Weight Distributions for Successive Cancellation Decoding of Polar Codes

Mars Davletshin, Nikita Polyanskii, and Rina Revenko

***Abstract*— In this paper, we derive the exact weight distributions for the successive cancellation decoding of polar codes. The results allow to get an estimate of the decoding error probability. Furthermore, we prove a statement on the minimal distance between cosets for the successive cancellation list decoding.**

Keywords: Polar codes, weight distribution, closest coset decoding, multistage decoding, partial order.

I. Introduction

Polar codes, introduced by Arikan [1], provably achieve the symmetric capacity of any binary-input memoryless symmetric channels (B-MSC) with encoding and decoding complexity $\Theta(N \log_2 N)$, where N is the block length of the code.

Multilevel codes are based on partitioning, thus multistage decoding is the most natural one to be performed. [2]. Polar codes with the successive cancellation (SC) decoding can be represented in this way [1], [3]. A typical multilevel code construction employs small codes to get a larger one. Whereas polar codes are obtained by taking a Kronecker power of a square kernel matrix and expurgating some rows using a specific criterion. It is known that polar codes with good distance properties turn out to have a poor performance under the SC decoding. To evaluate the error rate provided by a multistage decoder, it is essential to calculate the weight distribution (WD) between cosets (or spectrum of component codes) at all the stages [4]. However, only the minimal distance for the SC decoding of polar codes is known at present. The aim of our paper is to calculate WD at all the stages of the SC decoding.

A. Outline

The rest of the paper is organized as follows. In Section II, we give key definitions and notations of polar codes and WDs associated with the SC decoding. We derive WD and focus our attention on its first nonzero component in Section III. To obtain an algorithm calculating them, we exploit a similar idea as in [5], where an $|u|u+v|$ construction is investigated. Also, we find a natural connection between the first nonzero component of WDs and the partial order [6], [7]. The minimal distance between cosets for the SC list decoding is discussed in Section IV. Finally, we conclude with open problems.

Mars Davletshin is with the Moscow Research Center, Huawei Technologies Co., Ltd (email: davletshin.mars1@huawei.com).

Nikita Polyanskii is with CDISE, Skolkovo Institute of Science and Technology, and Department of Mathematics, Technion – Israel Institute of Technology. The research was conducted in part during May - October 2017 with the Moscow Research Center, Huawei Technologies Co., Ltd (email: nikitapolyansky@gmail.com).

Rina Revenko is with the Moscow Research Center, Huawei Technologies Co., Ltd (email: revenko.rina@huawei.com).

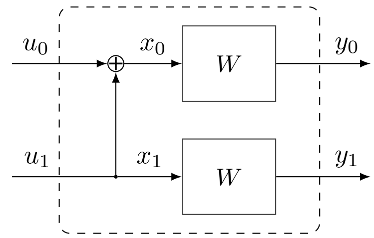

Fig. 1. The channel W_2.

II. Notations and Definitions

For simplicity of presentation we shall use zero-based numbering. A vector of length n is denoted by bold lowercase letters, such as $\boldsymbol{x}$ or $\boldsymbol{x}_0^{n-1}$, and the ith entry of the vector $\boldsymbol{x}$ is referred to as x_i. Given a binary vector $\boldsymbol{x}$, we define its support $\mathrm{supp}(\boldsymbol{x})$ as the set of coordinates in which the vector $\boldsymbol{x}$ has nonzero entries. Let $\mathrm{d}(\boldsymbol{x}, \boldsymbol{y})$ be the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, and $\mathrm{wt}(\boldsymbol{x})$ be the Hamming weight of $\boldsymbol{x}$. The set of integers from i to $j-1$, $0 \le i \le j$, is abbreviated to $[i, j)$ or simply $[j]$ if $i = 0$. Clearly, $\mathrm{wt}(\boldsymbol{x}) = \mathrm{d}(\boldsymbol{x}, \boldsymbol{0})$, where $\boldsymbol{0}$ is the all-zero vector. Given a $(N \times N)$ binary matrix X and $\mathcal{A} \subset [N]$, we write $X(\mathcal{A})$ to denote the $(|\mathcal{A}| \times N)$ submatrix of X formed by the rows of X with indexes in $\mathcal{A}$.

