

Weight Distributions for Successive Cancellation Decoding of Polar Codes

Rina Polyanskaya, Mars Davletshin, and Nikita Polyanskii

Abstract—In this paper, we derive the exact weight distributions that emerge during each stage of successive cancellation decoding of polar codes. Though we do not compute the distance spectrum of polar codes, the results allow us to get an estimate of the decoding error probability and to show a link between the first nonzero components of the weight distribution and the partial order between the synthetic channels. Also, we establish the minimal distance between two cosets associated with two paths that differ in two positions. This can be regarded as a first step toward analyzing the weight distributions for successive cancellation list decoding.

Index Terms—Polar codes, weight distribution, closest coset decoding, partial order, successive cancellation decoding.

I. INTRODUCTION

Polar codes, introduced by Arkan [1], provably achieve the symmetric capacity of any binary-input memoryless symmetric channels (B-MSC) with encoding and decoding complexity $\Theta(N \log N)$, where N is the block length of the code. Within the 5G standardization process, polar codes have been adopted as channel codes for uplink and downlink control information of the enhanced mobile broadband communication service.

Multilevel codes are based on partitioning and, thus, multistage decoding is the most natural one to be performed [2]. Polar codes with 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. 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 [1, Section X], [4] that polar codes with good distance properties turn out to have a poor performance under SC decoding. To evaluate the error rate provided by a multistage decoder, it is natural to calculate the weight distribution (WD) between cosets (or the distance spectrum of component codes) at all the stages [5]. However, only the minimal distance for SC decoding of polar codes is known at present [6, Lemma 6.2]. The aim of our paper is to calculate the WDs at all the stages of SC decoding.

We want to point out that we do not compute the distance spectrum of polar codes in this paper. Attempts pursuing the latter were undertaken by many authors. For example, the authors of [7] proposed to send the all-zero codeword over a noisy channel and decode the received vector utilizing the

successive cancellation list decoder [8]. After decoding, the weights of codewords from the list are calculated and a special weight function is updated. The paper [9] suggests a way how to compute a probabilistic weight distribution expression efficiently. The authors of [10] put forward an idea how to get an approximate distance spectrum of polar codes with large length using the spectrum of short ones and some probabilistic assumption on appearing ones in codewords.

A. Outline

The rest of the paper is organized as follows. In Section II, we give key definitions and notations of polar codes and the WDs associated with SC decoding. We derive the WDs and focus our attention on their first nonzero component in Section III. To obtain an algorithm computing the WDs, we exploit a similar idea as in [11], where a $|u|u+v|$ construction is investigated. Also, we find a natural connection between the first nonzero component of the WDs and the partial order [12], [13]. The minimal distance between cosets for SC list decoding is discussed in Section IV. Finally, we conclude with some open problems in Section V.

II. NOTATIONS AND DEFINITIONS

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

Let $W : \mathcal{X} \rightarrow \mathcal{Y}$ be a B-MSC 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., $W^N : \mathcal{X}^N \rightarrow \mathcal{Y}^N$ with transition probabilities

$$W^N(\mathbf{y}_0^{N-1} | \mathbf{x}_0^{N-1}) = \prod_{i=0}^{N-1} W(y_i | x_i).$$

Arkan used a construction based on the following kernel matrix

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

Rina Polyanskaya is with the Institute for Information Transmission Problems (email: rev-rina@yandex.ru).

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

Nikita Polyanskii is with the Technical University of Munich. The research was conducted in part during May - October 2017 with the Moscow Research Center, Huawei Technologies Co., Ltd (email: nikitapolysky@gmail.com). N. Polyanskii was funded in part by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Grant No. WA3907/1-1.

Given $N = 2^n$, we consider the $(N \times N)$ binary matrix $G_N := G_2^{\otimes n}$ by performing the n th Kronecker power of G_2 . We denote the i th row of G_N by \mathbf{g}_i . Usually a linear mapping $\mathbf{x} = \mathbf{x}(\mathbf{u}) : \mathcal{X}^n \rightarrow \mathcal{X}^n$ is defined by

$$\mathbf{x} = \mathbf{u} B_N G_N,$$

where B_N is the $(N \times N)$ bit-reversal permutation matrix defined in [1, Section VII-B], and the vectors \mathbf{x} , \mathbf{u} , and the vector space \mathcal{X}^n are over $GF(2)$. However, since $B_N G_N = G_N B_N$ [1, Proposition 16], the latter being a simple permutation on \mathbf{x} , we can dispense with B_N in this paper and assume

$$\mathbf{x} = \mathbf{u} G_N. \quad (1)$$

Let us produce a vector channel $W_N : \mathcal{X}^N \rightarrow \mathcal{Y}^N$ as follows

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

Given $i \in [0, N)$, we define the synthetic channel $W_N^{(i)} : \mathcal{X} \rightarrow \mathcal{Y}^N \times \mathcal{X}^i$ as

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

A. Polar Coding

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

Let \mathbf{x} be sent over W^N , and let a channel output \mathbf{y} be received. Given \mathcal{A} and \mathbf{y} , the decoder generates an estimate $\hat{\mathbf{u}}$ of \mathbf{u} . We shall briefly describe SC decoding as the sequential use of the closest coset decoding [14].

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

$$C^{(n)}(\mathbf{v}) := \sum_{j \in \text{supp}(\mathbf{v})} \mathbf{g}_j + \langle \mathbf{g}_i, \dots, \mathbf{g}_{N-1} \rangle,$$

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

$$\begin{aligned} C^{(n)}(\mathbf{v}, 0) &:= \sum_{j \in \text{supp}(\mathbf{v})} \mathbf{g}_j + \langle \mathbf{g}_{i+1}, \dots, \mathbf{g}_{N-1} \rangle, \\ C^{(n)}(\mathbf{v}, 1) &:= \mathbf{g}_i + C^{(n)}(\mathbf{v}, 0), \end{aligned}$$

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

At the beginning of the i th stage of SC decoding, we are given a binary vector $\hat{\mathbf{u}}_0^{i-1} \in \{0, 1\}^i$, which can be treated as an estimate of \mathbf{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 values $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 0)$ and $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 1)$, where the value $W_N^{(i)}(\mathbf{y}, \mathbf{u}_0^{i-1} | u_i)$ and the set $C^{(n)}(\mathbf{u}_0^{i-1}, u_i)$ are connected by

$$W_N^{(i)}(\mathbf{y}, \mathbf{u}_0^{i-1} | u_i) = \frac{1}{2^{N-1}} \sum_{\mathbf{v} \in C^{(n)}(\mathbf{u}_0^{i-1}, u_i)} W^N(\mathbf{y}|\mathbf{v}). \quad (3)$$

