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 300, 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]. For given $n$ and $k$, 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, traced back to the early days of information and coding theory [2], [3].

For given $n$ and $k$, if perfect or quasi-perfect binary $(n, k)$ codes exist, they are optimal for BSC [3]. For example, the optimal $(n, 1)$ codes are either perfect or quasi-perfect and hence are known [2]. Readers can find more about optimal codes in [2], [3]. More recently, Chen, Lin and Moser [4] gave the first proof of the optimal binary codes of 3 codewords for any block length $n$.

In this paper, we study binary $(n, 2)$ codes. The best linear $(n, 2)$ codes have been explicitly characterized for each block length $n$ [4], [5], and are conjectured to be optimal among all $(n, 2)$ codes in terms of the ML decoding [4]. 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 300, numerical evaluations show that linear codes are optimal, where the evaluation complexity is
Moreover, most ML decoding comparison results obtained in this paper about \((n, 2)\) codes 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 codeword different in one bit (see §IV) or 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 \(C\) 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 \(C\) as rows, we can form a \(2^k \times n\) binary matrix \(C\), which is used interchangeably with \(C\). For \(i = 1, \ldots, 2^k\), let \(c_i\) be the \(i\)th row of \(C\), i.e., a codeword of \(C\).

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

\[
d_C(y) = \min_{c \in C} w(c \oplus y).
\]

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

\[
p(y|x) = (1-\epsilon)^{w(x \oplus y)} \epsilon^{w(x \oplus y)}.
\]

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

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

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

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

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 \(\epsilon\), 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 have 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,

\[
\langle 1 \rangle^4 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T, \quad \langle 2 \rangle^4 = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T.
\]
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 \( i \)-th column of type \( \langle i \rangle \) and \( |i| = 1 \) for \( i = 1, \ldots, 7 \).

The analysis of the column types of \( C \) has been used in literature [4], [5]. Chen, Lin and Moser [4] compared different codes by induction in \( n \), i.e., increasing one column a time. Here, we compare two codes of the same length with difference in one or two positions in one codeword, and we find that it is also convenient to use the column representation in our analysis. The following facts about \((n,2)\) codes are straightforward [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 \( C \) of codewords \( c_1, \ldots, c_4 \) with \( w(c_i \oplus c_j) \) 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 §IV-B.

For example, suppose an \((n,2)\) code \( C \) has a column \( \langle 1 \rangle \) and \( w(c_3 \oplus c_4) \) 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

\[
w(c_2 \oplus c_3) = |2| + |3| + |4| + |5| \\
w(c_3 \oplus c_4) = |1| + |2| + |5| + |6| \\
w(c_2 \oplus c_4) = |1| + |3| + |4| + |6|.
\]

We claim that one of the above three weights must be even. For example, assume \( w(c_2 \oplus c_3) \) is odd. Then \( |3| + |4| \) and \( |2| + |5| \) are of different parity, so that one of \( w(c_3 \oplus c_4) \) and \( w(c_2 \oplus c_4) \) must be even.

Suppose \( w(c_2 \oplus c_3) \) is even. As \( |2| \) is positive, Theorem [II] implies a better code with \( |2| \) smaller and \( |6| \) bigger. 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:

\[
\begin{align*}
w(c_1 \oplus c_4) &= |1| + |3| + |5| \quad (2) \\
w(c_3 \oplus c_4) &= |1| + |5| + |6| \quad (3) \\
w(c_2 \oplus c_4) &= |1| + |3| + |6| \quad (4)
\end{align*}
\]

When \(|1|\) is odd, \(|3|, |5|, |6|\) 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 cases, 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 \S \[VI\]

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. Consider an arbitrary \((n, 2)\) code \(C\). 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 \(0 < |7| < |1| + |2| + |4|\) 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 $(1)$ column of $C$ by $(s)$, where $s = \arg\min_{i=3,5,6} |i|_C$. Then $\lambda_{C'} \geq \lambda_C$.

*Proof.* See §V-B. \hfill $\Box$

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 $(1)$ column of $C$ by $(s)$, where $s \in \{3, 5, 6\}$ has $|s|_C = 0$ or $1$. Then $\lambda_{C'} \geq \lambda_C$.

*Proof.* See §V-C. \hfill $\Box$

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 exist 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. \hfill $\Box$

**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 $(1)$ column of $C$ by $(s)$, where $s \in \{3, 5, 6\}$ with $|s|_C = 0$ or $1$. By Theorem 8 linear $(n, 2)$ codes are optimal for $n \leq 8$. \hfill $\Box$

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 $(1)$ column of $C$ by $(s)$ with $s = \arg\min_{i=3,5,6} |i|_C$. See an algorithm for verifying the optimality of linear $(n, 2)$ codes in §V-D where the evaluation complexity is $O(n^7)$. We have verified that for $n \leq 300$, linear codes are optimal.
III. An Approach of Comparing Two $(n, 2)$ Codes

We further define some notations. Let $C$ be an $(n, 2)$ code with the $j$th codeword/row $c_j$, $j = 1, \ldots, 4$. Denote

\begin{equation}
    d_j(y) = w(c_j \oplus y).
\end{equation}

For a binary vector $y$, denote $(y)_i$ or $y_i$ as the $i$th entry of $y$. For example, the 3rd entry of $(2)^4$ is $(2)^4_3 = ([0 \ 0 \ 1 \ 0]^T)_3 = 1$. For $i = 0, 1, \ldots, 15$, define $w_i(y) = \sum_{j \in \{i\}_C} y_j$ for $y \in \{0, 1\}^n$. When $y$ is clear from the context, we write $w_i = w_i(y)$. For a vector $y \in \{0, 1\}^n$, we can rewrite $d_j(y)$ defined in (5) as

\begin{equation}
    d_j(y) = \sum_{i=0}^{15} \left[ |i|/2 + (-1)^{(i)_4} \right] (w_i - |i|/2).
\end{equation}

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

\begin{align*}
    d_1(y) &= w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7, \\
    d_2(y) &= w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7, \\
    d_3(y) &= w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7, \\
    d_4(y) &= w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7,
\end{align*}

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 $c_i$ for each $i \in \mathcal{P}$. Denote by $c'_i$ the $i$th codeword/row of $C'$, $i = 1, \ldots, 4$. For $y \in \{0, 1\}^n$, let $f_i(y)$ be the vector obtained by flipping the first $t$ bits of $y$. We see that $c'_i = c_i$ for $i \in \mathcal{O}$ and $c'_i = f_i(c_i)$ for $i \in \mathcal{P}$.