Let $W : \mathcal{X} \to \mathcal{Y}$ be a B-MSC channel with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y}$, and transition probabilities $W(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. By W^N we denote the vector channel corresponding to N independent copies of W, i.e., $W^N : \mathcal{X}^N \to Y^N$ with transition probabilities $W^N(\boldsymbol{y}_0^{N-1}|\boldsymbol{x}_0^{N-1}) = \prod_{i=0}^{N-1} W(y_i|x_i)$.

Arikan used a construction based on the following kernel matrix

$$G_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Given $N = 2^n$, we consider a $(N \times N)$ binary matrix $G_N = G_2^{\otimes n}$ by performing the nth Kronecker power of G_2. We denote the ith row of G_N by $\boldsymbol{g}_i$. The matrix G_N defines a linear mapping $\mathcal{X}^n \to \mathcal{X}^n$ by

$$\boldsymbol{x} = \boldsymbol{u} G_N, \tag{1}$$

where the matrix G_N, the vectors $\boldsymbol{x}$, $\boldsymbol{u}$, and the vector space $\mathcal{X}^n$ are over $GF(2)$. Let us produce a vector channel $W_N : \mathcal{X}^N \to \mathcal{Y}^N$ as follows

$$W_N(\boldsymbol{y}|\boldsymbol{u}) = W^N(\boldsymbol{y}|\boldsymbol{u}G_N) = W^N(\boldsymbol{y}|\boldsymbol{x}).$$

In Figure 1, the channel W_2 is depicted. For $i \in [N]$, we define the synthetic channel $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^i$ as

$$W_N^{(i)}(\boldsymbol{y}, \boldsymbol{u}_0^{i-1}|u_i) := \sum_{\boldsymbol{u}_{i+1}^{N-1} \in \mathcal{X}^{N-i-1}} \frac{1}{2^{N-1}} W_N(\boldsymbol{y}|\boldsymbol{u}).$$

A. Polar Coding

The generator matrix of a polar code is given by $G_N(\mathcal{A})$ for some set $\mathcal{A} \subset [N]$, which is referred to as the information set. The indices $\mathcal{A}^c := [N] \setminus \mathcal{A}$ are usually called frozen ones and chosen carefully according to the reliabilities of the synthetic channels [1]. In other words, any message $\boldsymbol{u} \in \{0,1\}^N$ has $u_i = 0$ for all $i \in \mathcal{A}^c$, and is mapped to the codeword $\boldsymbol{x}$ by (1).

Let $\boldsymbol{x}$ be sent over W_N, and let a channel output $\boldsymbol{y}$ be received. Given $\mathcal{A}$ and $\boldsymbol{y}$, the decoder generates an estimate $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$. We shall briefly describe the SC decoding as the sequential use of the closest coset decoding [8].

For any binary vector $\boldsymbol{v} \in \{0,1\}^i$ and $i \le N$, let the set $C(\boldsymbol{v})$ induced by $\boldsymbol{v}$ be defined as follows

$$C(\boldsymbol{v}) := \sum_{j \in \operatorname{supp}(\boldsymbol{v})} \boldsymbol{g}_j + \langle \boldsymbol{g}_i, \ldots, \boldsymbol{g}_{N-1} \rangle,$$

where $\langle \cdot \rangle$ is a linear span of a set of vectors. By

$$C(\boldsymbol{v}, 0) := \sum_{j \in \operatorname{supp}(\boldsymbol{v})} \boldsymbol{g}_j + \langle \boldsymbol{g}_{i+1}, \ldots, \boldsymbol{g}_{N-1} \rangle,$$
$$C(\boldsymbol{v}, 1) := \boldsymbol{g}_i + C^{(m)}(\boldsymbol{v}, 0),$$

define the zero and the one cosets induced by $\boldsymbol{v}$, respectively. Obviously, the disjoint union of the zero and the one cosets coincides with $C(\boldsymbol{v})$.

