, $d_C(\mathbf{y}) = d_3$. By (7) – (10) and Lemma 13, 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

$$w_5 + w_6 < (|5| + |6|)/2, \tag{23}$$
$$w_1 + w_5 = i - (|3| + |6|)/2, \tag{24}$$
$$w_1 + w_5 - w_6 \leq (|1| + |5| - |6|)/2, \tag{25}$$
$$w_3 + w_6 = (|3| + |6|)/2. \tag{26}$$

We have

$$|\mathcal{Y}_3^A(i)| = \sum_{\substack{w_1 \geq 1, w_3, w_5, w_6: \\ (23),(24),(25),(26)}} \binom{|1|-1}{w_1 - 1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}.$$

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

$$w_5 + w_6 \geq (|5| + |6|)/2, \tag{27}$$
$$w_1 + \overline{w_6} = i - (|3| + |5|)/2, \tag{28}$$
$$w_1 + w_5 - w_6 \leq (|1| + |5| - |6|)/2, \tag{29}$$
$$w_3 - w_5 = (|3| - |5|)/2. \tag{30}$$

We have

$$|\mathcal{Y}_3^B(i)| = \sum_{\substack{w_1 \geq 1, w_3, w_5, w_6: \\ (27),(28),(29),(30)}} \binom{|1|-1}{w_1 - 1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}. \tag{31}$$

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 10, $d_C(\mathbf{y}) = d_3 - 1$. By (7) – (10) and Lemma 13, 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

$$w_3 = (n + |3| - 1)/2 - i, \tag{32}$$
$$w_1 + w_5 - w_6 = (|1| + |5| - |6| + 1)/2, \tag{33}$$
$$w_3 + w_6 \geq (|3| + |6|)/2 + 1, \tag{34}$$
$$w_3 - w_5 \geq (|3| - |5|)/2 + 1. \tag{35}$$

Hence

$$\alpha_i^5 = \sum_{\substack{w_1 \geq 1, w_3, w_5, w_6: \\ (32),(33),(34),(35)}} \binom{|1|-1}{w_1 - 1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}.$$

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 6 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

$$\sum_{i=0}^{d-1} \alpha_i^5 = \sum_{\mathcal{W}_5} \binom{|3|}{w_3}\binom{|5|}{\frac{|5|-|6|}{2} + w_6}\binom{|6|}{w_6} \tag{36}$$

where

$$\mathcal{W}_5 = \left\{ \begin{matrix} (34), \\ (33)+(35), \\ w_3 \geq \frac{n+|3|+1}{2} - d, \\ 0 \leq w_3 \leq |3|, 0 \leq w_6 \leq |6| \end{matrix} \right\} = \left\{ \begin{matrix} w_3 + w_6 \geq \frac{|3|+|6|}{2} + 1, \\ w_3 - w_6 \geq \frac{|3|-|6|}{2} + 1, \\ w_3 \geq \frac{n+|3|+1}{2} - d, \\ 0 \leq w_3 \leq |3|, 0 \leq w_6 \leq |6| \end{matrix} \right\}. \tag{37}$$

Similarly,

$$\begin{aligned}\sum_{i=1}^{d}\alpha_i^3 &\geq \sum_{i=1}^{d}|\mathcal{Y}_3^B(i)|\\ &= \sum_{\mathcal{W}_3}\binom{|3|}{w_3}\binom{|5|}{\frac{|5|-|3|}{2}+w_3}\binom{|6|}{w_6}\\ &= \sum_{\mathcal{W}_3'}\binom{|3|}{\frac{|3|-|6|}{2}+w_6'}\binom{|5|}{\frac{|5|-|6|}{2}+w_6'}\binom{|6|}{\frac{|6|-|3|}{2}+w_3'} \quad (38)\end{aligned}$$

where

$$\mathcal{W}_3 = \left\{\begin{array}{c}(27)+(30),\\(29)+(30),\\ w_6\geq\frac{n+|6|+1}{2}-d,\\ 0\leq w_3\leq|3|,0\leq w_6\leq|6|\end{array}\right\} = \left\{\begin{array}{c}w_3+w_6\geq\frac{|3|+|6|}{2}\\ w_3-w_6\leq\frac{|3|-|6|-1}{2}\\ w_6\geq\frac{n+|6|+1}{2}-d,\\ 0\leq w_3\leq|3|,0\leq w_6\leq|6|\end{array}\right\},$$

$$\mathcal{W}_3' = \left\{\begin{array}{c}w_3'+w_6'\geq\frac{|3|+|6|}{2},\quad w_3'-w_6'\geq\frac{|3|-|6|+1}{2}\\ w_3'\geq\frac{n+|3|+1}{2}-d,\\ \frac{|3|-|6|}{2}\leq w_3'\leq\frac{|3|+|6|}{2},\frac{|6|-|3|}{2}\leq w_6'\leq\frac{|3|+|6|}{2}\end{array}\right\},$$

and (38) is obtained by change of variables $w_3'-\frac{|3|}{2}=w_6-\frac{|6|}{2}$ and $w_6'-\frac{|6|}{2}=w_3-\frac{|3|}{2}$.

We show that $\mathcal{W}_5\subset\mathcal{W}_3'$. Due to $|3|\leq|6|$, we have $\frac{|3|-|6|}{2}\leq 0$ and $\frac{|3|+|6|}{2}\geq|3|$. For $(w_3,w_6)\in\mathcal{W}_5$, we have $w_3+w_6\geq\frac{|3|+|6|}{2}+1$, $w_3-w_6\geq\frac{|3|-|6|}{2}+1$ and $0\leq w_3\leq|3|$, which implies $\frac{|6|-|3|}{2}+1\leq w_6\leq\frac{|3|+|6|}{2}-1$. Thus

$$\frac{|3|-|6|}{2}\leq w_3\leq\frac{|3|+|6|}{2},\quad \frac{|6|-|3|}{2}\leq w_6\leq\frac{|3|+|6|}{2},$$

showing $(w_3,w_6)\in\mathcal{W}_3'$.