TABLE I
WEIGHT DISTRIBUTIONS $S_i^{(n)}$ FOR $n = 3$

Rows i	Weights w								
	0	1	2	3	4	5	6	7	8
0	0	8	0	56	0	56	0	8	0
1	0	0	16	0	32	0	16	0	0
2	0	0	8	0	16	0	8	0	0
3	0	0	0	0	16	0	0	0	0
4	0	0	4	0	0	0	4	0	0
5	0	0	0	0	4	0	0	0	0
6	0	0	0	0	2	0	0	0	0
7	0	0	0	0	0	0	0	0	1

Then the decoder makes a bit estimate \hat{u}_i of u_i : $\hat{u}_i = 0$ if $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 0) > W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 1)$, and $\hat{u}_i = 1$ if $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 0) < W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 1)$. For the case $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 0) = W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1} | 1)$, the decoder chooses the value of \hat{u}_i randomly and uniformly. This decision rule can be seen as choosing the “closest” (zero or one) coset to the received \mathbf{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 (e.g., see [1, Section VI], i.e., $\mathbf{u} = \mathbf{x} = \mathbf{0}$). At the i th stage, an error occurs if the decoder selects $C^{(n)}(\mathbf{0}_0^{i-1}, 1)$ instead of $C^{(n)}(\mathbf{0}_0^{i-1}, 0)$ (in the very beginning $C^{(n)}(1)$ instead of $C^{(n)}(0)$, respectively). Let us introduce the weight distribution for SC decoding. For $i \in [0, N)$ and $w \in [N]$, let $S_{i,w}^{(n)}$ be the number of words of weight w in $C^{(n)}(\mathbf{0}_0^{i-1}, 1)$ and $S_i^{(n)} := \left(S_{i,w}^{(n)} \right)_{w=0}^N$.

Example 1. In Table I we illustrate weight distributions $S_i^{(n)}$ for $n = 3$ and different i 's. Rows in this table correspond to values i 's and columns represent weights w 's, whereas the entry at position (i, w) is $S_{i,w}^{(n)}$. One can easily check that the weight distributions are symmetric in w , i.e. $S_{i,w}^{(n)} = S_{i,N-w}^{(n)}$ for all i except the case $i = N - 1$. This holds since \mathbf{g}_{N-1} is the all-one vector.

C. Approximate Upper Bound on the Error Probability

The weight distribution is useful for obtaining upper bounds of the error probability. One approximate bound is introduced in this subsection. Let $P_e(i)$ be the error probability at the i th decoding stage, i.e.,

$$\begin{aligned} P_e(i) &= \Pr \left\{ W_N^{(i)}(\mathbf{y}, \mathbf{0}_0^{i-1} | 0) < W_N^{(i)}(\mathbf{y}, \mathbf{0}_0^{i-1} | 1) \right\} \\ &\quad + \Pr \left\{ W_N^{(i)}(\mathbf{y}, \mathbf{0}_0^{i-1} | 0) = W_N^{(i)}(\mathbf{y}, \mathbf{0}_0^{i-1} | 1) \right\} / 2. \end{aligned}$$

This implies by (3)

$$P_e(i) \leq \Pr \left\{ W^N(\mathbf{y}|\mathbf{0}) \leq \sum_{\mathbf{v} \in C^{(n)}(\mathbf{0}_0^{i-1}, 1)} W^N(\mathbf{y}|\mathbf{v}) \right\}.$$

For the binary phase-shift keying (BPSK) transmission over the additive white Gaussian noise (AWGN) channel with small

enough variance σ^2 , the right-hand side of the inequality above can be well-approximated by

$$\Pr \left\{ W^N(\mathbf{y}|\mathbf{0}) \leq \max_{\mathbf{v} \in C^{(n)}(\mathbf{0}_0^{i-1}, 1)} W^N(\mathbf{y}|\mathbf{v}) \right\},$$

whereas the latter is upper bounded [15, Equations (4.109) and (4.110)] by the union bound

$$P_{ub}(i) := \sum_{w=1}^N \frac{1}{2} S_{i,w}^{(n)} \operatorname{erfc} \left(\sqrt{w/(2\sigma^2)} \right). \quad (4)$$

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

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

It is worth noting that there are several techniques allowing to calculate $P_e(i)$ with inherent inaccuracy and to bound $P_e(i)$. Among them are density evolution (DE) [16], degrading and upgrading algorithms [17] and Gaussian approximation [3].

III. WEIGHT DISTRIBUTION FOR SUCCESSIVE CANCELLATION DECODING

In this section, we first provide an algorithm for efficiently computing the WDs for SC decoding. After that, we derive a short formula for computing the first nonzero component of the WD. Finally, a link between the partial order and the first nonzero component of the WD is shown.

A. Algorithm for Computing the Weight Distributions

Our analysis in this subsection is similar to one in [11, Section 2], where the WD for the closest coset decoding of $|u|u+v|$ construction was established. Let us determine $S_{i,w}^{(n)}$, the number of words of weight w in $C^{(n)}(\mathbf{0}_0^{i-1}, 1)$. First, we recall that $G_N = G_2 \otimes G_{N/2}$ with $N/2 = 2^{n-1}$, i.e.,

$$G_N = \begin{pmatrix} G_{N/2} & 0 \\ G_{N/2} & G_{N/2} \end{pmatrix}.$$

So if $i \geq 2^{n-1}$, then any word in $C^{(n)}(\mathbf{0}_0^{i-1}, 1)$ represents a repetition of some word in $C^{(n-1)}(\mathbf{0}_0^{i-1-N/2}, 1)$. 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 word $\mathbf{x} \in C^{(n)}(\mathbf{0}_1^{i-1}, 1)$ can be uniquely represented in the form

$$\mathbf{x} = \left(\mathbf{g}_i + \sum_{j \in I_1} \mathbf{g}_j \right) + \sum_{j \in I_2} \mathbf{g}_j = (\mathbf{x}_1, \mathbf{0}) + (\mathbf{x}_2, \mathbf{x}_2), \quad (5)$$

where the index sets $I_1 \subset [i+1, 2^{n-1})$ and $I_2 \subset [2^{n-1}, 2^n)$, and $\mathbf{x}_1 \in C^{(n-1)}(\mathbf{0}_0^{i-1}, 1)$ and $\mathbf{x}_2 \in \{0, 1\}^{N/2}$. Moreover, since $G_{N/2}$ is full-rank, any $\mathbf{x}_2 \in \{0, 1\}^{N/2}$ can be represented as a sum of \mathbf{g}_j 's. So, any combination of $\mathbf{x}_1 \in C^{(n-1)}(\mathbf{0}_0^{i-1}, 1)$ and $\mathbf{x}_2 \in \{0, 1\}^{N/2}$ gives a unique $\mathbf{x} \in C^{(n)}(\mathbf{0}_0^{i-1}, 1)$. Hence, $S_{i,w}^{(n)}$ can be determined using the following statement.

Theorem 1. For $t \in \{0, \dots, N/2 - w'\}$, the contribution of \mathbf{x}_1 with $\operatorname{wt}(\mathbf{x}_1) = w'$ to $S_{i,w'+2t}^{(n)}$ is $2^{w'} \binom{N/2-w'}{t}$.