At the beginning of the ith stage of the SC decoding, we are given with a binary vector $\hat{\boldsymbol{u}}_0^{i-1} \in \{0,1\}^i$, which can be treated as an estimate of $\boldsymbol{u}_0^{i-1}$. If $i \in \mathcal{A}^c$, then the decoder makes a bit decision $\hat{u}_i = 0$. Otherwise, the decoder computes the log-likelihood ratio

$$L_i := \log \frac{\sum\limits_{\boldsymbol{v} \in C(\hat{\boldsymbol{u}}_0^{i-1}, 0)} W^N(\boldsymbol{y}|\boldsymbol{v})}{\sum\limits_{\boldsymbol{v} \in C(\hat{\boldsymbol{u}}_0^{i-1}, 1)} W^N(\boldsymbol{y}|\boldsymbol{v})}$$

and makes a bit estimate $\hat{u}_i$ of u_i: $\hat{u}_i = 0$ if $L_i \ge 0$, and $\hat{u}_i = 1$ otherwise. This decision can be seen as choosing the "closest" (zero or one) coset to the received $\boldsymbol{y}$. If the wrong coset is selected at some decoding stage, then this decoding error is propagated to the next stages.

B. Weight Distribution

Without loss of generality, we assume that the all-zero codeword is transmitted, i.e., $\boldsymbol{u} = \boldsymbol{x} = \boldsymbol{0}$. At the ith stage, the error occurs if the decoder selects $C(\boldsymbol{0}_0^{i-1}, 1)$ instead of $C(\boldsymbol{0}_0^{i-1}, 0)$. Let $S_{i,w}$ be the number of codewords of weight w in $C(\boldsymbol{0}_0^{i-1}, 1)$. Thus for BPSK transmission over the AWGN channel with variance σ^2 and $i \in \mathcal{A}$, we upper bound [9] the error probability $P_e(i)$ at the ith decoding stage as

$$P_e(i) \le \sum_{w=0}^{N} \frac{1}{2} S_{i,w} \operatorname{erfc}\left(\sqrt{w/(2\sigma^2)}\right), \tag{3}$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function defined by

$$\operatorname{erfc}(x) := \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt.$$

In Figure 2, we make use of more accurate bound, namely the sphere upper bound [10, Inequality (14)]

$$P_e(i) \le \min_{r>0} \Bigg\{ \sum_{w=0}^{\lfloor \sqrt{r} \rfloor} \frac{1}{2} S_{i,w} \left(\operatorname{erfc}\left(\sqrt{w/(2\sigma^2)}\right) \right. \tag{4}$$
$$\left. - \operatorname{erfc}\left(r/\sqrt{2\sigma^2}\right)\right) F_{\chi^2(N)}(r^2 - w)$$
$$+ (1 - F_{\chi^2(N)}(r^2)) \Bigg\},$$

where $F_{\chi^2(N)}(\cdot)$ is the cumulative distribution function of the chi-squared distribution with N degrees of freedom.

It is worth noting that there are several techniques allowing to calculate $P_e(i)$ with inherent inaccuracy. Among them are density evolution (DE) [11], [12] and Gaussian approximation (GA) [3]. However, the DE and GA computations give only some approximation of the error probability, whereas our method produces a fair upper bound.