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

\begin{equation}
    d'_i(y) = d_i(y) + \sum_{\tau=1}^{t} (-1)^{(s_{\tau})} (\overline{y}_{\tau} - y_{\tau}),
\end{equation}

where $\overline{y}_{\tau} = y_{\tau} \oplus 1$. For a nonempty subset $\mathcal{S} \subset \{1, 2, 3, 4\}$, let

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

We have

\begin{align*}
    d_C(y) &= \min\{d_O(y), d_P(y)\}, \\
    d'_C(y) &= \min\{d'_O(y), d'_P(y)\}, \\
    d_{C'}(f_i(y)) &= \min\{d_C(f_i(y)), d'_P(f_i(y))\} \\
    &= \min\{d'_O(y), d_P(y)\}.
\end{align*}

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 satisfies 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 \( g_i : \mathcal{Y}_i \to \mathcal{Y}'_i \) such that one of the following conditions holds:

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

Such a pair of partitions exists. For example, when \( i_0 = 2^n \), \( \mathcal{Y}_i = \mathcal{Y}'_i = \{(i)\} \) 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 used to prove Theorem 1 and 4, there exists such a pair of partitions with \( i_0 = 5 \).

In the following discussion, we write \( \min\{a, b\} \) as \( a \wedge b \). For a function \( g : \{0, 1\}^n \to \mathbb{R} \), we write \( \{y \in \{0, 1\}^n : g(y) \geq 0\} \) 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 \( \langle s \rangle \), \( 0 \leq s \leq 15 \). Let \( C' \) be the code formed by changing the first column of \( C \) to \( \langle s' \rangle \). Following the notations in [11] \( \mathcal{O} \) 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 \( \langle (s) \rangle_j \neq \langle (s') \rangle_j \). Assume \( s' \neq s \) and \( s' \neq 15 - s \), and hence both \( \mathcal{O} \) and \( \mathcal{P} \) are nonempty.

In this case, \( d'_i \) defined in (11) becomes

\[
d'_i(y) = d_i(y) + (-1)^{(s)}_{j_1}(y_1 - y_{1_1}).
\]  \hspace{1cm} (15)

Consider an example with \( s = 1 \) and \( s' = 3 \). Now \( \mathcal{O} = \{1, 2, 4\} \) and \( \mathcal{P} = \{3\} \). Substituting (14) into (15),

\[
d'_1(y) = d_1(y) - y_1 + \overline{y_1},
\]
\[
d'_2(y) = d_2(y) - y_1 + \overline{y_1},
\]
\[
d'_3(y) = d_3(y) - y_1 + \overline{y_1},
\]
\[
d'_4(y) = d_4(y) + y_1 - \overline{y_1}.
\]

and hence

\[
d_{\mathcal{O}}(y) = d_1 \land d_2 \land d_4,
\]  \hspace{1cm} (16)
\[
d_{\mathcal{P}}(y) = d_3,
\]  \hspace{1cm} (17)
\[
d'_{\mathcal{O}}(y) = [(d_1 \land d_2) - y_1 + \overline{y_1}] \land (d_4 + y_1 - \overline{y_1}),
\]  \hspace{1cm} (18)
\[
d'_{\mathcal{P}}(y) = d_3 - y_1 + \overline{y_1}.
\]  \hspace{1cm} (19)
We are ready to form the partitions. Define the following $5$ subsets of $\{0,1\}^n$:

\[
\begin{align*}
\mathcal{Y}_1 &= \{d_\circ \leq d_P < d'_P\} \cup \{d_\circ \leq d'_P \leq d_P, d'_\circ \leq d'_P\}, \\
\mathcal{Y}_2 &= \{d_P \leq d'_P, d_P < d_\circ\} \cup \{d'_P < d_P \leq d_\circ, d_P \leq d'_\circ\}, \\
\mathcal{Y}_3 &= \{d'_P = d_\circ < d_P = d_\circ\}, \\
\mathcal{Y}_4 &= \{d_P = d'_P = d_\circ < d'_\circ\}, \\
\mathcal{Y}_5 &= \{d'_P = d_\circ < d'_\circ = d_P\}.
\end{align*}
\] (20)

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

\[
\mathcal{Y}'_i = \{f_1(y) : y \in \mathcal{Y}_i\},
\]

where function $f_1$ (defined in \llap{III}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 \llap{III}III.

**Lemma 10.** For the $(n, 2)$ codes $C$ and $C'$ formulated above, both $\{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_5\}$ 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 $y \in \mathcal{Y}_1$, $d_C(y) = d_{C'}(y) = d_\circ$;
2. For $y \in \mathcal{Y}_2$, $d_C(y) = d_{C'}(y') = d_P$ where $y' \overset{\Delta}{=} f_1(y) \in \mathcal{Y}'_2$;
3. For $y \in \mathcal{Y}_3$, $d_C(y) = d_P = d_{C'}(y) + 1 = d'_P + 1$;
4. For $y \in \mathcal{Y}_4$, $d_C(y) = d_\circ = d_{C'}(y') = d_P$ where $y' \overset{\Delta}{=} f_1(y) \in \mathcal{Y}'_4$;
5. For $y \in \mathcal{Y}_5$, $d_C(y) + 1 = d_\circ + 1 = d_{C'}(y') = d_P$ where $y' \overset{\Delta}{=} f_1(y) \in \mathcal{Y}'_5$.

*Proof.* See Appendix \[A\].

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

\[
\alpha^i_d(C) = |\{y \in \mathcal{Y}_i : d_C(y) = d\}|.
\] (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

\[
\alpha_d(C) = \sum_{i=1}^5 \alpha^i_d(C).
\]

By the definition of $\mathcal{Y}_3$ and $\mathcal{Y}_5$, $\alpha^3_0(C) = 0$ and $\alpha^5_0(C) = 0$.

**Theorem 11.** For two $(n, 2)$ codes $C$ and $C'$ with only one column different, $\lambda_{C'} \geq \lambda_C$ if and only if

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

*Proof.* See Appendix \[A\].

**Corollary 12.** For two $(n, 2)$ codes $C$ and $C'$ with only one column different, $\lambda_{C'} \geq \lambda_C$ if for $d = 1, \ldots, n$,

\[
\sum_{i=1}^d \alpha^3_i(C) \geq \sum_{i=0}^{d-1} \alpha^5_i(C).
\]
Proof. See Appendix A.

If we can compare $C$ and $C'$ based on Corollary 12, their relation is universal in the sense that it does not depend on $\varepsilon$.

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 $w(c_3 \oplus c_4)$ 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 $O = \{1, 2, 4\}$ and $P = \{3\}$, and hence

$$Y_5^c = \{d_3^c = d_{\{1,2,4\}} < d_3 = d'_{\{1,2,4\}}\}.$$  

Assume $Y_5$ is nonempty and fix $y \in Y_5$. As $d_3^c(y) = d_3(y) - y_1 + \overline{y_1}$, we have $y_1 = 1$. Further, due to

$$d_{\{1,2,4\}}(y) = d_1 \land d_2 \land d_4,$$

$$d'_{\{1,2,4\}}(y) = (d_1 - 1) \land (d_2 - 1) \land (d_4 + 1),$$

we have $d_{\{1,2,4\}} = d_4$ and hence $d_3 = d_4 + 1$. By (6),

$$d_3(y) + d_4(y) = \sum_{i} \left[ |i|/2 + (-1)^{(i)}s (w_i - |i|/2) \right] + \sum_{i} \left[ |i|/2 + (-1)^{(i)}s (w_i - |i|/2) \right]$$

$$= \sum_{i:(i)_{3} \neq (i)_{4}} |i| + 2 \sum_{i:(i)_{3} = (i)_{4}} \left[ |i|/2 + (-1)^{(i)}s \left( w_i - |i|/2 \right) \right]$$

$$= w(c_3 \oplus c_4) + 2 \sum_{i:(i)_{3} = (i)_{4}} \left[ |i|/2 + (-1)^{(i)}s \left( w_i - |i|/2 \right) \right].$$

As $w(c_3 \oplus c_4)$ is even, we see that $d_3 + d_4$ is even, which is a contradiction to $d_3 = d_4 + 1$. Therefore, $Y_5 = \emptyset$ and hence by Corollary 12 $\lambda_{C'} \geq \lambda_C$.