By Lemma 14 in Appendix B, for $(w_3,w_6)\in\mathcal{W}_5$, we have

$$\binom{|3|}{w_3}\binom{|6|}{w_6}\leq\binom{|3|}{\frac{|3|-|6|}{2}+w_6}\binom{|6|}{\frac{|6|-|3|}{2}+w_3}.$$

Comparing (36) and (38), 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 12, $\lambda_{\mathcal{C}'}\geq\lambda_{\mathcal{C}}$, proving Theorem 6.

C. Class-I Codes with $|1|$ odd, $|3|=0,1$

Here we give a proof of Theorem 7 for the case $|3|=\min\{|3|,|5|,|6|\}$. Otherwise, we can perform row interchanging and column bit flipping (which do not change the ML decoding performance) so that C satisfies the condition.

1) $|3|=0$: When $|3|=0$, we have $w_3=0$ and n is odd. By (32), $\alpha_i^5=0$ if $i\neq\frac{n-1}{2}$. So when $d<\frac{n+1}{2}$, $\sum_{i=0}^{d-1}\alpha_i^5=0$ and hence $\sum_{i=1}^{d}\alpha_i^3\geq\sum_{i=0}^{d-1}\alpha_i^5$; when $d\geq\frac{n+1}{2}$,

$$\begin{aligned}\sum_{i=0}^{d-1}\alpha_i^5 &= \alpha_{\frac{n-1}{2}}^5\\ &= \sum_{\substack{w_1\geq1,(33)+(34),(34),\\ w_5=w_6-w_1+\frac{|1|+|5|-|6|+1}{2},\\ w_5\leq\frac{|5|}{2}-1}}\binom{|1|-1}{w_1-1}\binom{|5|}{w_5}\binom{|6|}{w_6}\\ &\leq \sum_{w_1\geq1,(33)+(34),(34)}\binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|}{2}}\binom{|6|}{w_6}\\ &= \sum_{\substack{w_1\geq1,w_6\geq\frac{|6|}{2}+1,\\ w_6-w_1\leq\frac{|6|-|1|-3}{2}}}\binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|}{2}}\binom{|6|}{w_6}, \quad (39)\end{aligned}$$

where (33) + (34) is $w_6-w_1\leq\frac{|6|-|1|-3}{2}$. Substituting $w_1'=|1|-w_1+1$ into (39), we obtain

$$\alpha_{\frac{n-1}{2}}^5\leq\sum_{\substack{1\leq w_1'\leq|1|,w_6\geq\frac{|6|}{2}+1,\\ w_1'\leq\frac{|1|+|6|-1}{2}-w_6}}\binom{|1|-1}{w_1'-1}\binom{|5|}{\frac{|5|}{2}}\binom{|6|}{w_6} \quad (40)$$

When $d\geq\frac{n+1}{2}$, we further have

$$\begin{aligned}\sum_{i=1}^{d}\alpha_i^3 &\geq \sum_{i=1}^{\frac{n+1}{2}}\alpha_i^3\\ &\geq \sum_{i=1}^{\frac{n+1}{2}}|\mathcal{Y}_3^B(i)|\\ &= \sum_{\substack{w_1\geq1,(27)+(30),\\ (29)+(30),w_1-w_6\leq\frac{|1|-|6|+1}{2}}}\binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|}{2}}\binom{|6|}{w_6}\\ &= \sum_{\substack{w_1\geq1,w_6\geq\frac{|6|}{2},\\ w_1\leq\frac{|1|-|6|-1}{2}+w_6}}\binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|}{2}}\binom{|6|}{w_6}, \quad (41)\end{aligned}$$

where (27) + (30) is $w_6\geq\frac{|6|}{2}$ and (29) + (30) is $w_1-w_6\leq\frac{|1|-|6|}{2}$, which is equivalent to $w_1-w_6\leq\frac{|1|-|6|-1}{2}$ as $|1|-|6|$ is odd.

As $\frac{|1|-|6|-1}{2}+w_6\geq\frac{|1|+|6|-1}{2}-w_6$ when $w_6\geq\frac{|6|}{2}$, comparing the RHS' of (40) and (41), we have $\sum_{i=1}^{d}\alpha_i^3\geq\sum_{i=0}^{d-1}\alpha_i^5$ for $d\geq\frac{n+1}{2}$. By Corollary 12, $\lambda_{\mathcal{C}'}\geq\lambda_{\mathcal{C}}$, proving the case when $|3|=0$.

2) $|3|=1$: When $|3|=1$, we have $w_3=0$ or 1, $|5|$ and $|6|$ are odd, and n is even. In (32), $w_3=1$ when $i=\frac{n}{2}-1$,

and hence

$$\alpha^5_{\frac{n}{2}-1} = \sum_{\substack{w_1\ge1, w_6\ge\frac{|6|+1}{2},(32)+(35),\\ w_5=w_6-w_1+\frac{|1|+|5|-|6|+1}{2},\\ w_5\le\frac{|5|-1}{2}}} \binom{|1|-1}{w_1-1}\binom{|5|}{w_5}\binom{|6|}{w_6}$$

$$\le \sum_{w_1\ge1, w_6\ge\frac{|6|+1}{2},(32)+(35)} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|+1}{2}}\binom{|6|}{w_6}$$

$$= \sum_{\substack{w_1\ge1, w_6\ge\frac{|6|+1}{2},\\ w_6-w_1\le\frac{|6|-|1|-2}{2}}} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|+1}{2}}\binom{|6|}{w_6}, \quad (42)$$

$$= \sum_{\substack{|1|\ge w'_1\ge1, w_6\ge\frac{|6|+1}{2},\\ w'_1\le\frac{|6|+|1|}{2}-w_6}} \binom{|1|-1}{w'_1-1}\binom{|5|}{\frac{|5|+1}{2}}\binom{|6|}{w_6}. \quad (43)$$

where (32) + (35) is $w_6-w_1\le\frac{|6|-|1|-2}{2}$, and (43) is obtained by substituting $w'_1=|1|-w_1+1$ into (42).