III. Weight Distribution for Successive Cancellation Decoding

In this section we give an algorithm for computing WDs for the SC decoding. Our analysis is similar to one in [5, Section 2], where WD for the closest coset decoding of $|u|u+v|$ construction was established. Given $N = 2^n$ let us determine $S_{i,w}^{(n)} := S_{i,w}$, the number of codewords of weight w in $C^{(n)}(\boldsymbol{0}_0^{i-1}, 1) := C(\boldsymbol{0}_0^{i-1}, 1)$. If $i \ge 2^{n-1}$, then any codeword in $C^{(n)}(\boldsymbol{0}_0^{i-1}, 1)$ represents a repetition of some codeword in $C^{(n-1)}(\boldsymbol{0}_0^{i-1-N/2}, 1)$. Thus $S_{i,w}^{(n)} = 0$ for odd w, and $S_{i,w}^{(n)} = S_{i-N/2,w/2}^{(n-1)}$ for even w. If $i < 2^{n-1}$, then any codeword $\boldsymbol{x} \in C^{(n)}(\boldsymbol{0}_1^{i-1}, 1)$ can be uniquely represented in the form

$$\boldsymbol{x} = \boldsymbol{g}_i + \sum_{j \in I_1} \boldsymbol{g}_j + \sum_{j \in I_2} \boldsymbol{g}_j = (\boldsymbol{x}_1, \boldsymbol{0}) + (\boldsymbol{x}_2, \boldsymbol{x}_2), \tag{5}$$

where the index sets $I_1 \subset [i+1, 2^{n-1})$ and $I_2 \subset [2^{n-1}, 2^n)$, and the codewords $\boldsymbol{x}_1 \in C^{(n-1)}(\boldsymbol{0}_0^{i-1}, 1)$ and $\boldsymbol{x}_2 \in \{0,1\}^{N/2}$. Moreover any pair of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ gives $\boldsymbol{x} \in C^{(n)}(\boldsymbol{0}_0^{i-1}, 1)$. Hence $S_{i,w}^{(n)}$ can be determined using the following statement.

Theorem 1. *For $t \in \{0, \ldots, N/2 - w_1\}$, the contribution of $\boldsymbol{x}_1$ with $\operatorname{wt}(\boldsymbol{x}_1) = w_1$ to $S_{i,w_1+2t}^{(n)}$ is $2^{w_1}\binom{N/2-w_1}{t}$.*

Proof of Theorem 1. It is easy to check that

$$\operatorname{wt}(\boldsymbol{x}) = \operatorname{wt}(\boldsymbol{x}_2) + \operatorname{d}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \operatorname{wt}(\boldsymbol{x}_1)$$
$$+ (\operatorname{wt}(\boldsymbol{x}_2) + \operatorname{d}(\boldsymbol{x}_1, \boldsymbol{x}_2) - \operatorname{wt}(\boldsymbol{x}_1)) \ge \operatorname{wt}(\boldsymbol{x}_1). \tag{6}$$

Algorithm 1 Computing the weight distributions

Input: length $N = 2^n$

Output: weight distributions $((S_{i,w} : w \in \{0, \ldots, N\}))_{i=1}^N$

Initialisation :

1: $S^{(0)} := \left(S_{0,0}^{(0)}, S_{0,1}^{(0)}\right)$, where $S_{0,0}^{(0)} := 0$ and $S_{0,1}^{(0)} := 1$
2: **for** Kroneker's power $j = 1$ to n **do**
3:   **for** row index $i = 0$ to $2^j - 1$ **do**
4:     **if** $(i < 2^{j-1})$ **then**
5:       **for** weight $w = 0$ to 2^j **do**
6:         $S_{i,w}^{(j)} := \sum_{t=(w-2^{j-1})^+}^{\lfloor w/2 \rfloor} S_{i,w-2t}^{(j-1)} \binom{2^{j-1}+2t-w}{t} 2^{w-2t}$
7:       **end for**
8:     **else**
9:       **for** weight $w = 0$ to 2^{j-1} **do**
10:         $S_{i,2w}^{(j)} := S_{i-2^{j-1},w}^{(j-1)}$
11:       **end for**
12:     **end if**
13:   **end for**
14:   $S^{(j)} := \left(\left(S_{i,w}^{(j)} : w \in \{0, \ldots, 2^j\}\right)\right)_{i=0}^{2^j-1}$
15: **end for**
16: **return** $S^{(n)}$