Proof of Theorem 1. Let $\mathbf{x} \in C^{(n)}(\mathbf{0}_0^{i-1}, 1)$. By (5), it is easy to check that

$$\begin{aligned} \operatorname{wt}(\mathbf{x}) &= \operatorname{wt}(\mathbf{x}_2) + \operatorname{d}(\mathbf{x}_1, \mathbf{x}_2) = \operatorname{wt}(\mathbf{x}_1) \\ &+ (\operatorname{wt}(\mathbf{x}_2) + \operatorname{d}(\mathbf{x}_1, \mathbf{x}_2) - \operatorname{wt}(\mathbf{x}_1)) \geq \operatorname{wt}(\mathbf{x}_1). \end{aligned} \quad (6)$$

We observe that the sum in the parentheses is equal to double the number of coordinates i so that $x_{1,i} = 0$ and $x_{2,i} = 1$. Given \mathbf{x}_1 with $\operatorname{wt}(\mathbf{x}_1) = w'$, there are $\binom{N/2-w'}{t}$ different choices for placing t ones in \mathbf{x}_2 among $N/2 - w'$ coordinates corresponding to zeros in \mathbf{x}_1 . By (5), \mathbf{x}_2 could have anything in the remaining w' coordinates corresponding to ones in \mathbf{x}_1 . Therefore the total number of choices for \mathbf{x}_2 is $2^{w'} \binom{N/2-w'}{t}$. \square

Summarizing the arguments given above, the WDs can be calculated in a recursive manner with the help of Algorithm 1. We start the algorithm by initializing $\mathbf{S}_0^{(0)} = (0, 1)$ as for the base case, we have $G_0 = (1)$ and $C^{(0)}(1) = \{(1)\}$.

Remark 1. Let $T_{i,w}^{(n)}$ denote the number of words of weight w in $C^{(n)}(\mathbf{0}_0^{i-1}, 0)$. The WDs for the zero coset can be obtained by Algorithm 1 by only changing the initialization step from $\mathbf{S}_0^{(0)} := (0, 1)$ to $\mathbf{T}_0^{(0)} := (1, 0)$ as $C^{(0)}(0) = \{(0)\}$. Let us briefly check this. If $i \geq 2^{n-1}$, then any word in $C^{(n)}(\mathbf{0}_0^{i-1}, 0)$ represents a repetition of some word in $C^{(n-1)}(\mathbf{0}_0^{i-1-N/2}, 0)$. Thus, $T_{i,w}^{(n)} = 0$ for odd w , and $T_{i,w}^{(n)} = T_{i-N/2,w/2}^{(n-1)}$ for even w . If $i < 2^{n-1}$, then any word $\mathbf{x} \in C^{(n)}(\mathbf{0}_1^{i-1}, 0)$ can be uniquely represented in the form

$$\mathbf{x} = \left(\sum_{j \in I_1} \mathbf{g}_j \right) + \sum_{j \in I_2} \mathbf{g}_j = (\mathbf{x}_1, \mathbf{0}) + (\mathbf{x}_2, \mathbf{x}_2),$$

where the index sets $I_1 \subset [i+1, 2^{n-1})$ and $I_2 \subset [2^{n-1}, 2^n)$, and $\mathbf{x}_1 \in C^{(n-1)}(\mathbf{0}_0^{i-1}, 0)$ and $\mathbf{x}_2 \in \{0, 1\}^{N/2}$. Moreover, for $t \in \{0, \dots, N/2 - w'\}$, the contribution of \mathbf{x}_1 with $\operatorname{wt}(\mathbf{x}_1) = w'$ to $T_{i,w'+2t}^{(n)}$ is again $2^{w'} \binom{N/2-w'}{t}$.

Remark 2. Algorithm 1 provides a practical way to determine WDs associated with SC decoding. Let us take some $p \geq 1$ and assume that we have a precomputed look-up table of appropriate binomial coefficients. Then the complexity of computing the first p nonzero components of $\mathbf{S}_i^{(n)}$ for all $i \in [0, N]$ is $O(p^2 N)$. Indeed, the complexity of the for-loops on lines 6 and 10 can be reduced to $O(p^2)$ and $O(p)$, respectively, if only the first p nonzero components are computed. Now we check this claim for line 6. Suppose that the first nonzero component of $\mathbf{S}_i^{(j-1)}$ corresponds to weight $w_i^{(j-1)}$. By line 7, the weights of the first p nonzero components of $\mathbf{S}_i^{(j)}$ are between $w_i^{(j-1)}$ and $w_i^{(j-1)} + 2p - 2$ and to compute one of them we need to sum at most p numbers. Hence, the loop on line 6 has complexity $O(p^2)$. Similar arguments work out for line 11. The total number of required computations of the algorithm is also linear with the number of times when we are in the for-loop on line 4, which is equal to $2N - 2$. This implies that for $p = O(1)$, the complexity of getting approximate upper bounds on the error probabilities for all the subchannels based on the formula (4), restricted to the first p nonzero terms, is $O(N)$.

Algorithm 1 Computing the weight distributions

Input: length $N = 2^n$
Output: weight distributions $\{S_i^{(n)} : i \in [0, N)\}$
Initialization :

```

1:  $S_0^{(0)} \leftarrow (S_{0,0}^{(0)}, S_{0,1}^{(0)})$  with  $S_{0,0}^{(0)} \leftarrow 0$  and  $S_{0,1}^{(0)} \leftarrow 1$ 
2: for Kronecker's power  $j = 1$  to  $n$  do
3:    $N_j \leftarrow 2^j$ 
4:   for row index  $i = 0$  to  $N_j - 1$  do
5:     if  $(i < N_j/2)$  then
6:       for weight  $w = 0$  to  $N_j$  do
7:          $S_{i,w}^{(j)} \leftarrow \sum_{\substack{0 \leq w' \leq \min(w, N_j/2) \\ w' \equiv w \pmod{2}}} S_{i,w'}^{(j-1)} \binom{N_j/2 - w'}{(w - w')/2} 2^{w'}$ 
8:       end for
9:     else
10:      for weight  $w = 0$  to  $N_j/2$  do
11:         $S_{i,2w}^{(j)} \leftarrow S_{i - N_j/2, w}^{(j-1)}$ 
12:         $S_{i,2w+1}^{(j)} \leftarrow 0$ 
13:      end for
14:    end if
15:  end for
16: return  $\{S_i^{(n)} : i \in [0, N)\}$ 

```

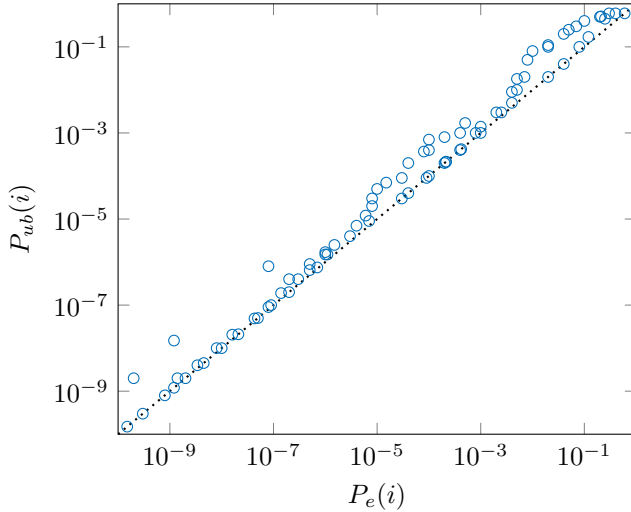


Fig. 1. The approximate union bound $P_{ub}(i)$ and the bit-channel probabilities of error $P_e(i)$. Only those pairs $(P_e(i), P_{ub}(i))$ for which the error probability $P_e(i)$ is greater than 10^{-10} threshold are shown. Additionally, the dashed line (x, x) for $x \in [10^{-10}, 10^0]$ is plotted.