In (32), $w_3=0$ when $i=\frac{n}{2}$, and hence

$$\alpha^5_{\frac{n}{2}} = \sum_{\substack{w_1\ge1, w_6\ge\frac{|6|+3}{2},(33)+(35),\\ w_5=w_6-w_1+\frac{|1|+|5|-|6|+1}{2},\\ w_5\le\frac{|5|-3}{2}}} \binom{|1|-1}{w_1-1}\binom{|5|}{w_5}\binom{|6|}{w_6}$$

$$\le \sum_{w_1\ge1, w_6\ge\frac{|6|+3}{2},(33)+(35)} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|-1}{2}}\binom{|6|}{w_6}$$

$$= \sum_{\substack{w_1\ge1, w_6\ge\frac{|6|+3}{2},\\ w_6-w_1\le\frac{|6|-|1|}{2}-2}} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|-1}{2}}\binom{|6|}{w_6} \quad (44)$$

$$= \sum_{\substack{|1|\ge w'_1\ge1, w_6\ge\frac{|6|+3}{2},\\ w'_1\le-w_6+\frac{|6|+|1|-2}{2}}} \binom{|1|-1}{w'_1-1}\binom{|5|}{\frac{|5|-1}{2}}\binom{|6|}{w_6}. \quad (45)$$

where (33) + (35) means $w_6-w_1\le\frac{|6|-|1|}{2}-2$, and (45) is obtained by substituting $w'_1=|1|-w_1+1$ into (44).

Following (31), we have

$$\left|\{w_3=1\}\cap\left(\bigcup_{i\le\frac{n}{2}}\mathcal{Y}_3^B(i)\right)\right| = \sum_{\substack{w_1\ge1,\\ (27)+(30),\ (29)+(30),\\ w_1-w_6\le\frac{n}{2}-\frac{1+|5|}{2}-|6|}} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|+1}{2}}\binom{|6|}{w_6}$$

$$= \sum_{\substack{w_1\ge1, w_6\ge\frac{|6|-1}{2},\\ w_1\le w_6+\frac{|1|-|6|-1}{2}}} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|+1}{2}}\binom{|6|}{w_6} \quad (46)$$

where (27) + (30) implies $w_6\ge\frac{|6|-1}{2}$, (29) + (30) implies $w_1-w_6\le\frac{|1|-|6|-1}{2}$, and (46) follows that $\frac{|1|-|6|-1}{2}\le\frac{n}{2}-\frac{1+|5|}{2}-|6|$. Since $w_6+\frac{|1|-|6|-1}{2}\ge-w_6+\frac{|6|+|1|}{2}$ when $w_6\ge\frac{|6|+1}{2}$, comparing the RHS' of (43) and (46), we get

$$\alpha^5_{\frac{n}{2}-1}\le\left|\{w_3=1\}\cap\left(\bigcup_{i\le\frac{n}{2}}\mathcal{Y}_3^B(i)\right)\right|. \quad (47)$$

Following (31), we have

$$\left|\{w_3=0\}\cap\left(\bigcup_{i\le\frac{n}{2}+1}\mathcal{Y}_3^B(i)\right)\right| = \sum_{\substack{w_1\ge1,\\ (27)+(30),\ (29)+(30),\\ w_1-w_6\le\frac{n}{2}+1-\frac{1+|5|}{2}-|6|}} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|-1}{2}}\binom{|6|}{w_6}$$

$$= \sum_{\substack{w_1\ge1, w_6:w_6\ge\frac{|6|+1}{2},\\ w_1\le w_6+\frac{|1|-|6|+1}{2}}} \binom{|1|-1}{w_1-1}\binom{|5|}{\frac{|5|-1}{2}}\binom{|6|}{w_6}, \quad (48)$$

where (27) + (30) implies $w_6\ge\frac{|6|+1}{2}$, (29) + (30) implies $w_1-w_6\le\frac{|1|-|6|+1}{2}$, and (48) follows that $\frac{|1|-|6|+1}{2}\le\frac{n}{2}+1-\frac{1+|5|}{2}-|6|$. Since $w_6+\frac{|1|-|6|+1}{2}\ge-w_6+\frac{|6|+|1|-2}{2}$ when $w_6\ge\frac{|6|+1}{2}$, comparing the RHS' of (45) and (48), we get

$$\alpha^5_{\frac{n}{2}}\le\left|\{w_3=0\}\cap\left(\bigcup_{i\le\frac{n}{2}+1}\mathcal{Y}_3^B(i)\right)\right|. \quad (49)$$

When $d<\frac{n}{2}$, $\sum_{i=0}^{d-1}\alpha_i^5=0\le\sum_{i=1}^{d}\alpha_i^3$. When $d=\frac{n}{2}$, by (47),

$$\sum_{i=0}^{d-1}\alpha_i^5=\alpha^5_{\frac{n}{2}-1}\le\left|\{w_3=1\}\cap\left(\bigcup_{i\le\frac{n}{2}}\mathcal{Y}_3^B(i)\right)\right|\le\sum_{i=1}^{\frac{n}{2}}\alpha_i^3.$$

When $d\ge\frac{n}{2}+1$, by (47) and (49),

$$\sum_{i=0}^{d-1}\alpha_i^5=\alpha^5_{\frac{n}{2}-1}+\alpha^5_{\frac{n}{2}}$$
$$\le\left|\{w_3=1\}\cap\left(\bigcup_{i\le\frac{n}{2}}\mathcal{Y}_3^B(i)\right)\right|+\left|\{w_3=0\}\cap\left(\bigcup_{i\le\frac{n}{2}+1}\mathcal{Y}_3^B(i)\right)\right|$$
$$\le\sum_{i=1}^{d}\alpha_i^3.$$

Thus we have $\sum_{i=1}^{d}\alpha_i^3\ge\sum_{i=0}^{d-1}\alpha_i^5$ for $1\le d\le n$. By Corollary 12, $\lambda_{\mathcal{C}'}\ge\lambda_{\mathcal{C}}$, proving the case when $|3|=1$.