We observe that the sum in the brackets is equal to the double number of coordinates i so that $x_{1,i} = 0$ and $x_{2,i} = 1$. Given $\boldsymbol{x}_1$ with $\mathrm{wt}(\boldsymbol{x}_1) = w_1$, there are $\binom{N/2-w_1}{t}$ different choices for placing t ones in $\boldsymbol{x}_2$ among $N/2-w_1$ coordinates corresponding to zeros in $\boldsymbol{x}_1$. Also $\boldsymbol{x}_2$ could have anything in the remaining w_1 coordinates corresponding to ones in $\boldsymbol{x}_1$. Therefore the total number of choices for $\boldsymbol{x}_2$ is $2^{w_1}\binom{N/2-w_1}{t}$. □

Summarizing the arguments given above, WDs can be calculated in a recurrent manner with the help of Algorithm 1, where the notation $(k)^+$ stands for $\max(0, k)$.

Remark 1. *Algorithm 1 provides a practical way to determine WDs for SC. For instance, given $p \geq 1$ the complexity of computing the p first nonzero components of $(S_{i,w})_{w=0}^N$ for $i \in [N]$ is $O(N)$ as $N \to \infty$.*

Example 1. *Let us illustrate the bound* (4) *by taking the code length $N = 128$ and the synthetic channel with index $i = 72$. Using Algorithm 1 we compute the weight distributions and depict the sphere bound* (4) *on the decoding error probability $P_e(72)$ along with this probability, calculated with the help of DE, in Figure 2.*

A. First Nonzero Component

It is well known that the first nonzero component of $(S_{i,w})_{w=0}^N$ corresponds to $w_i = \mathrm{wt}(\boldsymbol{g}_i)$. We extend this line of research and find an explicit formula for S_{i,w_i}. Let $b(j,k)$ be the jth bit in the binary representation of k, and $p(j,k)$ is a partial sum of the first j bits, i.e.,

$$k = \sum_{j=0}^{n-1} b(j,k) 2^j \quad \text{and} \quad p(j,k) = \sum_{s=0}^{k} b(s,k).$$

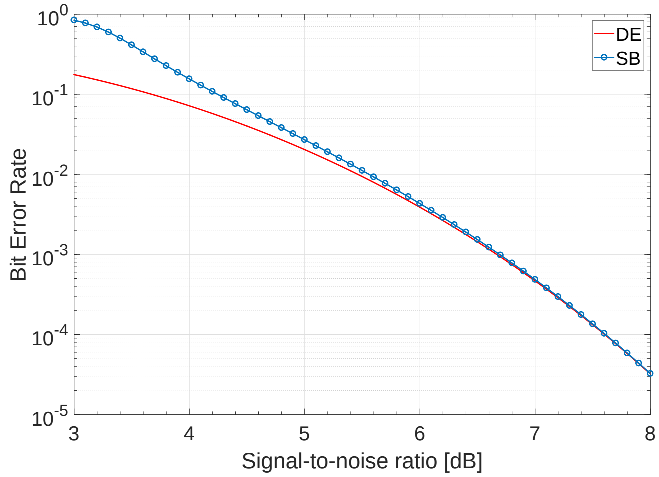

Fig. 2. The DE computations and the sphere bound on $P_e(72)$ under SC

Theorem 2. *Given $N = 2^n$ and $i \in [N]$ the first nonzero component of $(S_{i,w})_{w=0}^N$ is S_{i,w_i} with w_i so that $\log_2(w_i) = p(n-1, i)$. Moreover*

$$\log_2 S_{i,w_i} = \sum_{j=0}^{n-1} 2^{p(j,i)}(1 - b(j,i)). \tag{7}$$

Proof of Theorem 2. We shall prove the statement of this theorem by induction on n. The base case is evident as $S_{0,1} = 1$ for $n = 0$. We assume for a moment that the equality (7) holds for $S_{i,w_i} =: S_{i,w_i^{(n)}}^{(n)}$. Following line 6 of Algorithm 1, we get

$$S_{i,w_i^{(n)}}^{(n+1)} = S_{i,w_i^{(n)}}^{(n)} 2^{w_i} \quad \text{and} \quad S_{i,w}^{(n+1)} = 0$$