Example 2. Let us illustrate the bound (4) by taking code length $N = 256$ and the AWGN channel with variance $\sigma^2 = 0.158$ (SNR is around 8.01 dB). Using Algorithm 1 we compute the weight distributions and depict the pairs $(P_e(i), P_{ub}(i))$ in Figure 1, where the approximate union bound $P_{ub}(i)$ on the decoding error probability $P_e(i)$ is computed with the help of (4). We compute $P_e(i)$ with the help of DE.

Example 3. Now let us take the code length $N = 128$ and the synthetic channel with index $i = 72$. Using Algorithm 1 we compute the weight distributions and depict $P_{ub}(72)$ the

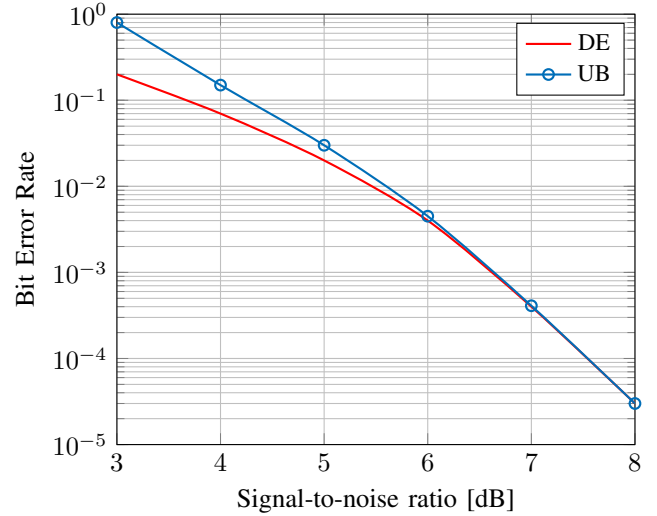


Fig. 2. The DE computations and the approximate union bound on $P_e(72)$ under SC decoding.

approximate union bound (4) on the decoding error probability $P_e(72)$ along with this probability, calculated with the help of DE, in Figure 2.

B. First Nonzero Component

Let $s_i^{(n)}$ be the first nonzero component of $S_i^{(n)}$. It is known (e.g., see [6, Chapter 6]) that the first nonzero component of $S_i^{(n)}$ corresponds to weight $\text{wt}(\mathbf{g}_i)$ and, thus, $s_i^{(n)} = S_{i, \text{wt}(\mathbf{g}_i)}^{(n)} > 0$, i.e.,

$$S_i^{(n)} = \underbrace{(0, 0, \dots, 0)}_{\text{wt}(\mathbf{g}_i)} (s_i^{(n)}, \dots).$$

We extend this line of research and find an explicit formula for $s_i^{(n)}$. Given $j \in [0, n)$ and $i \in [0, N)$, let $b_j(i)$ be the j th bit in the binary representation of integer i , and $p_j(i)$ be the partial sum of the first $j + 1$ bits, i.e., integers i and $p_j(i)$ can be represented as

$$i = \sum_{j=0}^{n-1} b_j(i) 2^j, \quad p_j(i) = \sum_{s=0}^j b_s(i).$$

Theorem 2. Given $N = 2^n$ and $i \in [0, N)$, the first nonzero component of $S_i^{(n)}$ corresponds to weight $\text{wt}(\mathbf{g}_i) = 2^{p_{n-1}(i)}$ and equals $s_i^{(n)}$, where

$$\log_2 s_i^{(n)} = \sum_{j=0}^{n-1} (1 - b_j(i)) 2^{p_j(i)}. \quad (7)$$

Proof of Theorem 2. We shall prove the statement of this theorem by induction on n . The base case is evident as $S_{0,1}^{(0)} = 1$. We assume for a moment that the equality (7) holds for some pair (n, i) with $i \in [0, N)$. Let \mathbf{g}_i be the i th row in the matrix G_N . We derive the statement of this theorem for $(n + 1, i)$ and $(n + 1, i + N)$.

First, let us prove the case $(n+1, i)$. By line 7 of Algorithm 1, for $i < N = 2^n$, we have

$$S_{i,w}^{(n+1)} := \sum_{\substack{0 \leq w' \leq \min(w, N) \\ w' \equiv w \pmod{2}}} S_{i,w'}^{(n)} \binom{N-w'}{(w-w')/2} 2^{w'}. \quad (8)$$

Therefore, we get that $S_{i,w}^{(n+1)} = 0$ for all $w < \text{wt}(\mathbf{g}_i)$ since $S_{i,w'}^{(n)} = 0$ for all $w' < \text{wt}(\mathbf{g}_i)$. For $w = \text{wt}(\mathbf{g}_i)$, the nonzero contribution to $S_{i,w}^{(n+1)}$ in the right-hand side of (8) comes from the only term indexed by $w' = w$. Thus, we have $s_i^{(n+1)} = s_i^{(n)} 2^{\text{wt}(\mathbf{g}_i)}$ and since $b_n(i) = 0$ so that $2^{p_n(i)} = 2^{p_{n-1}(i)} = \text{wt}(\mathbf{g}_i)$,

$$\begin{aligned} \log_2 s_i^{(n+1)} &= \log_2 s_i^{(n)} + \text{wt}(\mathbf{g}_i) \\ &= \sum_{j=0}^{n-1} (1 - b_j(i)) 2^{p_j(i)} + \text{wt}(\mathbf{g}_i) \\ &= \sum_{j=0}^n (1 - b_j(i)) 2^{p_j(i)}. \end{aligned}$$