V. ANALYSIS OF CLASS-I CODES

Recall the definition of Class-I codes in Definition 1. 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 $Y_3$ and $Y_5$

Guided by Theorem 11 and Corollary 12, we first study $Y_3$ and $Y_5$ defined in (20) and (21).
Lemma 13. For a Class-I \((n, 2)\) code \(C\) with the first column (1) and \(C'\) obtained by replacing the first column of \(C\) to \((3)\),

\[
\mathcal{Y}_3 = \{y_1 = 1, d_4 \geq d_1 \land d_2 = d_3\},
\]

\[
\mathcal{Y}_5 = \{y_1 = 1, d_1 \land d_2 \geq d_4 + 2 = d_3 + 1\}.
\]

Proof. For code \(C\) and \(C'\) defined above, we have \((16) - (19)\). For \(y \in \mathcal{Y}_3\), \(d_3 - y_1 + \overline{y}_1 < d_4\) implies \(y_1 = 1\), and \(d_3 = d_1 \land d_2 \land d_4\) and \(d_3 - 1 = [(d_1 \land d_2) - 1] \land (d_4 + 1)\) together implies \(d_1 \land d_2 \leq d_4\) and \(d_3 = d_1 \land d_2\).

For \(y \in \mathcal{Y}_5\), \(d_3 - y_1 + \overline{y}_1 < d_3\) implies \(y_1 = 1\), and \(d_4 - 1 = d_1 \land d_2 \land d_4\) and \(d_3 = [(d_1 \land d_2) - 1] \land (d_4 + 1)\) together implies \(d_4 + 1 \leq (d_1 \land d_2) - 1\) and \(d_3 - 1 = d_4\). \(\Box\)

1) Characterization of \(\alpha_2^3\): For \(y \in \mathcal{Y}_3\), by Lemma \((10)\) \(d_C(y) = d_3\). By \((7) - (10)\) and Lemma \(13\) we have the following necessary and sufficient condition for \(y \in \mathcal{Y}_3\) with \(d_C(y) = i\): \(y_1 = 1\) and

\[
w_1 + \overline{w}_1 = i - w_5 - \overline{w}_6,
\]

\[
w_1 + \overline{w}_1 \leq w_5 + w_6 - w_5 - \overline{w}_6,
\]

\[
w_3 - \overline{w}_3 = w_5 + \overline{w}_6 - (w_5 + w_6) \land (\overline{w}_5 + \overline{w}_6).
\]

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 \(y\) satisfying \(y_1 = 1\) and

\[
w_5 + w_6 < (|5| + |6|)/2, \quad (23)
\]

\[
w_1 + w_5 = i - (|3| + |6|)/2, \quad (24)
\]

\[
w_1 + w_5 - w_6 \leq (|1| + |5| - |6|)/2, \quad (25)
\]

\[
w_3 + w_6 = (|3| + |6|)/2. \quad (26)
\]

We have

\[
|\mathcal{Y}_3^A(i)| = \sum_{w_1 > 1, w_5 > 0, w_3 > 0} \left(\binom{|1| - 1}{w_1 - 1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}\right).
\]

Define \(\mathcal{Y}_3^B(i)\) as the collection of \(y\) satisfying \(y_1 = 1\) and

\[
w_5 + w_6 \geq (|5| + |6|)/2, \quad (27)
\]

\[
w_1 + \overline{w}_6 = i - (|3| + |5|)/2, \quad (28)
\]

\[
w_1 + w_5 - w_6 \leq (|1| + |5| - |6|)/2, \quad (29)
\]

\[
w_3 - w_5 = (|3| - |5|)/2. \quad (30)
\]

We have

\[
|\mathcal{Y}_3^B(i)| = \sum_{w_1 > 1, w_3 > 0, w_5 > 0} \left(\binom{|1| - 1}{w_1 - 1}\binom{|3|}{w_3}\binom{|5|}{w_5}\binom{|6|}{w_6}\right). \quad (31)
\]

We see that \(\alpha_2^3 = |\mathcal{Y}_3^A(i)| + |\mathcal{Y}_3^B(i)|\).
2) Characterization of $\alpha_i^5$: For $y \in \mathcal{Y}_5$, by Lemma[10] $d_C(y) = d_3 - 1$. By (7) – (10) and Lemma[13] we have the following necessary and sufficient condition for $y \in \mathcal{Y}_5$ with $d_C(y) = i$: $y_1 = 1$ and
\[
w_1 + w_3 = i + 1 - w_5 - w_6,
\]
\[
w_1 - w_3 = w_5 + w_6 - w_5 - w_6 + 1,
\]
\[
w_3 - w_3 \geq w_5 + w_6 - (w_5 + w_6) \wedge (w_5 + w_6) + 1,
\]
which can be further simplified as $y_1 = 1$ and
\[
w_3 = (n + |3| - 1)/2 - i,
\]
\[
w_1 + w_5 - w_6 = (|1| + |5| - |6| + 1)/2,
\]
\[
w_3 + w_6 \geq (|3| + |6|)/2 + 1,
\]
\[
w_3 - w_5 \geq (|3| - |5|)/2 + 1.
\]

Hence
\[
\alpha_i^5 = \sum_{w_3 > 1, w_1, w_5, w_6: 32, 33, 34, 35} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|3|}{w_3} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right).
\]

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
\[
\sum_{i=0}^{d-1} \alpha_i^5 \sum_{w_3} \left( \frac{|3|}{w_3} \right) \left( \frac{|5|}{|5| - |3|} + w_3 \right) \left( \frac{|6|}{w_6} \right),
\]
where
\[
\mathcal{W}_5 = \left\{ \frac{|5| - |3|}{2} - d, \frac{|3| + |6|}{2} \right\} = \left\{ \frac{w_3 + w_6 \geq (|3| + |6|)}{2} + 1, \right\}
\]
\[
0 \leq w_3 \leq |3|, 0 \leq w_6 \leq |6|
\]

Similarly,
\[
\sum_{i=1}^{d} \alpha_i^3 \geq \sum_{i=1}^{d} \left| \mathcal{Y}_3^B(i) \right|
\]
\[
= \sum_{w_3} \left( \frac{|3|}{w_3} \right) \left( \frac{|5|}{|5| - |3|} + w_3 \right) \left( \frac{|6|}{w_6} \right)
\]
\[
= \sum_{w_3} \left( \frac{|3|}{2} + w_6' \right) \left( \frac{|5|}{2} + w_6' \right) \left( \frac{|6|}{2} + w_6' \right)
\]
where

\[
W_3 = \begin{cases} 
\frac{3|+6|}{2}, & w_6 \geq \frac{|3|+6|}{2} - d, \\
0 \leq w_5 \leq \frac{|3|}{2}, 0 \leq w_6 \leq \frac{6}{2},
\end{cases} \quad \begin{cases} 
\frac{|3|+6|}{2}, & w_3 - w_6 \geq \frac{|3|+6|}{2} - d, \\
0 \leq w_3 \leq \frac{|3|}{2}, 0 \leq w_6 \leq \frac{6}{2},
\end{cases}
\]

and (38) is obtained by change of variables \(w'_5 = \frac{|3|}{2} - w_6 - \frac{6}{2}\) and \(w'_6 = \frac{6}{2} - w_3 - \frac{|3|}{2}\).