for $i < N = 2^n$ and $w < w_i^{(n)}$. Similarly we have

$$S_{i,2w_i^{(n)}}^{(n+1)} = S_{i-N,w_i^{(n)}}^{(n)} \quad \text{and} \quad S_{i,w}^{(n+1)} = 0$$

for $i \geq N$ and $w < 2w_i^{(n)}$. In other words, if the nth bit $b(n,i)$ in the binary expansion equals 0, then the first nonzero component of WD parametrized by i and $n+1$ has $w_i^{(n+1)} = w_i^{(n)}$ and is $2^{p(j,i)}$ times as much as one of the previous WD. In other case, the value of the first nonzero component on the next level parametrized by $i, i \geq N$, and $n+1$ remains the same as one parametrized by $i-N$ and n. Combining the arguments given above completes the proof of the inductive step. □

It was observed [6], [7] that there is a partial order between the synthetic channels, which holds for any B-MSC channel. We rephrase this result using our notations.

Theorem 3 (The partial order [6, Definition 8])**.** *For any B-MSC W, if $p(k,i) \leq p(k,j)$ for all $k \in \{0, \ldots, n-1\}$ and given $i, j \in [N]$, then the synthetic channel $W_N^{(i)}$ is stochastically degraded by $W_N^{(j)}$.*

Applying Theorem 2 we find a natural connection between the partial order given in Theorem 3 and the first nonzero components of WDs. Namely, if $p(k,i) \le p(k,j)$ for all $k \in [n]$ and given $i,j \in [N]$, then there holds either $w_i < w_j$, or $w_i = w_j = w$ and $S_{i,w} < S_{j,w}$. In other words, if the ith synthetic channel is worse than the jth one by the partial order, then the first nonzero component of WD at the ith decoding stage of SC is worse than that at the jth decoding stage. Unfortunately, we are not able to prove any converse statement.

IV. Toward Weight Distribution for Successive Cancellation List Decoding

Let us briefly recall the high level description of the successive cancellation list (SCL) decoder [13] with the list size L. At i-th decoding stage for $i \in \mathcal{A}$, we split each path $\hat{\boldsymbol{u}}_0^{i-1}$ with the path metric PM into two paths by taking $\hat{u}_i = 0$ and $\hat{u}_i = 1$ and assigning the path metrics for the two paths as $PM_0 := PM - |L_i|(1-\mathrm{sgn}(L_i))/2$ and $PM_1 := PM - |L_i|(1+\mathrm{sgn}(L_i))/2$, respectively. Since the number of paths is doubled, we keep only the L most likely paths at each stage. The pruning criterion is based on the values of the path metrics. If a bit $i \in \mathcal{A}^c$, then, for any path $\hat{\boldsymbol{u}}_0^{i-1}$ with the path metric PM, the decoder makes a bit decision $\hat{u}_i = 0$ and assign the path metric $PM_0 := PM - |L_i|(1-\mathrm{sgn}(L_i))/2$ to this path.

Assume that after the j-th decoding stage, the SCL decoder keeps (at least) the following two paths: true path $\mathbf{0}_0^j$ and path $\boldsymbol{u}_0^j$ mistaken in only two positions i and j, $i < j$, i.e., $\mathrm{supp}(\boldsymbol{u}_0^j) = \{i,j\}$. Our goal is to estimate the minimal distance between sets induced by these two pathes. For simplicity of notation we abbreviate $C(\boldsymbol{u}_0^j)$ by $C(i,j)$.

Theorem 4. *The minimal weight of any codeword in $C(i,j)$ is*

$$\min_{\boldsymbol{x} \in C(i,j)} \mathrm{wt}(\boldsymbol{x}) = \mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j). \tag{8}$$

Proof of Theorem 4. We shall prove the statement of this theorem by induction on n, $N = 2^n$. The base case $n = 1$ is obviously true. Now assume that the equation (8) holds for every $C^{(n)}(i,j) := C(i,j)$, where $n \le \bar{n}$. We prove that it holds for $n = \bar{n}+1$. Let us consider one of the three cases.