Similarly, we deal with the case $(n+1, i+N)$. Let $i' := i + N \geq N$. According to line 11 of Algorithm 1, we have

$$S_{i',2w}^{(n+1)} = S_{i,w}^{(n)}, \quad S_{i',2w+1}^{(n+1)} = 0. \quad (9)$$

We obtain $S_{i',w}^{(n+1)} = 0$ for all $w < 2\text{wt}(\mathbf{g}_i)$ since $S_{i,w}^{(n)} = 0$ for all $w < \text{wt}(\mathbf{g}_i)$. Thus, the first nonzero component of $\mathbf{S}_{i'}^{(n+1)}$ corresponds to the weight $2\text{wt}(\mathbf{g}_i) = \text{wt}(\mathbf{g}_{i'})$, where $\mathbf{g}_{i'}$ is the i' 'th row in the matrix

$$G_{2N} = \begin{pmatrix} G_N & 0 \\ G_N & G_N \end{pmatrix}.$$

Moreover, by (9) we have $s_{i'}^{(n+1)} = s_i^{(n)}$ and since $b_n(i') = 1$

$$\begin{aligned} \log_2 s_{i'}^{(n+1)} &= \log_2 s_i^{(n)} \\ &= \sum_{j=0}^{n-1} (1 - b_j(i)) 2^{p_j(i)} \\ &= \sum_{j=0}^n (1 - b_j(i')) 2^{p_j(i')}. \end{aligned}$$

This proves the inductive step and completes the proof. \square

C. Connection with the Partial Order

It was observed [12], [13] that there is a partial order between the synthetic channels, which holds for any B-MSC. Let us rephrase this result using our notation.

Theorem 3 (The partial order [12, Definition 8]). $W_N^{(i)}$ is stochastically degraded by $W_N^{(j)}$ if there exists a finite sequence of integers $a_0, a_1, \dots, a_\ell \in [0, N)$, $\ell \geq 0$, such that $a_0 = i$, $a_\ell = j$ and for all $m \in [0, \ell)$, one of the following two properties holds:

- 1) There exist two indices $u = u(m), w = w(m) \in [0, n)$ so that $u < w$ and

$$\begin{aligned} b_k(a_m) &= b_k(a_{m+1}) \text{ for all } k \in [0, n) \setminus \{u, w\}, \\ b_u(a_m) &= b_w(a_{m+1}) = 1, \\ b_w(a_m) &= b_u(a_{m+1}) = 0; \end{aligned}$$

- 2) $b_k(a_m) \leq b_k(a_{m+1})$ for all $k \in [0, n)$.

Note that for any such sequence a_0, a_1, \dots, a_ℓ , we have the property

$$\sum_{k=n-1-t}^{n-1} b_k(a_m) \leq \sum_{k=n-1-t}^{n-1} b_k(a_{m+1})$$

for all $t \in [0, n)$, $m \in [0, \ell)$. In particular, this means

$$\sum_{k=n-1-t}^{n-1} b_k(i) \leq \sum_{k=n-1-t}^{n-1} b_k(j) \text{ for all } t \in [0, n). \quad (10)$$

Theorem 2 shows us a natural one-way connection between the partial order given in Theorem 3 and the first nonzero components of WDs. Let $i < j$ be fixed. If the i 'th synthetic channel is stochastically degraded by the j 'th one by the partial order, then the property (10) holds. In particular, this implies that $p_{n-1}(i) \leq p_{n-1}(j)$. In case the latter inequality is strict, by Theorem 2, we deduce that $\text{wt}(\mathbf{g}_i) < \text{wt}(\mathbf{g}_j)$ and the first nonzero component of $\mathbf{S}_j^{(n)}$ corresponds to a larger weight than the first nonzero component of $\mathbf{S}_i^{(n)}$:

$$\begin{aligned} \mathbf{S}_i^{(n)} &= (\underbrace{0, \dots, 0}_{\text{wt}(\mathbf{g}_i)}, s_i^{(n)}, \dots), \\ \mathbf{S}_j^{(n)} &= (\underbrace{0, \dots, 0}_{\text{wt}(\mathbf{g}_i)}, \underbrace{0, \dots, 0}_{\text{wt}(\mathbf{g}_j) - \text{wt}(\mathbf{g}_i)}, s_j^{(n)}, \dots). \end{aligned}$$

In case $p_{n-1}(i) = p_{n-1}(j)$, we have $\text{wt}(\mathbf{g}_i) = \text{wt}(\mathbf{g}_j)$ and the inequality (10) can be rewritten in a simple way

$$\begin{aligned} p_t(i) &= p_{n-1}(i) - \sum_{k=t+1}^{n-1} b_k(i) \\ &\geq p_{n-1}(j) - \sum_{k=t+1}^{n-1} b_k(j) = p_t(j) \end{aligned}$$

for all $t \in [0, n)$. Now we shall prove that $s_i^{(n)} > s_j^{(n)}$. To this end, it is enough to show that for a_m and a_{m+1} that fulfill the first property in Theorem 3, we have $s_{a_m}^{(n)} > s_{a_{m+1}}^{(n)}$. Indeed, if two distinct integers a_m and a_{m+1} satisfy the second property in Theorem 3, then

$$p_{n-1}(i) \leq p_{n-1}(a_m) < p_{n-1}(a_{m+1}) = p_{n-1}(j).$$

By Theorem 2, we know

$$\begin{aligned} \log_2 s_{a_m}^{(n)} &= \sum_{k=0}^{n-1} (1 - b_k(a_m)) 2^{p_k(a_m)}, \\ \log_2 s_{a_{m+1}}^{(n)} &= \sum_{k=0}^{n-1} (1 - b_k(a_{m+1})) 2^{p_k(a_{m+1})}. \end{aligned}$$

By the first property in Theorem 3, $b_k(a_m) = b_k(a_{m+1})$ for all $k \in [0, n) \setminus \{u, w\}$ with $u < w$ and $b_u(a_m) = b_w(a_{m+1}) = 1$ and $b_w(a_m) = b_u(a_{m+1}) = 0$. Thus, we obtain

$$\begin{aligned} \log_2 s_{a_m}^{(n)} - \log_2 s_{a_{m+1}}^{(n)} &= \sum_{k=u+1}^w (1 - b_k(a_m)) 2^{p_k(a_m)} \\ &\quad - \sum_{k=u}^{w-1} (1 - b_k(a_{m+1})) 2^{p_k(a_{m+1})} \\ &= 2^{p_w(a_m)} - 2^{p_u(a_{m+1})} \\ &\quad + \sum_{k=u+1}^{w-1} (1 - b_k(a_m)) (2^{p_k(a_m)} - 2^{p_k(a_{m+1})}) > 0, \end{aligned}$$

where we used the property that $p_k(a_m) = p_k(a_{m+1}) + 1$ for $k \in [u+1, w)$ and $p_w(a_m) \geq p_u(a_m) = p_u(a_{m+1}) + 1$. Combining the arguments above we arrive to the following statement.

Theorem 4. Let $W_N^{(i)}$ be stochastically degraded by $W_N^{(j)}$ with $i < j$ by Theorem 3. Then one of the two statements holds:

- 1) The first nonzero component of $S_j^{(n)}$ corresponds to a larger weight than the first nonzero component of $S_i^{(n)}$, i.e., $\text{wt}(\mathbf{g}_j) > \text{wt}(\mathbf{g}_i)$.
- 2) The first nonzero component of $S_j^{(n)}$ corresponds to the same weight as the first nonzero component of $S_i^{(n)}$, i.e., $w := \text{wt}(\mathbf{g}_j) = \text{wt}(\mathbf{g}_i)$, and we have $s_i^{(n)} > s_j^{(n)}$.