We show that \(W_5 \subset W'_3\). Due to \(|3| \leq |6|\), we have \(\frac{|3|+6|}{2} \leq |6| \leq |3|\). For \((w_3, w_6) \in W_5\), 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 \((w_3, w_6) \in W'_3\).

By Lemma 14 in Appendix B for \((w_3, w_6) \in W_5\), we have

\[
\left( \begin{array}{c} |3| \\ w_3 \\ |6| \\ w_6 \end{array} \right) \leq \left( \begin{array}{c} |3| \\ w_3 + w_6 \\ |6| \\ w_3 \end{array} \right) = \left( \begin{array}{c} |3| \\ w_3 \\ |6| \\ w_6 \end{array} \right)
\]

Comparing (36) and (38), we obtain \(\sum_{i=1}^{d} \alpha^5_i \geq \sum_{i=0}^{d-1} \alpha^5_i\) for any \(d = 1, \ldots, n\). By Corollary 12 \(\lambda_{C'} \geq \lambda_C\), 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), \(\alpha^5_i = 0\) if \(i \neq \frac{n-1}{2}\). So when \(d < \frac{n+1}{2}\), \(\sum_{i=0}^{d-1} \alpha^5_i = 0\) and hence \(\sum_{i=1}^{d} \alpha^5_i \geq \sum_{i=0}^{d-1} \alpha^5_i\); when \(d \geq \frac{n+1}{2}\),

\[
\sum_{i=0}^{d-1} \alpha^5_i = \alpha^5_{\frac{n+1}{2}} = \sum_{w_5 = \frac{|3|+6|}{2}}^{\frac{|3|+6|}{2} - \frac{|3|}{2} - 1} \frac{\left( |1| - 1 \right) \left( |5| \right) \left( |6| \right)}{w_5} \left( |3| \right) \left( |6| \right)
\]

where (33) is \(w_6 - w_1 \leq \frac{|6| - |1| - |3|}{2}\). Substituting \(w'_1 = |1| - w_1 + 1\) into (39), we obtain

\[
\alpha^5_{\frac{n+1}{2}} \leq \sum_{1 \leq w'_1 \leq |1|, |w_6| \geq |5| + 1, \quad w'_1 \leq \frac{|3|+6|}{2} - w_6} \frac{\left( |1| - 1 \right) \left( |5| \right) \left( |6| \right)}{w_6}
\]
When $d \geq \frac{n+1}{2}$, we further have

$$\sum_{i=1}^{d} \alpha_i^3 \geq \sum_{i=1}^{\frac{n+1}{2}} |\mathcal{Y}_{d'}(i)|$$

$$= \sum_{\begin{subarray}{c} w_1 \geq 1, w_6 \geq 6 \end{subarray}} \sum_{\begin{subarray}{c} w_5 = w_6 - w_1 + \frac{|6| - |5| - |6| - 1}{2} \end{subarray}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_6} \right) \left( \frac{|6|}{w_6} \right)$$

$$= \sum_{\begin{subarray}{c} w_1 \geq 1, w_6 \geq 6 \end{subarray}} \sum_{\begin{subarray}{c} w_5 = w_6 - w_1 + \frac{|6| - |5| - |6| - 1}{2} \end{subarray}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_6} \right) \left( \frac{|6|}{w_6} \right),$$

where (27) + (30) is $w_6 \geq \frac{|6|}{2}$ and (29) + (30) is $w_1 - w_6 \leq \frac{|1| - |6| - 1}{2}$, 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 $\sum_{i=1}^{d} \alpha_i^3 \geq \sum_{i=0}^{d-1} \alpha_i^3$ for $d \geq \frac{n+1}{2}$. By Corollary 12, $\lambda_{C'} \geq \lambda_C$, 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

$$\alpha_{\frac{n}{2} - 1}^5 = \sum_{\begin{subarray}{c} w_1 \geq 1, w_6 \geq 6 \end{subarray}} \sum_{\begin{subarray}{c} w_5 = w_6 - w_1 + \frac{|6| - |5| - |6| - 1}{2} \end{subarray}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)$$

$$\leq \sum_{\begin{subarray}{c} w_1 \geq 1, w_6 \geq 6 \end{subarray}} \sum_{\begin{subarray}{c} w_5 = w_6 - w_1 + \frac{|6| - |5| - |6| - 1}{2} \end{subarray}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_6} \right) \left( \frac{|6|}{w_6} \right)$$

$$= \sum_{\begin{subarray}{c} w_1 \geq 1, w_6 \geq 6 \end{subarray}} \sum_{\begin{subarray}{c} w_5 = w_6 - w_1 + \frac{|6| - |5| - |6| - 1}{2} \end{subarray}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_6} \right) \left( \frac{|6|}{w_6} \right),$$

$$= \sum_{\begin{subarray}{c} w_1 \geq 1, w_6 \geq 6 \end{subarray}} \sum_{\begin{subarray}{c} w_5 = w_6 - w_1 + \frac{|6| - |5| - |6| - 1}{2} \end{subarray}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_6} \right) \left( \frac{|6|}{w_6} \right),$$

where (32) + (33) is $w_6 - w_1 \leq \frac{|6| - |1| - 2}{2}$, 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

\[
\alpha_{\frac{n}{2}}^5 = \sum_{\begin{array}{l}
w_1 \geq 1, w_6 \geq \frac{|6|+3}{2} \\
|\{i\} - \frac{|6|}{2} - 2 \leq w_6 \leq \frac{|6|+1}{2} - \left|\frac{|6|+3}{2}\right| - 3
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)
\]

\[
\leq \sum_{\begin{array}{l}
w_1 \geq 1, w_6 \geq \frac{|6|+3}{2} \\
|\{i\} - \frac{|6|}{2} - 2 \leq w_6 \leq \frac{|6|+1}{2} - \left|\frac{|6|+3}{2}\right| - 3
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)
\]

\[
= \sum_{\begin{array}{l}
w_1 \geq 1, w_6 \geq \frac{|6|+3}{2} \\
|\{i\} - \frac{|6|}{2} - 2 \leq w_6 \leq \frac{|6|+1}{2} - \left|\frac{|6|+3}{2}\right| - 3
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)
\]