Case 1: $2^{n-1} \le i < j < 2^n$. Let i' and j' be the residues of i and j modulo 2^{n-1}, respectively. Given $\alpha_r \in \{0,1\}$ for $r \in [j+1, 2^n)$, the weight of any binary vector $\boldsymbol{x} \in C^{(n)}(i,j)$ represented by

$$\boldsymbol{x} = \boldsymbol{g}_i + \boldsymbol{g}_j + \sum_{r \in [j+1, 2^n)} \alpha_r \boldsymbol{g}_r \tag{9}$$

is exactly two times larger as the weight of the binary vector

$$\boldsymbol{x}' = \boldsymbol{g}_{i'} + \boldsymbol{g}_{j'} + \sum_{r \in [j'+1, 2^{n-1})} \alpha_{r+2^{n-1}} \boldsymbol{g}_r,$$

Therefore we deduce from the induction that

$$\begin{aligned}\min_{\boldsymbol{x} \in C^{(n)}(i,j)} \mathrm{wt}(\boldsymbol{x}) &= 2 \min_{\boldsymbol{x}' \in C^{(n-1)}(i',j')} \mathrm{wt}(\boldsymbol{x}') \\ &= 2\,\mathrm{wt}(\boldsymbol{g}_{i'} + \boldsymbol{g}_{j'}) = \mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j).\end{aligned}$$

Case 2: $i < j < 2^{n-1}$. Any binary vector in $\boldsymbol{x} \in C(i,j)$ can be written in the form $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$, where

$$\boldsymbol{x}_1 = \boldsymbol{g}_i + \boldsymbol{g}_j + \sum_{r \in [j+1, 2^{n-1})} \alpha_r \boldsymbol{g}_r, \quad \boldsymbol{x}_2 = \sum_{r \in [2^{n-1}, 2^n - 1)} \alpha_r \boldsymbol{g}_r.$$

Applying (5)-(6) and the inductive assumption, we establish $\mathrm{wt}(\boldsymbol{x}) \ge \mathrm{wt}(\boldsymbol{x}_1) \ge \mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j)$.

Case 3: $i < 2^{n-1} \le j < 2^n$. Let ℓ be the minimal integer such that $g_{i,2^\ell} = 0$. By I denote the collection of indices k such that $k = k_m 2^\ell + k_r$, $0 \le k_r < 2^\ell$, and k_m is odd. Hence $g_{i,k} = 0$ for any $k \in I$ and $|I| = 2^{n-1}$. If $\ell = n-1$, then $\boldsymbol{g}_i = (\mathbf{1}, \mathbf{0})$, and the weight of any vector $\boldsymbol{x} \in C(i,j)$ represented by (9) is exactly $\mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j) = \mathrm{wt}(\boldsymbol{g}_i) = 2^{n-1}$. If $\ell < n-1$, then we consider two subcases.

Case 3.a: $g_{j,k} = 0$ for all $k \in I$. For any given vector $\boldsymbol{x} \in C(i,j)$, we split the terms in (9) into two groups: in the first one, $g_{r,k} = 0$ for all $k \in I$; the remaining terms go to the second group. Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be the sums of the vectors in the first and the second groups, respectively. For $\boldsymbol{x}_1$, the vector projection of all the terms onto coordinates indexed by $I^c := [2^n] \setminus I$ maintains the assumption in the induction. Hence $\mathrm{wt}(\boldsymbol{x}_1) \ge \mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j)$. Since the terms $\boldsymbol{g}_r$ included in $\boldsymbol{x}_2$ satisfy the equality $g_{r,k} = g_{r,k-2^\ell}$ for all $k \in I$, we apply the arguments as in (5)-(6) and conclude that $\mathrm{wt}(\boldsymbol{x}) = \mathrm{wt}(\boldsymbol{x}_1 + \boldsymbol{x}_2) \ge \mathrm{wt}(\boldsymbol{x}_1) \ge \mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j)$