IV. TOWARD WEIGHT DISTRIBUTION FOR SUCCESSIVE CANCELLATION LIST DECODING

Let us briefly recall the high level description of the successive cancellation list (SCL) decoder [8] with list size L . At the i th decoding stage for $i \in \mathcal{A}$, we split each path $\hat{\mathbf{u}}_0^{i-1}$ from the list of candidates, abbreviated by \mathcal{L} , into two paths by taking $\hat{u}_i = 0$ and $\hat{u}_i = 1$ and calculate two values $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|0)$ and $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|1)$ by (3). Since the number of paths is doubled, we keep in \mathcal{L} only the L most likely paths at each stage. The pruning criterion is based on the values

$$\left\{ W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|0), W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|1) \right\}_{\hat{\mathbf{u}}_0^{i-1} \in \mathcal{L}}.$$

If index $i \in \mathcal{A}^c$, then for any path $\hat{\mathbf{u}}_0^{i-1}$ from \mathcal{L} , the decoder makes a bit decision $\hat{u}_i = 0$ and keeps this $\hat{\mathbf{u}}_0^i$ in \mathcal{L} .

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 \mathbf{u}_0^j mistaken in only two positions i and j , $i < j$, i.e., $\text{supp}(\mathbf{u}_0^j) = \{i, j\}$. Our goal is to estimate the minimal distance between sets induced by these two paths. For simplicity of notation we abbreviate $C^{(n)}(\mathbf{u}_0^j)$ by $C^{(n)}(i, j)$. Also, recall that

$$C^{(n)}(i, j) = \mathbf{g}_i + \mathbf{g}_j + \langle \mathbf{g}_{j+1}, \dots, \mathbf{g}_{N-1} \rangle.$$

Theorem 5. Let $N = 2^n$. For any $i, j \in [0, N)$ such that $i < j$ and any $\mathbf{x} \in C^{(n)}(i, j)$, we have $\text{wt}(\mathbf{x}) \geq \text{wt}(\mathbf{g}_i + \mathbf{g}_j)$, where

$$\text{wt}(\mathbf{g}_i + \mathbf{g}_j) = \text{wt}(\mathbf{g}_i) + \text{wt}(\mathbf{g}_j) - 2^{t_{i,j}+1} \quad (11)$$

with $t_{i,j} := \sum_{k=0}^{n-1} b_k(i)b_k(j)$.

Before we start proving the theorem, let us introduce several useful definitions. Recall the notion $b_r(k)$, the r th bit in the binary representation of integer k . Let $\bar{\mathbf{g}}_0 := (1, 0)$ and $\bar{\mathbf{g}}_1 := (1, 1)$. In other words, the first row of G_2 is $\bar{\mathbf{g}}_0$ and the second one is $\bar{\mathbf{g}}_1$. By induction, any row \mathbf{g}_k in the matrix $G_N = G_2^{\otimes n}$ can be shown to be a Kronecker product of vectors

$$\mathbf{g}_k = \bar{\mathbf{g}}_{b_{n-1}(k)} \otimes \bar{\mathbf{g}}_{b_{n-2}(k)} \otimes \dots \otimes \bar{\mathbf{g}}_{b_0(k)}. \quad (12)$$

Let ℓ be an arbitrary integer such that $0 \leq \ell < n$. By $I_\ell \subset [0, N)$ denote the collection of indices $k \in [0, N)$ such that $b_\ell(k) = 1$. Define the complement of I_ℓ by $I_\ell^c := [0, N) \setminus I_\ell$. We note that $|I_\ell| = |I_\ell^c| = N/2$.

Define the function $f_\ell : [0, N) \rightarrow [0, N/2)$ that maps an integer $k \in [0, N)$ to the integer $f_\ell(k)$ by removing the ℓ th bit in the binary representation of k

$$f_\ell(k) := \sum_{r=0}^{\ell-1} b_r(k) 2^r + \sum_{r=\ell+1}^{n-1} b_r(k) 2^{r-1}.$$

Note that in the second sum, $b_r(k)$ is multiplied by 2^{r-1} and not 2^r . Moreover, in the special case where $\ell = n-1$, we have $f_{n-1}(k) = k - b_{n-1}(k) 2^{n-1}$.

Let vector \mathbf{x} be of length N and let $I \subset [0, N)$ be a set of indices in ascending order. Then, $\mathbf{x}|_I$ is a projection of \mathbf{x} onto the coordinates indexed by I . That is, $\mathbf{x}|_I$ is a vector of length $|I|$ in which any index of \mathbf{x} that is not in I is removed. For example, if $N = 8$, $\mathbf{x} = (0, 0, 1, 0, 1, 1, 0, 1)$, and $I = \{1, 3, 5, 7\}$, then $\mathbf{x}|_I = (0, 0, 1, 1)$.

Denote $Q := G_{N/2}$ with rows $\{\mathbf{q}_k, k \in [0, N/2)\}$. From the structure (12) and the definition of I_ℓ^c , we note that

$$\begin{aligned} \mathbf{g}_k|_{I_\ell^c} &= \bar{\mathbf{g}}_{b_{n-1}(k)} \otimes \dots \otimes \bar{\mathbf{g}}_{b_{\ell+1}(k)} \otimes 1 \otimes \dots \otimes \bar{\mathbf{g}}_{b_0(k)} \\ &= \mathbf{q}_{f_\ell(k)}. \end{aligned} \quad (13)$$

Similarly, from the structure (12) and the definition of I_ℓ , we get that

$$\mathbf{g}_k|_{I_\ell} = \begin{cases} \mathbf{0}_0^{N/2-1} & \text{if } b_\ell(k) = 0, \\ \mathbf{q}_{f_\ell(k)} & \text{if } b_\ell(k) = 1. \end{cases} \quad (14)$$

Observe that by (13)-(14), $\text{wt}(\mathbf{g}_k|_{I_\ell^c}) = \text{wt}(\mathbf{g}_k)$ if $b_\ell(k) = 0$.

The following example demonstrates the definitions and concepts introduced above.