(44)

\[
= \sum_{\begin{array}{l}
w_1 \geq 1, w_6 \geq \frac{|6|+3}{2} \\
|\{i\} - \frac{|6|}{2} - 2 \leq w_6 \leq \frac{|6|+1}{2} - \left|\frac{|6|+3}{2}\right| - 3
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)
\]

(45)

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

Following (31), we have

\[
\left| \left\{ w_3 = 1 \right\} \cap \left( \bigcup_{\frac{n}{2} \leq i \leq 5} \mathcal{A}_{\alpha_{\frac{n}{2}}}^B (i) \right) \right|
\]

\[
= \sum_{\begin{array}{l}
w_1 \geq 1 \\
\frac{29}{5} + \frac{30}{5}, w_6 \leq \frac{|6|+1}{2} - \left|\frac{|6|+3}{2}\right| - 3
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)
\]

(46)

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

\[
\alpha_{\frac{n}{2} - 1}^5 \leq \left| \left\{ w_3 = 1 \right\} \cap \left( \bigcup_{\frac{n}{2} \leq i \leq 5} \mathcal{A}_{\alpha_{\frac{n}{2}}}^B (i) \right) \right|.
\]

(47)

Following (31), we have

\[
\left| \left\{ w_3 = 0 \right\} \cap \left( \bigcup_{\frac{n}{2} \leq i \leq 1} \mathcal{A}_{\alpha_{\frac{n}{2}}}^B (i) \right) \right|
\]

\[
= \sum_{\begin{array}{l}
w_1 \geq 1 \\
\frac{29}{5} + \frac{30}{5}, w_6 \leq -w_6 + \frac{6+1}{2} - \left|\frac{6+3}{2}\right|
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right)
\]

(48)

\[
= \sum_{\begin{array}{l}
w_1 \geq 1, w_6 \geq \frac{|6|+1}{2} \\
w_6 \leq w_6 + \frac{|6|+1}{2} - \left|\frac{6+3}{2}\right|
\end{array}} \left( \frac{|1| - 1}{w_1 - 1} \right) \left( \frac{|5|}{w_5} \right) \left( \frac{|6|}{w_6} \right).
\]
where (27) + (30) implies \( w_6 \geq \frac{|6|+1}{2} \), (29) + (30) implies \( w_1 - w_6 \leq \frac{|1|-|6|+1}{2} \), and (48) follows that \( \frac{|1|-|6|+1}{2} \leq \frac{n}{2} + 1 - \frac{1+|5|}{2} - |6| \). Since \( w_6 + \frac{|1|-|6|+1}{2} \geq -w_6 + \frac{|6|+|1|-2}{2} \) when \( w_6 \geq \frac{|6|+1}{2} \), comparing the RHS’ of (45) and (48), we get

\[
\alpha_{\frac{n}{2}}^5 \leq \{ w_3 = 0 \} \cap \left( \bigcup_{i \leq \frac{n}{2}+1} \mathcal{Y}_3^B(i) \right).
\]

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

\[
\sum_{i=0}^{d-1} \alpha_i^5 = \alpha_{\frac{n}{2}}^5 - 1 \leq \{ w_3 = 1 \} \cap \left( \bigcup_{i \leq \frac{n}{2}} \mathcal{Y}_3^B(i) \right) = \sum_{i=1}^{\frac{n}{2}} \alpha_i^3.
\]

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

\[
\sum_{i=0}^{d-1} \alpha_i^5 = \alpha_{\frac{n}{2}}^5 + 1 \leq \{ w_3 = 1 \} \cap \left( \bigcup_{i \leq \frac{n}{2}} \mathcal{Y}_3^B(i) \right) + \{ w_3 = 0 \} \cap \left( \bigcup_{i \leq \frac{n}{2}+1} \mathcal{Y}_3^B(i) \right) = \sum_{i=1}^{d} \alpha_i^3.
\]

Thus we have \( \sum_{i=1}^{d} \alpha_i^3 \geq \sum_{i=0}^{d-1} \alpha_i^5 \) for \( 1 \leq d \leq n \). By Corollary 12 \( \lambda_C \geq \lambda_{C'} \), proving the case when \( |3| = 1 \).

D. Algorithm for Verifying Optimal Codes

Based on Theorem 8 we give an algorithm for verifying whether linear \((n, 2)\) codes are optimal for fixed \( n \) (see the pseudocode in Algorithm 1). The algorithm checks each Class-I \((n, 2)\) code \( C \) specified by \((w_1, w_3, w_5, w_6)\) with the first column being \( \langle 1 \rangle \) and \(|3|_C \leq |5|_C \leq |6|_C \), and compares \( C \) with \( C' \) obtained by replacing the first column of \( C \) to \( \langle 3 \rangle \). Other \((n, 2)\) Class-I codes can be converted to ones of the above type by flipping columns and interchanging rows, and hence do not need to be checked again. By Theorem 8 if for each code \( C \) checked by the algorithm we have \( \sum_{i=1}^{d} \alpha_i^3(C) \geq \sum_{i=0}^{d-1} \alpha_i^5(C) \) for \( d = 1, \ldots, n \), which implies \( \lambda_C \leq \lambda_{C'} \) by Corollary 12 then linear \((n, 2)\) codes are optimal.

Evaluating Algorithm (V-D) we have verified that for \( n \leq 300 \), linear codes are optimal. To total number of types of Class-I codes to evaluate is \( O(n^4) \). For each type, there are less than \( 2n \) values \( \alpha_i^3/\alpha_i^5 \) to evaluate, each of which has complexity \( O(n^2) \). Therefore, the complexity of the algorithm is \( O(n^7) \).

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 \( \langle 1 \rangle \) and \( \langle 7 \rangle \). Let \( C' \) be the code obtained by flipping the first two bits of \( c_3 \), so that the first two columns of \( C' \) are \( \langle 3 \rangle \) and \( \langle 5 \rangle \). (Other cases of Theorem 4 can be converted to this case by interchanging rows.)

Following the notations in III \( O = \{ 1, 2, 4 \} \), \( P = \{ 3 \} \), and

\[
d'_i(y) = d_i(y) + (-1)^{(1)}(y_1 - y_1) + (-1)^{(7)}(y_2 - y_2).
\]
Algorithm 1: Check optimality of linear \((n,2)\) codes

**input**: \(n\)

**output**: \(a\) (If \(a = -1\), linear \((n,2)\) codes are optimal.)

Initialize \(a = -1\):

for \(n_1 = 3, 5, 7, \ldots \) and \(n_1 \leq n\) do

for \(n_3 = 2, 3, \ldots, \left\lfloor \frac{n-n_1}{3} \right\rfloor\) do

for \(n_5 = n_3, n_3 + 2, \ldots \) and \(n_5 \leq \left\lfloor \frac{n-n_1-n_3}{5} \right\rfloor\) do

for \(n_6 = n_5, n_5 + 2, \ldots \) and \(n_6 \leq n - n_1 - n_3 - n_5\) do

Compute \(\alpha_i^3\) and \(\alpha_i^5\), \(i = 0, \ldots, n\) for code \(C\) with \(|1|_C = n_1, |3|_C = n_3, |5|_C = n_5\).