Case 3.b: $g_{j,k} = g_{j,k-2^\ell}$ for all $k \in I$. For any given vector $\boldsymbol{x} \in C(i,j)$, we consider its projections $\boldsymbol{x}|_I$ and $\boldsymbol{x}|_{I^c}$ onto coordinates indexed by I and I^c, respectively. For the vector $\boldsymbol{x}|_{I^c}$ and all the terms in (9) restricted onto I^c, we apply the inductive assumption. Therefore $\mathrm{wt}(\boldsymbol{x}|_{I^c}) \ge \mathrm{wt}(\boldsymbol{g}_i|_{I^c} + \boldsymbol{g}_j|_{I^c})$. For the vector $\boldsymbol{x}|_I$, we deduce from the induction that $\mathrm{wt}(\boldsymbol{x}|_I) \ge \mathrm{wt}(\boldsymbol{g}_j|_I)$. Finally, we have $\mathrm{wt}(\boldsymbol{x}) = \mathrm{wt}(\boldsymbol{x}|_I) + \mathrm{wt}(\boldsymbol{x}|_{I^c}) \ge \mathrm{wt}(\boldsymbol{g}_i|_{I^c} + \boldsymbol{g}_j|_{I^c}) + \mathrm{wt}(\boldsymbol{g}_j|_I) = \mathrm{wt}(\boldsymbol{g}_i + \boldsymbol{g}_j)$. $\square$

V. Open Problems

The upper bounds (3)-(4) take into account only WD of the one coset $C(\mathbf{0}_0^i, 1)$. The drawback of this approach is evident: for low and medium signal-to-noise ratio, the bounds could not be tight. Based on Algorithm 1, WD of any zero coset $C(\mathbf{0}_0^i, 0)$ can be calculated. However, we do not know how to use it in order to get a more accurate bound of the error probability.

It is still unknown how to calculate efficiently the minimal weight (and the first nonzero component) of a set $C(\boldsymbol{u}_0^j)$ with an arbitrary $\boldsymbol{u}_0^j$. We believe that such an analysis can be helpful for constructing polar codes under the SC list decoding. In addition, it may be reasonable to use polar codes with dynamic frozen symbols [14].

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] G. J. Pottie and D. P. Taylor, "Multilevel codes based on partitioning," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 87–98, 1989.

[3] P. Trifonov, “Efficient design and decoding of polar codes,” *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3221–3227, 2012.

[4] H. Herzberg, “On the spectrum of distances of a multilevel code, decoded by a multistage decoder,” *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1736–1740, 1997.

[5] M. P. Fossorier and S. Lin, “Weight distribution for closest coset decoding of $|u|u + v|$ constructed codes,” *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1028–1030, 1997.

[6] C. Schürch, “A partial order for the synthesized channels of a polar code,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 220–224.

[7] M. Bardet, V. Dragoi, A. Otmani, and J.-P. Tillich, “Algebraic properties of polar codes from a new polynomial formalism,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 230–234.

[8] F. Hemmati, “Closest coset decoding of $|u|u+v|$ codes,” *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, pp. 982–988, 1989.

[9] J. G. Proakis, *Digital Communication*, 4th ed. McGraw Hill, 2004.

[10] H. Herzberg and G. Poltyrev, “Techniques of bounding the probability of decoding error for block coded modulation structures,” *IEEE Transactions on Information Theory*, vol. 40, no. 3, pp. 903–911, 1994.

[11] R. Mori and T. Tanaka, “Performance and construction of polar codes on symmetric binary-input memoryless channels,” in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 1496–1500.

[12] I. Tal and A. Vardy, “How to construct polar codes,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, 2013.

[13] ——, “List decoding of polar codes,” *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.

[14] P. Trifonov and V. Miloslavskaya, “Polar codes with dynamic frozen symbols and their decoding by directed search,” in *Information Theory Workshop (ITW), 2013 IEEE*. IEEE, 2013, pp. 1–5.