VI. Proof of Theorem 4

In this section, we prove Theorem 4 using another case of the general approach in §III. Let C be an $(n,2)$ code with the first two columns $\langle1\rangle$ and $\langle7\rangle$. Let C' be the code obtained by flipping the first two bits of $\mathbf{c}_3$, so that the first two columns of C' are $\langle3\rangle$ and $\langle5\rangle$. (Other cases of Theorem 4 can be converted to this case by interchanging rows.)

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

$$d'_i(\mathbf{y})=d_i(\mathbf{y})+(-1)^{\langle1\rangle_i}(\overline{\mathbf{y}_1}-\mathbf{y}_1)+(-1)^{\langle7\rangle_i}(\overline{\mathbf{y}_2}-\mathbf{y}_2).$$

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

$$d'_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}). \tag{50}$$

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

$$d'_2(\mathbf{y}) - d_2(\mathbf{y}) = d'_3(\mathbf{y}) - d_3(\mathbf{y}) = \pm 2, \tag{51}$$

and hence

$$\begin{aligned}&(d'_{\mathcal{O}}(\mathbf{y}) - d_{\mathcal{O}}(\mathbf{y}))(d'_{\mathcal{P}}(\mathbf{y}) - d_{\mathcal{P}}(\mathbf{y}))\\ &= (d'_{\{1,2,4\}}(\mathbf{y}) - d_{\{1,2,4\}}(\mathbf{y}))(d'_3(\mathbf{y}) - d_3(\mathbf{y})) \geq 0.\end{aligned} \tag{52}$$

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 $\{0,1\}^n$: First, we show that

$$\mathcal{Y}_4 \cup \mathcal{Y}_5 = \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}, d_{\mathcal{P}} > d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\}, \tag{53}$$

and 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 $\{0,1\}^n$.

To show (53), since $\mathcal{Y}_4 \subseteq \{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\}$, we have

$$\begin{aligned}\mathcal{Y}_4 &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d'_{\mathcal{P}} < d_{\mathcal{O}} \wedge d'_{\mathcal{O}} < d_{\mathcal{P}}\} \cap \{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\}\\ &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}, d_{\mathcal{P}} > d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\} \cap\\ &\quad \{d_{\mathcal{O}} \wedge d'_{\mathcal{O}} > d'_{\mathcal{P}}\}.\end{aligned} \tag{54}$$

Denote

$$\begin{aligned}\mathcal{A}_1 = \{&\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}, d_{\mathcal{P}} > d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\} \cap\\ &\{d_{\mathcal{O}} \wedge d'_{\mathcal{O}} \leq d'_{\mathcal{P}}\}.\end{aligned} \tag{55}$$

For $\mathbf{y} \in \mathcal{A}_1$, we have $d_{\mathcal{O}}(\mathbf{y}) > d'_{\mathcal{O}}(\mathbf{y})$ which implies $d_{\mathcal{P}}(\mathbf{y}) > d'_{\mathcal{P}}(\mathbf{y})$ by (51) and (52), and hence

$$d'_{\mathcal{O}}(\mathbf{y}) \leq d_{\mathcal{P}}(\mathbf{y}) \wedge d'_{\mathcal{P}}(\mathbf{y}) < d_{\mathcal{O}}(\mathbf{y}), \quad d'_{\mathcal{O}}(\mathbf{y}) < d_{\mathcal{P}}(\mathbf{y}).$$

Thus we have $\mathbf{y} \in \mathcal{Y}_5$ and then $\mathcal{A}_1 \subseteq \mathcal{Y}_5$. For $\mathbf{y} \in \mathcal{Y}_5$, we have $d_{\mathcal{O}}(\mathbf{y}) > d'_{\mathcal{O}}(\mathbf{y})$ by the definition above, which implies $d_{\mathcal{P}}(\mathbf{y}) > d'_{\mathcal{P}}(\mathbf{y})$ by (51) and (52). Then we obtain

$$d_{\mathcal{P}}(\mathbf{y}) > d_{\mathcal{O}}(\mathbf{y}) \wedge d'_{\mathcal{O}}(\mathbf{y}) = d'_{\mathcal{O}}(\mathbf{y}) \leq d'_{\mathcal{P}}(\mathbf{y}) < d_{\mathcal{O}}(\mathbf{y}).$$

Thus $\mathbf{y} \in \mathcal{A}_1$ and then $\mathcal{Y}_5 \subseteq \mathcal{A}_1$. Therefore, $\mathcal{Y}_5 = \mathcal{A}_1$. From (54) and (55), we obtain (53).

We further show that

$$\mathcal{Y}'_3 \cup \mathcal{Y}'_4 \subseteq \mathcal{Y}_3 \cup \mathcal{Y}_4. \tag{56}$$

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 $\{0,1\}^n$.