\(|6|_C = n_6;\)

if \(\sum_{i=1}^d \alpha_i^3 < \sum_{i=0}^{d-1} \alpha_i^5\) for some \(d \in \{1, \ldots, n\}\) then

\(a = 1\); break;

end

end

end

end

When \(y_1 = y_2\), we have

\[d'_P(y) = d_P(y). \tag{50}\]

When \(y_1 \neq y_2\), we have \(d'_i(y) = d_i(y), d'_4(y) = d_4(y)\), and

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

and hence

\[(d'_C(y) - d_C(y))(d'_P(y) - d_P(y)) = (d'_{1,2,4}(y) - d_{1,2,4}(y))(d'_3(y) - d_3(y)) \geq 0. \tag{52}\]

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

\[\mathcal{Y}_1 = \{y_1 = y_2\},\]

\[\mathcal{Y}_2 = \{y_1 \neq y_2, d_C \leq d_P \land d'_P\},\]

\[\mathcal{Y}_3 = \{y_1 \neq y_2, d_C > d_P \land d'_P, d_P \leq d_C \land d'_C\},\]

\[\mathcal{Y}_4 = \{y_1 \neq y_2, d'_P < d_C \land d'_C < d_P\},\]

\[\mathcal{Y}_5 = \{y_1 \neq y_2, d'_C \leq d_P \land d'_P < d_C, d'_C < d_P\}.
\]

Recall the function \(f_2\) defined in \([3]\) that flips the first two bits of a binary vector. For \(i = 3, 4\), let

\[\mathcal{Y}'_i = \{f_2(y) : y \in \mathcal{Y}_i\}.\]
We justify that \( Y_1, Y_2, Y_3, Y_4, Y_5 \) form a partition of \( \{0, 1\}^n \) and \( Y_1, Y_2, Y_3', Y_4', Y_5' \) form a partition of \( \{0, 1\}^n \):

First, we show that
\[
Y_4 \cup Y_5 = \{ y_1 \neq y_2, d_O > d_P \land d'_P, d_P > d_O \land d'_O \},
\]
and then we obtain \( \bigcup_{i=1}^5 Y_i = \{0, 1\}^n \). Moreover, \( Y_1, \ldots, Y_5 \) are all disjoint by checking the definition. Thus \( Y_1, \ldots, Y_5 \) form a partition of \( \{0, 1\}^n \).

To show (53), since \( Y_4 \subseteq \{d_O > d_P \land d'_P\} \), we have
\[
Y_4 = \{ y_1 \neq y_2, d'_P < d_O \land d'_O < d_P \} \cap \{d_O > d_P \land d'_P\}
= \{ y_1 \neq y_2, d_O > d_P \land d'_P, d_P > d_O \land d'_O \} \cap
\{d_O \land d'_O > d'_P\}.
\]
Denote
\[
\mathcal{A}_1 = \{ y_1 \neq y_2, d_O > d_P \land d'_P, d_P > d_O \land d'_O \} \cap
\{d_O \land d'_O \leq d'_P\}.
\]
For \( y \in \mathcal{A}_1 \), we have \( d_O(y) > d'_O(y) \) which implies \( d_P(y) > d'_P(y) \) by (51) and (52), and hence
\[
d'_O(y) \leq d_P(y) \land d'_P(y) < d_O(y), \quad d'_O(y) < d_P(y).
\]
Thus we have \( y \in Y_5 \) and then \( \mathcal{A}_1 \subseteq Y_5 \). For \( y \in Y_5 \), we have \( d_O(y) > d'_O(y) \) by the definition above, which implies \( d_P(y) > d'_P(y) \) by (51) and (52). Then we obtain
\[
d_P(y) > d_O(y) \land d'_O(y) = d'_O(y) \leq d'_P(y) < d_O(y).
\]
Thus \( y \in A_1 \) and then \( Y_3 \subseteq A_1 \). Therefore, \( Y_3 = A_1 \). From (54) and (55), we obtain (53).

We further show that
\[
Y_3' \cup Y_4' \subseteq Y_3 \cup Y_4.
\]
Since \( f_2 \) is an one-to-one mapping, we get \( Y_3' \cup Y_4' = Y_3 \cup Y_4 \). Therefore, \( Y_1, Y_2, Y_3', Y_4', Y_5' \) form a partition of \( \{0, 1\}^n \).

To show (56), we see
\[
Y_3' = \{ y_1 \neq y_2, d_P < d_O \land d'_O < d'_P \} \subseteq Y_3,
\]
Following the similar argument as in YP, the above claims are justified as follows:

\[
\mathcal{Y}_3' \setminus \mathcal{Y}_4 = \{ y_1 \neq y_2, d'_\mathcal{O} > d_P \land d'_P, d'_P \leq d_\mathcal{O} \land d'_\mathcal{O} \} \cap \\
(\{ d_\mathcal{O} \land d'_\mathcal{O} \geq d_P \} \cup \{ d'_P \geq d_\mathcal{O} \land d'_\mathcal{O} \}) \\
= \{ y_1 \neq y_2, d'_\mathcal{O} > d_P \land d'_P, d'_P \leq d_\mathcal{O} \land d'_\mathcal{O}, \\
d_\mathcal{O} \land d'_\mathcal{O} \geq d_P \} \cup \{ y_1 \neq y_2, d'_\mathcal{O} > d_P \land d'_P, \\
d'_P \leq d_\mathcal{O} \land d'_\mathcal{O}, d'_P \geq d_\mathcal{O} \land d'_\mathcal{O} \} \\
= \{ y_1 \neq y_2, d_P \land d'_P < \max(d_P, d'_P) \leq d_\mathcal{O} \land d'_\mathcal{O} \} \cup \\
\{ y_1 \neq y_2, d_\mathcal{O} \land d'_\mathcal{O} = d'_P, d_P \land d'_P < d'_\mathcal{O} \} \tag{58}
\]

where in the last equality \( d_P \land d'_P < \max(d_P, d'_P) \) follows from \((51)\). By \((52)\), when \( y_1 \neq y_2 \), if \( d_\mathcal{O} < d'_\mathcal{O} \), then \( d_P < d'_P \); and if \( d_P > d'_P \), 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 \( y \in \mathcal{Y}_1 \), \( d_C(y) = d_C'(y) \);
2) For \( y \in \mathcal{Y}_2 \), \( d_C(y) = d_C'(y) = d_\mathcal{O} \);
3) For \( y \in \mathcal{Y}_3 \), \( d_C(y) = d_C'(f_2(y)) = d_P \);
4) For \( y \in \mathcal{Y}_4 \), \( d_C(y) = d_\mathcal{O} \land d_P \geq d_C'(f_2(y)) = d'_\mathcal{O} \);
5) For \( y \in \mathcal{Y}_5 \), \( d_C(y) = d_\mathcal{O} \land d_P \geq d_C'(y) = d'_P \).