Example 4. Let $N = 8$, $\ell = 0$, $i = 2 = 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$ and $j = 5 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$. Then the vectors \mathbf{g}_i and \mathbf{g}_j can be represented using (12) as

$$\begin{aligned} \mathbf{g}_i &= \bar{\mathbf{g}}_0 \otimes \bar{\mathbf{g}}_1 \otimes \bar{\mathbf{g}}_0 = (1, 0, 1, 0, 0, 0, 0, 0), \\ \mathbf{g}_j &= \bar{\mathbf{g}}_1 \otimes \bar{\mathbf{g}}_0 \otimes \bar{\mathbf{g}}_1 = (1, 1, 0, 0, 1, 1, 0, 0). \end{aligned}$$

Similarly, we have that $\mathbf{q}_1 = \bar{\mathbf{g}}_0 \otimes \bar{\mathbf{g}}_1 = (1, 1, 0, 0)$ and $\mathbf{q}_2 = \bar{\mathbf{g}}_1 \otimes \bar{\mathbf{g}}_0 = (1, 0, 1, 0)$. The set $I_\ell = \{1, 3, 5, 7\}$ consists of all integers in $[0, N)$ whose ℓ th bit (least significant bit in this case) in the binary representation is one (1). The set $I_\ell^c = [0, 8) \setminus I_\ell = \{0, 2, 4, 6\}$. One can see that

$$\begin{aligned} f_\ell(0) &= f_\ell(1) = 0, & f_\ell(2) &= f_\ell(3) = 1, \\ f_\ell(4) &= f_\ell(5) = 2, & f_\ell(6) &= f_\ell(7) = 3. \end{aligned}$$

TABLE II
EXAMPLE DEMONSTRATING LEMMA 6

(i, j)	(0, 1)	(0, 2)	(0, 3)	(1, 3)	(2, 3)
base two	(00, 01)	(00, 10)	(00, 11)	(01, 11)	(10, 11)
ℓ	1	0	0	0	1
n_0	0	1	1	1	0

By (13) and (14) the restrictions of \mathbf{g}_i and \mathbf{g}_j to coordinates indexed by I_ℓ^c and I_ℓ are the vectors

$$\begin{aligned}\mathbf{g}_i|_{I_\ell^c} &= (1, 0) \otimes (1, 1) \otimes (1) = \mathbf{q}_{f_\ell(i)} = \mathbf{q}_1, \\ \mathbf{g}_i|_{I_\ell} &= (1, 0) \otimes (1, 1) \otimes (0) = \mathbf{0}_0^3, \\ \mathbf{g}_j|_{I_\ell} &= \mathbf{g}_j|_{I_\ell^c} = (1, 1) \otimes (1, 0) \otimes (1) = \mathbf{q}_{f_\ell(j)} = \mathbf{q}_2.\end{aligned}$$

Lemma 6. Given $i, j \in [0, N)$ with $i < j$ and $N = 2^n$, let $n_0 \in [0, n)$ be the largest index such that $b_{n_0}(i) = 0$ and $b_{n_0}(j) = 1$. Then one of the two statements holds:

- 1) There exists $\ell \in [0, n) \setminus \{n_0\}$ such that $b_\ell(i) \leq b_\ell(j)$ and $f_\ell(i) < f_\ell(j)$.
- 2) $i = N/2 - 1$ and $j = N/2$.

Proof of Lemma 6. If $n_0 < n - 1$, we can take $\ell = n - 1$ and the first property on the lemma holds. Indeed, $b_\ell(i) = b_\ell(j)$ and

$$f_\ell(i) = i - b_{n-1}(i)2^{n-1} < j - b_{n-1}(j)2^{n-1} = f_\ell(j).$$

Now assume $n_0 = n - 1$. If the second statement holds, we are done. Otherwise, there must exist $\ell < n - 1$ for which $b_\ell(i) \leq b_\ell(j)$. In this case, we have $f_\ell(i) < 2^{n-2} \leq f_\ell(j)$. \square

Let us illustrate how the above lemma works.

Example 5. Let $N = 4$. We provide Table II showing the integer n_0 from Lemma 6 and a possible choice of ℓ for a given pair (i, j) . For the case $(i, j) = (1, 2)$, we have the second property in Lemma 6.

Lemma 7. Given $j, k \in [0, N)$ with $j < k$, let ℓ be an arbitrary integer from $[0, n)$ such that $b_\ell(j) \geq b_\ell(k)$. Then $f_\ell(j) < f_\ell(k)$.

Proof of Lemma 7. Since $j < k$, for the largest integer n_0 so that $b_{n_0}(j) \neq b_{n_0}(k)$, we have $b_{n_0}(j) = 0 < 1 = b_{n_0}(k)$. Thus, necessarily $\ell \neq n_0$. Let n'_0 be the largest integer such that $b_{n'_0}(f_\ell(j)) \neq b_{n'_0}(f_\ell(k))$. If $\ell < n_0$ then $n'_0 = n_0 - 1$, and if $\ell > n_0$ then $n'_0 = n_0$. In either case, $b_{n'_0}(f_\ell(j)) = b_{n_0}(j) = 0 < 1 = b_{n_0}(k) = b_{n'_0}(f_\ell(k))$ and, thus, $f_\ell(j) < f_\ell(k)$. \square

Proof of Theorem 5. We shall prove the inequality $\text{wt}(\mathbf{x}) \geq \text{wt}(\mathbf{g}_i + \mathbf{g}_j)$ for any $\mathbf{x} \in C^{(n)}(i, j)$ by induction on n , where $N = 2^n$. Showing (11) is postponed to the end of the proof. The base case $n = 1$ is obviously true as $i, j \in [0, 2)$ and condition $i < j$ can be satisfied only for $i = 0, j = 1$. Thus, the set $C^{(1)}(i, j)$ contains only the single element $\mathbf{g}_0 + \mathbf{g}_1$, and, hence, its minimal weight equals $\text{wt}(\mathbf{g}_0 + \mathbf{g}_1)$. Assume that the statement holds for $n - 1$. We prove that it is true for n .

Set $N = 2^n$ and let $i, j \in [0, N)$ such that $i < j$. Any binary vector $\mathbf{x} \in C^{(n)}(i, j)$ can be determined by coefficients $(\alpha_k)_{k=j+1}^{N-1} \in \{0, 1\}^{N-j-1}$ in the following way

$$\mathbf{x} = \mathbf{g}_i + \mathbf{g}_j + \sum_{k \in [j+1, N)} \alpha_k \mathbf{g}_k.$$

By Lemma 6, we can have two possibilities. If $i = N/2 - 1$, then, by the structure of G_N and since $j > i$, $\text{wt}(\mathbf{g}_i + \mathbf{g}_j) = \text{wt}(\mathbf{g}_i) = N/2$. As in (5)-(6), $\text{wt}(\mathbf{x}) \geq \text{wt}(\mathbf{g}_i) = N/2$. Thus, in this case the statement is true. Otherwise, $i \neq N/2 - 1$ and the first statement of Lemma 6 must hold. Namely, there exists $\ell \in [0, n)$ such that $b_\ell(i) \leq b_\ell(j)$ and $f_\ell(i) < f_\ell(j)$. Define

$$\begin{aligned}\mathbf{x}_0 &:= \mathbf{g}_i \mathbb{1}_0(b_\ell(i)) + \mathbf{g}_j \mathbb{1}_0(b_\ell(j)) + \sum_{k \in [j+1, N)} \alpha_k \mathbf{g}_k \mathbb{1}_0(b_\ell(k)), \\ \mathbf{x}_1 &:= \mathbf{g}_i \mathbb{1}_1(b_\ell(i)) + \mathbf{g}_j \mathbb{1}_1(b_\ell(j)) + \sum_{k \in [j+1, N)} \alpha_k \mathbf{g}_k \mathbb{1}_1(b_\ell(k)),\end{aligned}\tag{15}$$

where the indicator function $\mathbb{1}_c(x) = 1$ if $x = c$ and $\mathbb{1}_c(x) = 0$ if $x \neq c$. It is readily seen that $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1$. From (13)-(14) we note that

$$\mathbf{x}_0|_{I_\ell} = \mathbf{0}_0^{N/2-1}, \quad \mathbf{x}_1|_{I_\ell} = \mathbf{x}_1|_{I_\ell^c}.$$