To show (56), we see

$$\mathcal{Y}'_4 = \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{P}} < d_{\mathcal{O}} \wedge d'_{\mathcal{O}} < d'_{\mathcal{P}}\} \subseteq \mathcal{Y}_3, \tag{57}$$

and

$$\begin{aligned}\mathcal{Y}'_3 \setminus \mathcal{Y}_4 &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d'_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}, d'_{\mathcal{P}} \leq d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\} \cap\\ &\quad (\{d_{\mathcal{O}} \wedge d'_{\mathcal{O}} \geq d_{\mathcal{P}}\} \cup \{d'_{\mathcal{P}} \geq d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\})\\ &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d'_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}, d'_{\mathcal{P}} \leq d_{\mathcal{O}} \wedge d'_{\mathcal{O}},\\ &\quad d_{\mathcal{O}} \wedge d'_{\mathcal{O}} \geq d_{\mathcal{P}}\} \cup \{\mathbf{y}_1 \neq \mathbf{y}_2, d'_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}},\\ &\quad d'_{\mathcal{P}} \leq d_{\mathcal{O}} \wedge d'_{\mathcal{O}}, d'_{\mathcal{P}} \geq d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\}\\ &= \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{P}} \wedge d'_{\mathcal{P}} < \max(d_{\mathcal{P}}, d'_{\mathcal{P}}) \leq d_{\mathcal{O}} \wedge d'_{\mathcal{O}}\} \cup\\ &\quad \{\mathbf{y}_1 \neq \mathbf{y}_2, d_{\mathcal{O}} \wedge d'_{\mathcal{O}} = d'_{\mathcal{P}}, d_{\mathcal{P}} \wedge d'_{\mathcal{P}} < d'_{\mathcal{O}}\}\end{aligned} \tag{58}$$

where in the last equality $d_{\mathcal{P}} \wedge d'_{\mathcal{P}} < \max(d_{\mathcal{P}}, d'_{\mathcal{P}})$ follows from (51). By (52), when $\mathbf{y}_1 \neq \mathbf{y}_2$, if $d_{\mathcal{O}} < d'_{\mathcal{O}}$, then $d_{\mathcal{P}} < d'_{\mathcal{P}}$; and if $d_{\mathcal{P}} > d'_{\mathcal{P}}$, then $d_{\mathcal{O}} \geq d'_{\mathcal{O}}$. Hence, we can verify that both terms to union in (58) are subsets of $\mathcal{Y}_3$. Therefore, $\mathcal{Y}'_3 \setminus \mathcal{Y}_4 \subset \mathcal{Y}_3$, which together with (57), proves (56).

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 $\lambda_{C'} \geq \lambda_C$.

The above claims are justified as follows:

1) For $\mathbf{y} \in \mathcal{Y}_1$, as $\mathbf{y}_1 = \mathbf{y}_2$, we have $d'_{\mathcal{P}} = d_{\mathcal{P}}$ by (50), and hence
$$d_C(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y}) \wedge d'_{\mathcal{P}}(\mathbf{y}) = d_{C'}(\mathbf{y}).$$
2) For $\mathbf{y} \in \mathcal{Y}_2$, by the definition of $\mathcal{Y}_2$, we have $d_{\mathcal{O}} \leq d_{\mathcal{P}} \wedge d'_{\mathcal{P}}$, and hence $d_C(\mathbf{y}) = d_{C'}(\mathbf{y}) = d_{\mathcal{O}}$.
3) For $\mathbf{y} \in \mathcal{Y}_3$, we have $d_{\mathcal{P}} \leq d_{\mathcal{O}} \wedge d'_{\mathcal{O}}$ by the definition of $\mathcal{Y}_3$. We then have
$$\begin{aligned}d_C(\mathbf{y}) &= d_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}),\\ d_{C'}(f_2(\mathbf{y})) &= d'_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}).\end{aligned}$$
4) For $\mathbf{y} \in \mathcal{Y}_4$, we have $\mathbf{y}_1 \neq \mathbf{y}_2, d'_{\mathcal{P}} < d_{\mathcal{O}} \wedge d'_{\mathcal{O}} < d_{\mathcal{P}}$ by the definition of $\mathcal{Y}_4$. By (52), $d_{\mathcal{O}} \wedge d'_{\mathcal{O}} = d'_{\mathcal{O}}$, which implies $d_{C'}(f_2(\mathbf{y})) = d'_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d'_{\mathcal{O}}(\mathbf{y})$ and hence
$$d_C(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) \geq d'_{\mathcal{O}}(\mathbf{y}) = d_{C'}(f_2(\mathbf{y})).$$
5) For $\mathbf{y} \in \mathcal{Y}_5$, we have $\mathbf{y}_1 \neq \mathbf{y}_2, d'_{\mathcal{O}} \leq d_{\mathcal{P}} \wedge d'_{\mathcal{P}} < d_{\mathcal{O}}$. By (51) and (52), $d'_{\mathcal{P}} < d_{\mathcal{P}}$. Then we have
$$d_C(\mathbf{y}) = d_{\mathcal{O}} \wedge d_{\mathcal{P}} \geq d_{\mathcal{O}} \wedge d'_{\mathcal{P}} = d_{C'}(\mathbf{y})$$
and
$$d_{C'}(\mathbf{y}) = d'_{\mathcal{P}}(\mathbf{y}).$$

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.

Appendix A
Proofs

Proof of Lemma 10. By checking the definition, we see that $\mathcal{Y}_1, \dots \mathcal{Y}_5$ are all disjoint. To show they form a partition, we can verify that

$$\mathcal{Y}_1 \cup \mathcal{Y}_4 \cup \mathcal{Y}_5 = \{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\},$$
$$\mathcal{Y}_2 \cup \mathcal{Y}_3 = \{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\}$$

and hence $(\mathcal{Y}_1 \cup \mathcal{Y}_4 \cup \mathcal{Y}_5) \cup (\mathcal{Y}_2 \cup \mathcal{Y}_3) = \{0,1\}^n$.