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 \( y \in \mathcal{Y}_1 \), as \( y_1 = y_2 \), we have \( d'_P = d_P \) by \((50)\), and hence

\[
d_C(y) = d_\mathcal{O}(y) \land d_P(y) = d_\mathcal{O}(y) \land d'_P(y) = d_C'(y).
\]

2) For \( y \in \mathcal{Y}_2 \), by the definition of \( \mathcal{Y}_2 \), we have \( d_\mathcal{O} \leq d_P \land d'_P \), and hence \( d_C(y) = d_C'(y) = d_\mathcal{O} \).

3) For \( y \in \mathcal{Y}_3 \), we have \( d_P \leq d_\mathcal{O} \land d'_\mathcal{O} \) by the definition of \( \mathcal{Y}_3 \). We then have

\[
d_C(y) = d_\mathcal{O}(y) \land d_P(y) = d_P(y), \\
d_C'(f_2(y)) = d'_\mathcal{O}(y) \land d_P(y) = d_P(y).
\]

4) For \( y \in \mathcal{Y}_4 \), we have \( y_1 \neq y_2, d'_P < d_\mathcal{O} \land d'_\mathcal{O} < d_P \) by the definition of \( \mathcal{Y}_4 \). By \((52)\), \( d_\mathcal{O} \land d'_\mathcal{O} = d'_\mathcal{O} \), which implies \( d_C'(f_2(y)) = d'_\mathcal{O}(y) \land d_P(y) = d'_\mathcal{O}(y) \) and hence

\[
d_C(y) = d_\mathcal{O}(y) \land d_P(y) \geq d'_\mathcal{O}(y) = d_C'(f_2(y)).
\]

5) For \( y \in \mathcal{Y}_5 \), we have \( y_1 \neq y_2, d'_\mathcal{O} \leq d_P \land d'_P < d_\mathcal{O} \). By \((51)\) and \((52)\), \( d'_P < d_P \). Then we have

\[
d_C(y) = d_\mathcal{O} \land d_P \geq d_\mathcal{O} \land d'_P = d_C'(y)
\]

and

\[
d_C'(y) = d'_P(y).
\]
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 \(Y_1, \ldots, Y_5\) are all disjoint. To show they form a partition, we can verify that

\[
Y_1 \cup Y_4 \cup Y_5 = \{d_O \leq d_P \land d'_P\},
\]

\[
Y_2 \cup Y_3 = \{d_O > d_P \land d'_P\}
\]

and hence \((Y_1 \cup Y_4 \cup Y_5) \cup (Y_2 \cup Y_3) = \{0, 1\}^n\).

We first prove that \(Y_1 \cup Y_4 \cup Y_5 = \{d_O \leq d_P \land d'_P\}\). Notice that the three sets can be rewritten as

\[
Y_1 = \{d_O \leq d_P < d'_P\} \cup \{d_O \leq d'_P \leq d_P, d'_O \leq d'_P\}
\]

\[
= (\{d_O \leq d_P \land d'_P\} \cap \{d_P < d'_P\}) \cup (\{d_O \leq d_P \land d'_P\}
\]

\[
\cap \{d'_P \leq d_P, d'_O \leq d'_P\})
\]

(59)

\[
Y_4 = \{d_P = d'_P = d_O < d'_O\}
\]

(a)

\[
= \{d_O \leq d_P \land d'_P\} \cap \{d'_P \leq d_P, d'_O > d'_P, d_P = d'_P\}
\]

(60)

\[
Y_5 = \{d'_P = d_O < d'_O = d_P\}
\]

(b)

\[
= \{d_O \leq d_P \land d'_P\} \cap \{d'_P \leq d_P, d'_O > d'_P, d_P > d'_P\}
\]

(61)

For \(\forall y\), we have

\[
|d_S(y) - d'_S(y)| \leq 1,
\]

(62)

which can be obtained by the definition. Then if \(d_O \leq d'_P < d'_O\), we will have \(d_O = d'_P\) and thus the equality (a) in (60) holds. Furthermore, if \(d'_O > d'_P\) we have \(d_O \geq d'_P\), and if \(d'_O > d'_P\), we have \(d_O \geq d'_P\) and thus the equality (b) in (61) holds. By (60) and (61), we have

\[
Y_4 \cup Y_5 = \{d_O \leq d_P \land d'_P\} \cap \{d'_P \leq d_P, d'_O > d'_P\}.
\]

From (59), this further implies

\[
Y_1 \cup Y_4 \cup Y_5 = \{d_O \leq d_P \land d'_P\}.
\]
We now prove \( \mathcal{Y}_2 \cup \mathcal{Y}_3 = \{d_\mathcal{O} > d_P \land d'_P\} \). First we rewrite the two sets as

\[
\mathcal{Y}_2 = \{d_P \leq d'_P, d_P < d_\mathcal{O}\} \cup \{d'_P < d_P \leq d_\mathcal{O}, d_P \leq d'_\mathcal{O}\} \\
= (\{d_\mathcal{O} > d_P \land d'_P\} \cap \{d_P \leq d'_P\}) \cup (\{d_\mathcal{O} > d_P \land d'_P\} \\
\cap \{d_P > d'_P, d_P \leq d_\mathcal{O}, d_P \leq d'_\mathcal{O}\} \\
= (\{d_\mathcal{O} > d_P \land d'_P\} \cap \{d_P \leq d'_P\}) \cup (\{d_\mathcal{O} > d_P \land d'_P\} \\
\cap \{d_P > d'_P, d_P \leq d_\mathcal{O}\}).
\]

(63)

\[
\mathcal{Y}_3 = \{d'_P = d'_{\mathcal{O}} < d_P = d_\mathcal{O}\} \\
= \{d_\mathcal{O} > d_P \land d'_P\} \cap \{d_P > d'_P, d_P > d'_\mathcal{O}\}.
\]

(64)

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

By definition,

\[
\mathcal{Y}_2' = \{d'_P \leq d_P, d'_P < d'_{\mathcal{O}}\} \cup \{d_P < d'_P \leq d'_{\mathcal{O}}, d'_P \leq d_\mathcal{O}\}, \\
\mathcal{Y}_4 = \{d_P = d'_P = d'_{\mathcal{O}} \land d_P < d_\mathcal{O}\}, \\
\mathcal{Y}_5 = \{d_P = d'_{\mathcal{O}} \land d_P = d'_P\}.
\]

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).

\[
d_{C}(y) = \min \{d_\mathcal{O}(y), d_P(y)\} \\
d_{C'}(y) = \min \{d_\mathcal{O}(y), d'_P(y)\} \\
d_{C'}(f_1(y)) = \min \{d'_{\mathcal{O}}(y), d_P(y)\}.
\]

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

\[
d_{C'}(y') = d_\mathcal{O}(y') \land d'_P(y') \\
= d'_\mathcal{O}(y) \land d_P(y) = d_P(y).
\]

To prove claim 3), for \( y \in \mathcal{Y}_3 \), \( d_{C}(y) = d_\mathcal{O}(y) \land d_P(y) = d_P(y) \) by the definition of \( \mathcal{Y}_3 \). Moreover,

\[
d_{C'}(y) = d_\mathcal{O}(y) \land d'_P(y) = d'_P(y) < d_{C}(y).
\]

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

\[
d_{C}(y) = d_\mathcal{O}(y) \land d_P(y) = d_P(y), \\
d_{C'}(y') = d'_{\mathcal{O}}(y) \land d'_P(y) = d_P(y).
\]
To prove claim 5), for \( y \in \mathcal{Y}_5 \), \( d_C(y) = d_C(y) \land d_P(y) = d_C(y) \) by the definition of \( \mathcal{Y}_5 \). Moreover,

\[
d_C(y') = d_C(y) \land d_P(y) = d_C(y) > d_C(y).
\]

By \( 62 \), we have \( d'_C = d_C + 1 \).