We first prove that $\mathcal{Y}_1 \cup \mathcal{Y}_4 \cup \mathcal{Y}_5 = \{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\}$. Notice that the three sets can be rewritten as

$$\begin{aligned}
\mathcal{Y}_1 &= \{d_{\mathcal{O}} \le d_{\mathcal{P}} < d'_{\mathcal{P}}\} \cup \{d_{\mathcal{O}} \le d'_{\mathcal{P}} \le d_{\mathcal{P}}, d'_{\mathcal{O}} \le d'_{\mathcal{P}}\} \\
&= (\{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d_{\mathcal{P}} < d'_{\mathcal{P}}\}) \cup (\{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \\
&\quad \cap \{d'_{\mathcal{P}} \le d_{\mathcal{P}}, d'_{\mathcal{O}} \le d'_{\mathcal{P}}\}), \qquad (59) \\
\mathcal{Y}_4 &= \{d_{\mathcal{P}} = d'_{\mathcal{P}} = d_{\mathcal{O}} < d'_{\mathcal{O}}\} \\
&\overset{(a)}{=} \{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d'_{\mathcal{P}} \le d_{\mathcal{P}}, d'_{\mathcal{O}} > d'_{\mathcal{P}}, d_{\mathcal{P}} = d'_{\mathcal{P}}\}, \qquad (60) \\
\mathcal{Y}_5 &= \{d'_{\mathcal{P}} = d_{\mathcal{O}} < d'_{\mathcal{O}} = d_{\mathcal{P}}\} \\
&\overset{(b)}{=} \{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d'_{\mathcal{P}} \le d_{\mathcal{P}}, d'_{\mathcal{O}} > d'_{\mathcal{P}}, d_{\mathcal{P}} > d'_{\mathcal{P}}\}. \qquad (61)
\end{aligned}$$

For $\forall \mathbf{y}$, we have

$$|d_{\mathcal{S}}(\mathbf{y}) - d'_{\mathcal{S}}(\mathbf{y})| \le 1, \qquad (62)$$

which can be obtained by the definition. Then if $d_{\mathcal{O}} \le d'_{\mathcal{P}} < d'_{\mathcal{O}}$, we will have $d_{\mathcal{O}} = d'_{\mathcal{P}}$ and thus the equality (a) in (60) holds. Furthermore, if $d'_{\mathcal{O}} > d'_{\mathcal{P}}$ we have $d_{\mathcal{O}} \ge d'_{\mathcal{P}}$, and if $d'_{\mathcal{O}} > d'_{\mathcal{P}}$, we have $d_{\mathcal{O}} \ge d'_{\mathcal{P}}$ and thus the equality (b) in (61) holds. By (60) and (61), we have

$$\mathcal{Y}_4 \cup \mathcal{Y}_5 = \{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d'_{\mathcal{P}} \le d_{\mathcal{P}}, d'_{\mathcal{O}} > d'_{\mathcal{P}}\}.$$

From (59), this further implies

$$\mathcal{Y}_1 \cup \mathcal{Y}_4 \cup \mathcal{Y}_5 = \{d_{\mathcal{O}} \le d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\}.$$

We now prove $\mathcal{Y}_2 \cup \mathcal{Y}_3 = \{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\}$. First we rewrite the two sets as

$$\begin{aligned}
\mathcal{Y}_2 &= \{d_{\mathcal{P}} \le d'_{\mathcal{P}}, d_{\mathcal{P}} < d_{\mathcal{O}}\} \cup \{d'_{\mathcal{P}} < d_{\mathcal{P}} \le d_{\mathcal{O}}, d_{\mathcal{P}} \le d'_{\mathcal{O}}\} \\
&= (\{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d_{\mathcal{P}} \le d'_{\mathcal{P}}\}) \cup (\{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \\
&\quad \cap \{d_{\mathcal{P}} > d'_{\mathcal{P}}, d_{\mathcal{P}} \le d_{\mathcal{O}}, d_{\mathcal{P}} \le d'_{\mathcal{O}}\}) \\
&\overset{(a)}{=} (\{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d_{\mathcal{P}} \le d'_{\mathcal{P}}\}) \cup (\{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \\
&\quad \cap \{d_{\mathcal{P}} > d'_{\mathcal{P}}, d_{\mathcal{P}} \le d'_{\mathcal{O}}\}), \qquad (63) \\
\mathcal{Y}_3 &= \{d'_{\mathcal{P}} = d'_{\mathcal{O}} < d_{\mathcal{P}} = d_{\mathcal{O}}\} \\
&\overset{(b)}{=} \{d_{\mathcal{O}} > d_{\mathcal{P}} \wedge d'_{\mathcal{P}}\} \cap \{d_{\mathcal{P}} > d'_{\mathcal{P}}, d_{\mathcal{P}} > d'_{\mathcal{O}}\}. \qquad (64)
\end{aligned}$$

By (62), we can get $d_{\mathcal{P}} \le d_{\mathcal{O}}$ if $d_{\mathcal{O}} > d'_{\mathcal{P}}, d_{\mathcal{P}} > d'_{\mathcal{P}}$. Then the equality (a) in (63) holds. Similarly we can justify (b) in (64) by (62).

By definition,

$$\begin{aligned}
\mathcal{Y}'_2 &= \{d'_{\mathcal{P}} \le d_{\mathcal{P}}, d'_{\mathcal{P}} < d'_{\mathcal{O}}\} \cup \{d_{\mathcal{P}} < d'_{\mathcal{P}} \le d'_{\mathcal{O}}, d'_{\mathcal{P}} \le d_{\mathcal{O}}\}, \\
\mathcal{Y}'_4 &= \{d_{\mathcal{P}} = d'_{\mathcal{P}} = d'_{\mathcal{O}} < d_{\mathcal{O}}\}, \\
\mathcal{Y}'_5 &= \{d_{\mathcal{P}} = d'_{\mathcal{O}} < d_{\mathcal{O}} = d'_{\mathcal{P}}\}.
\end{aligned}$$

It can be verified that $(\mathcal{Y}'_2 \cup \mathcal{Y}'_4 \cup \mathcal{Y}'_5) \cap (\mathcal{Y}_1 \cup \mathcal{Y}_3) = \emptyset$. As f_1 is a one-to-one mapping, $\mathcal{Y}'_2 \cup \mathcal{Y}'_4 \cup \mathcal{Y}'_5 = \mathcal{Y}_2 \cup \mathcal{Y}_4 \cup \mathcal{Y}_5$. Hence, we conclude that $\mathcal{Y}_1, \mathcal{Y}'_2, \mathcal{Y}_3, \mathcal{Y}'_4, \mathcal{Y}'_5$ form a partition of $\{0,1\}^n$.

We use the following facts in the proof of claim 1) – 5).

$$\begin{aligned}
d_C(\mathbf{y}) &= \min\{d_{\mathcal{O}}(\mathbf{y}), d_{\mathcal{P}}(\mathbf{y})\} \\
d_{C'}(\mathbf{y}) &= \min\{d_{\mathcal{O}}(\mathbf{y}), d'_{\mathcal{P}}(\mathbf{y})\} \\
d_{C'}(f_1(\mathbf{y})) &= \min\{d'_{\mathcal{O}}(\mathbf{y}), d_{\mathcal{P}}(\mathbf{y})\}.
\end{aligned}$$

To prove the claim 1), for $\mathbf{y} \in \mathcal{Y}_1$, by the definition of $\mathcal{Y}_1$, $d_{\mathcal{O}} \le \min\{d_{\mathcal{P}}, d'_{\mathcal{P}}\}$. Hence $d_C(\mathbf{y}) = d_{C'}(\mathbf{y}) = d_{\mathcal{O}}$. To prove claim 2), for $\mathbf{y} \in \mathcal{Y}_2$, by the definition of $\mathcal{Y}_2$, $d_{\mathcal{P}} \le \min\{d_{\mathcal{O}}, d'_{\mathcal{O}}\}$, and hence $d_C(\mathbf{y}) = d_{\mathcal{P}}$. Further,

$$\begin{aligned}
d_{C'}(\mathbf{y}') &= d_{\mathcal{O}}(\mathbf{y}') \wedge d'_{\mathcal{P}}(\mathbf{y}') \\
&= d'_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}).
\end{aligned}$$

To prove claim 3), for $\mathbf{y} \in \mathcal{Y}_3$, $d_C(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y})$ by the definition of $\mathcal{Y}_3$. Moreover,

$$d_{\mathbf{C}'}(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y}) \wedge d'_{\mathcal{P}}(\mathbf{y}) = d'_{\mathcal{P}}(\mathbf{y}) < d_C(\mathbf{y}).$$

By (62), we have $d_{\mathcal{P}} = d'_{\mathcal{P}} + 1$. To prove claim 4), for $\mathbf{y} \in \mathcal{Y}_4$, by the definition of $\mathcal{Y}_4$,

$$\begin{aligned}
d_C(\mathbf{y}) &= d_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}), \\
d_{C'}(\mathbf{y}') &= d'_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{P}}(\mathbf{y}).
\end{aligned}$$

To prove claim 5), for $\mathbf{y} \in \mathcal{Y}_5$, $d_C(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d_{\mathcal{O}}(\mathbf{y})$ by the definition of $\mathcal{Y}_5$. Moreover,

$$d_{\mathbf{C}'}(\mathbf{y}') = d'_{\mathcal{O}}(\mathbf{y}) \wedge d_{\mathcal{P}}(\mathbf{y}) = d'_{\mathcal{O}}(\mathbf{y}) > d_C(\mathbf{y}).$$

By (62), we have $d'_{\mathcal{O}} = d_{\mathcal{O}} + 1$. □

Proof of Theorem 11. 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 $\alpha_d(C') = \sum_{i=1}^5 \alpha_d^i(C')$ where

$$\begin{aligned}
\alpha_d^1(C') &= |\{\mathbf{y} \in \mathcal{Y}_1 : d_{C'}(\mathbf{y}) = d\}| = \alpha_d^1(C), \\
\alpha_d^2(C') &= |\{\mathbf{y} \in \mathcal{Y}'_2 : d_{C'}(\mathbf{y}) = d\}| = \alpha_d^2(C), \\
\alpha_d^3(C') &= |\{\mathbf{y} \in \mathcal{Y}_3 : d_{C'}(\mathbf{y}) = d\}| = \begin{cases} \alpha_{d+1}^3(C) & d < n \\ 0 & d = n \end{cases}, \\
\alpha_d^4(C') &= |\{\mathbf{y} \in \mathcal{Y}'_4 : d_{C'}(\mathbf{y}) = d\}| = \alpha_d^4(C), \\
\alpha_d^5(C') &= |\{\mathbf{y} \in \mathcal{Y}'_5 : d_{C'}(\mathbf{y}) = d\}| = \begin{cases} \alpha_{d-1}^5(C) & d \ge 1 \\ 0 & d = 0 \end{cases}.
\end{aligned}$$

The second equality in each line follows from Lemma 10. Together with (22), we write

$$\begin{aligned}\lambda_{\mathcal{C}'} - \lambda_{\mathcal{C}} &= \frac{1}{|\mathcal{C}|}\sum_{d=0}^{n}(\alpha_d(C') - \alpha_d(C))(1-\epsilon)^{n-d}\epsilon^d \\ &= \frac{1}{|\mathcal{C}|}\sum_{d=0}^{n}\sum_{i=1}^{5}(\alpha_d^i(C') - \alpha_d^i(C))(1-\epsilon)^{n-d}\epsilon^d \\ &= \frac{1}{|\mathcal{C}|}\sum_{d=0}^{n}\sum_{i=3,5}(\alpha_d^i(C') - \alpha_d^i(C))(1-\epsilon)^{n-d}\epsilon^d,\end{aligned}$$