**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 \( \alpha_d(C') = \sum_{i=1}^5 \alpha'_d(C') \) where

\[
\alpha'_d(C') = \left| \{ y \in \mathcal{Y}_i : d_{C'}(y) = \alpha \} \right| \alpha_d(C),
\]

\[
\alpha'_d(C') = \left| \{ y \in \mathcal{Y}_2 : d_{C'}(y) = \alpha \} \right| \alpha_d(C),
\]

\[
\alpha'_d(C') = \left| \{ y \in \mathcal{Y}_3 : d_{C'}(y) = \alpha \} \right| \alpha_d(C),
\]

\[
\alpha'_d(C') = \left| \{ y \in \mathcal{Y}_4 : d_{C'}(y) = \alpha \} \right| \alpha_d(C),
\]

\[
\alpha'_d(C') = \left| \{ y \in \mathcal{Y}_5 : d_{C'}(y) = \alpha \} \right| \alpha_d(C).
\]

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

\[
\lambda_{C'} - \lambda_C = \frac{1}{|C'|} \sum_{d=0}^n (\alpha_d(C') - \alpha_d(C))(1 - \epsilon)^{n-d} \epsilon^d
\]

\[
= \frac{1}{|C'|} \sum_{d=0}^n \sum_{i=1}^5 (\alpha'_d(C') - \alpha'_d(C))(1 - \epsilon)^{n-d} \epsilon^d
\]

\[
= \frac{1}{|C'|} \sum_{d=0}^n \sum_{i=3,5} (\alpha'_d(C') - \alpha'_d(C))(1 - \epsilon)^{n-d} \epsilon^d,
\]

By substituting \( \alpha'_d(C') = \alpha'_d+1(C) \) and \( \alpha'_d(C') = \alpha'_d-1(C) \), we see that \( \lambda_{C'} \geq \lambda_C \) if and only if

\[
\sum_{d=0}^n [\alpha'_d+1(C) - \alpha'_d(C) + \alpha'_d-1(C) - \alpha'_d(C)] \left( \frac{\epsilon}{1-\epsilon} \right)^d \geq 0,
\]

where the LHS can be further simplified as

\[
\sum_{d=1}^n [\alpha'_d(C) - \alpha'_d-1(C)] \left( \frac{\epsilon}{1-\epsilon} \right)^{d-1} \left( 1 - \frac{\epsilon}{1-\epsilon} \right).
\]

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

**Proof of Corollary 22** Let \( \epsilon_0 = \frac{\epsilon}{1-\epsilon} \) and let \( \Psi_d = \sum_{i=1}^d [\alpha'_d(C) - \alpha'_d-1(C)] \) for \( d = 1, \ldots, n \) and \( \Psi_0 = 0 \). Write

\[
\sum_{d=1}^n [\alpha'_d(C) - \alpha'_d-1(C)] \left( \frac{\epsilon}{1-\epsilon} \right)^{d-1}
\]

\[
= \sum_{d=1}^n (\Psi_d - \Psi_{d-1}) \epsilon_0^{d-1}
\]

\[
= \Psi_n \epsilon_0^{n-1} + \sum_{d=1}^{n-1} \Psi_d (\epsilon_0^{d-1} - \epsilon_0^d).
\]

Note that for \( 0 < \epsilon < \frac{1}{4} \), \( \epsilon_0^d = \left( \frac{\epsilon}{1-\epsilon} \right)^d \) is a strictly decreasing function of \( d \). By Theorem 11 we can prove the sufficient conditions of the corollary.
Appendix B

A Lemma

A similar result has been proved in [4]. Here we provide a proof for completeness.

Lemma 14. Suppose $|3| \leq |6|$ of the same parity. For $(w_3, w_6) \in W_5$ (defined in (37)),

$$
\left( \frac{|3|}{w_3} \right) \left( \frac{|6|}{w_6} \right) \leq \left( \frac{|3|}{|3|-|6|} + \tilde{w}_6 \right) \left( \frac{|6|}{|6|-|3|} + \tilde{w}_3 \right).
$$

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

$$
\left( \frac{|3|}{\frac{|3|}{2} + \tilde{w}_3} \right) \left( \frac{|6|}{\frac{|6|}{2} + \tilde{w}_6} \right) \leq \left( \frac{|3|}{\frac{|3|}{2} + \tilde{w}_3} \right) \left( \frac{|6|}{\frac{|6|}{2} + \tilde{w}_6} \right).
$$

We have by the definition of $W_5$ in (37),

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

We write

$$
\left( \frac{|3|}{\frac{|3|}{2} + \tilde{w}_3} \right) \left( \frac{|6|}{\frac{|6|}{2} + \tilde{w}_6} \right) = \frac{|3| \cdots (|3| - |3| + 1) \cdots (|6| - |6| + 1)}{(|3| + \tilde{w}_3) \cdots (|3| + \tilde{w}_3 + 1) \cdots (|6| + \tilde{w}_6 + 1)}
$$

$$
= \frac{|3| - \tilde{w}_3 \cdots (|3| - \tilde{w}_3 + 1)}{(|3| + \tilde{w}_3) \cdots (|3| + \tilde{w}_3 + 1)} \cdot \frac{|6| + \tilde{w}_6 + 1}{(|6| + \tilde{w}_6) \cdots (|6| + \tilde{w}_6 + 1)}
$$

$$
= \prod_{i=1}^{\tilde{w}_3 - \tilde{w}_6} \frac{|3| - i}{|3| + \tilde{w}_3 + \frac{i}{2}} \cdot \frac{|6| + i}{|6| + \frac{i}{2}}
$$

$$
\leq 1,
$$

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

$$
\left( \frac{|3|}{2} + \frac{i}{2} \right) \tilde{w}_6 - \tilde{w}_3 \left( \frac{|6|}{2} + \frac{i}{2} \right) = - \left( \frac{|3|}{2} + \frac{i}{2} \right) \tilde{w}_3 + \tilde{w}_6 \left( \frac{|6|}{2} + \frac{i}{2} \right)
$$

$$
= (\tilde{w}_3 + \tilde{w}_6) \left( \frac{|3|}{2} - \frac{|6|}{2} \right) \leq 0
$$

where the inequality follows from $\tilde{w}_3 + \tilde{w}_6 \geq 1$ 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] 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.
[5] 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.