By substituting $\alpha_d^3(C') = \alpha_{d+1}^3(C)$ and $\alpha_d^5(C') = \alpha_{d-1}^5(C)$, we see that $\lambda_{\mathcal{C}'} \geq \lambda_{\mathcal{C}}$ if and only if

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

where the LHS can be further simplified as

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

The theorem is proved by checking that in the above argument, the relation $\geq$ can be replaced by $>$. □

Proof of Corollary 12. Let $\epsilon_0 = \frac{\epsilon}{1-\epsilon}$ and let $\Psi_d = \sum_{i=1}^{d}\left[\alpha_i^3(C) - \alpha_{i-1}^5(C)\right]$ for $d = 1, \ldots, n$ and $\Psi_0 = 0$. Write

$$\begin{aligned}&\sum_{d=1}^{n}[\alpha_d^3(C) - \alpha_{d-1}^5(C)]\left(\frac{\epsilon}{1-\epsilon}\right)^{d-1} \\ &= \sum_{d=1}^{n}(\Psi_d - \Psi_{d-1})\epsilon_0^{d-1} \\ &= \Psi_n\epsilon_0^{n-1} + \sum_{d=1}^{n-1}\Psi_d(\epsilon_0^{d-1} - \epsilon_0^d).\end{aligned}$$

Note that for $0 < \epsilon < \frac{1}{2}$, $\epsilon_0^d = \left(\frac{\epsilon}{1-\epsilon}\right)^d$ is a strictly decreasing function of d. By Theorem 11, we can prove the sufficient conditions of the corollary. □

APPENDIX B
A LEMMA

Lemma 14. *Suppose $|3| \leq |6|$ of the same parity. For $(w_3, w_5) \in \mathcal{W}_5$ (defined in (37)),*

$$\binom{|3|}{w_3}\binom{|6|}{w_6} \leq \binom{|3|}{\frac{|3|-|6|}{2} + w_6}\binom{|6|}{\frac{|6|-|3|}{2} + w_3}.$$

Proof. Let $\hat{w}_i = w_i - \frac{|i|}{2}$, $i = 3, 6$. The inequality to prove becomes

$$\binom{|3|}{\frac{|3|}{2} + \hat{w}_3}\binom{|6|}{\frac{|6|}{2} + \hat{w}_6} \leq \binom{|3|}{\frac{|3|}{2} + \hat{w}_6}\binom{|6|}{\frac{|6|}{2} + \hat{w}_3}.$$

We have by the definition of $\mathcal{W}_5$ in (37),

$$\hat{w}_3 + \hat{w}_6 \geq 1, \hat{w}_3 - \hat{w}_6 \geq 1, \hat{w}_3 \geq \frac{n+1}{2} - d.$$

We write

$$\begin{aligned}&\frac{\binom{|3|}{\frac{|3|}{2}+\hat{w}_3}\binom{|6|}{\frac{|6|}{2}+\hat{w}_6}}{\binom{|3|}{\frac{|3|}{2}+\hat{w}_6}\binom{|6|}{\frac{|6|}{2}+\hat{w}_3}} = \frac{\frac{|3|\cdots(\frac{|3|}{2}-\hat{w}_3+1)}{(\frac{|3|}{2}+\hat{w}_3)!}\frac{|6|\cdots(\frac{|6|}{2}-\hat{w}_6+1)}{(\frac{|6|}{2}+\hat{w}_6)!}}{\frac{|3|\cdots(\frac{|3|}{2}-\hat{w}_6+1)}{(\frac{|3|}{2}+\hat{w}_6)!}\frac{|6|\cdots(\frac{|6|}{2}-\hat{w}_3+1)}{(\frac{|6|}{2}+\hat{w}_3)!}} \\ &= \frac{(\frac{|3|}{2}-\hat{w}_6)\cdots(\frac{|3|}{2}-\hat{w}_3+1)}{(\frac{|3|}{2}+\hat{w}_3)\cdots(\frac{|3|}{2}+\hat{w}_6+1)} \cdot \frac{(\frac{|6|}{2}+\hat{w}_3)\cdots(\frac{|6|}{2}+\hat{w}_6+1)}{(\frac{|6|}{2}-\hat{w}_6)\cdots(\frac{|6|}{2}-\hat{w}_3+1)} \\ &= \prod_{i=1}^{\hat{w}_3-\hat{w}_6}\frac{\frac{|3|}{2}-\hat{w}_3+i}{\frac{|3|}{2}+\hat{w}_6+i}\frac{\frac{|6|}{2}+\hat{w}_6+i}{\frac{|6|}{2}-\hat{w}_3+i} \\ &= \prod_{i=1}^{\hat{w}_3-\hat{w}_6}\frac{(\frac{|3|}{2}+i)(\frac{|6|}{2}+i) - \hat{w}_3\hat{w}_6 + (\frac{|3|}{2}+i)\hat{w}_6 - \hat{w}_3(\frac{|6|}{2}+i)}{(\frac{|3|}{2}+i)(\frac{|6|}{2}+i) - \hat{w}_3\hat{w}_6 - (\frac{|3|}{2}+i)\hat{w}_3 + \hat{w}_6(\frac{|6|}{2}+i)} \\ &\leq 1,\end{aligned}$$

where the last inequality is obtained by comparing the last two terms of the denominator and the nominator:

$$\begin{aligned}&\left(\frac{|3|}{2}+i\right)\hat{w}_6 - \hat{w}_3\left(\frac{|6|}{2}+i\right) - \left(-\left(\frac{|3|}{2}+i\right)\hat{w}_3 + \hat{w}_6\left(\frac{|6|}{2}+i\right)\right) \\ &= (\hat{w}_3+\hat{w}_6)\left(\frac{|3|}{2} - \frac{|6|}{2}\right) \\ &\leq 0\end{aligned}$$

where the inequality follows from $\hat{w}_3 + \hat{w}_6 \geq 1$ and $|3| \leq |6|$. □

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