From this it follows

$$\begin{aligned}\text{wt}(\mathbf{x}) &= \text{wt}(\mathbf{x}|_{I_\ell}) + \text{wt}(\mathbf{x}|_{I_\ell^c}) \\ &= \text{wt}(\mathbf{x}_1|_{I_\ell}) + \text{wt}(\mathbf{x}|_{I_\ell^c}) \\ &= \text{wt}(\mathbf{x}_1|_{I_\ell}) + \text{wt}(\mathbf{x}_0|_{I_\ell^c} + \mathbf{x}_1|_{I_\ell}).\end{aligned}\tag{16}$$

Recall that $b_\ell(i) \leq b_\ell(j)$ and $f_\ell(i) < f_\ell(j)$. We distinguish the further analysis between two cases: $b_\ell(j) = 0$ and $b_\ell(j) = 1$.

Case 1: $b_\ell(j) = 0$. It follows that $b_\ell(i) = 0$. By applying (13) to \mathbf{x}_0 from (15), we get

$$\mathbf{x}_0|_{I_\ell^c} = \mathbf{q}_{f_\ell(i)} + \mathbf{q}_{f_\ell(j)} + \sum_{k \in [j+1, N)} \alpha_k \mathbf{q}_{f_\ell(k)} \mathbb{1}_0(b_\ell(k)).$$

By Lemma 7, $f_\ell(j) < f_\ell(k)$ for k such that $k > j$ and $b_\ell(k) = b_\ell(j) = 0$, and, thus, $\mathbf{x}_0|_{I_\ell^c} \in C^{(n-1)}(f_\ell(i), f_\ell(j))$. Therefore, we are able to apply the inductive assumption for $\mathbf{x}_0|_{I_\ell^c}$ and obtain that

$$\begin{aligned}\text{wt}(\mathbf{x}_0|_{I_\ell^c}) &\geq \text{wt}(\mathbf{q}_{f_\ell(i)} + \mathbf{q}_{f_\ell(j)}) \\ &= \text{wt}((\mathbf{g}_i + \mathbf{g}_j)|_{I_\ell^c}) \\ &= \text{wt}(\mathbf{g}_i + \mathbf{g}_j),\end{aligned}\tag{17}$$

where the first equality follows from (13) and the right-most one is implied by (14) and $b_\ell(i) = b_\ell(j) = 0$. Finally, by (16) we conclude with

$$\begin{aligned}\text{wt}(\mathbf{x}) &= \text{wt}(\mathbf{x}_1|_{I_\ell}) + \text{wt}(\mathbf{x}_0|_{I_\ell^c} + \mathbf{x}_1|_{I_\ell}) \\ &\geq \text{wt}(\mathbf{x}_0|_{I_\ell^c}) \\ &\geq \text{wt}(\mathbf{g}_i + \mathbf{g}_j),\end{aligned}$$

where the first inequality is an obvious observation that the weight of $(\mathbf{x} + \mathbf{y}, \mathbf{y})$ is at least $\text{wt}(\mathbf{x})$ (see similar arguments in (5)-(6)), and the right-most one is implied by (17).

Case 2: $b_\ell(j) = 1$. By applying (13) to \mathbf{x} , we derive

$$\mathbf{x}|_{I_\ell^c} = \mathbf{q}_{f_\ell(i)} + \mathbf{q}_{f_\ell(j)} + \sum_{k \in [j+1, N]} \alpha_k \mathbf{q}_{f_\ell(k)}.$$

By Lemma 7 and since $b_\ell(j) = 1 \geq b_\ell(k)$ for any k , we obtain that $f_\ell(j) < f_\ell(k)$ for $k > j$ and, thus, $\mathbf{x}|_{I_\ell^c} \in C^{(n-1)}(f_\ell(i), f_\ell(j))$. Then we apply the inductive assumption for $\mathbf{x}|_{I_\ell^c}$ and derive that

$$\text{wt}(\mathbf{x}|_{I_\ell^c}) \geq \text{wt}(\mathbf{q}_{f_\ell(i)} + \mathbf{q}_{f_\ell(j)}). \quad (18)$$

Now we take a look on the weight of $\mathbf{x}_1|_{I_\ell}$, where \mathbf{x}_1 is defined in (15). By (14) we have

$$\begin{aligned} \mathbf{x}_1|_{I_\ell} &= \mathbf{q}_{f_\ell(i)} \mathbb{1}_1(b_\ell(i)) + \mathbf{q}_{f_\ell(j)} \\ &+ \sum_{k \in [j+1, N]} \alpha_k \mathbf{q}_{f_\ell(k)} \mathbb{1}_1(b_\ell(k)). \end{aligned}$$

For $b_\ell(i) = 0$, we obtain $\mathbf{x}_1|_{I_\ell} \in C^{(n-1)}(\mathbf{0}_0^{f_\ell(j)-1}, 1)$ and by Theorem 2, we have

$$\text{wt}(\mathbf{x}_1|_{I_\ell}) \geq \text{wt}(\mathbf{q}_{f_\ell(j)}) = \text{wt}(\mathbf{q}_{f_\ell(i)} \mathbb{1}_1(b_\ell(i)) + \mathbf{q}_{f_\ell(j)}). \quad (19)$$

For $b_\ell(i) = 1$, we get $\mathbf{x}_1|_{I_\ell} \in C^{(n-1)}(f_\ell(i), f_\ell(j))$ and

$$\begin{aligned} \text{wt}(\mathbf{x}_1|_{I_\ell}) &\geq \text{wt}(\mathbf{q}_{f_\ell(i)} + \mathbf{q}_{f_\ell(j)}) \\ &= \text{wt}(\mathbf{q}_{f_\ell(i)} \mathbb{1}_1(b_\ell(i)) + \mathbf{q}_{f_\ell(j)}). \end{aligned} \quad (20)$$

Finally by (16) we conclude with

$$\begin{aligned} \text{wt}(\mathbf{x}) &= \text{wt}(\mathbf{x}_1|_{I_\ell}) + \text{wt}(\mathbf{x}|_{I_\ell^c}) \\ &\geq \text{wt}(\mathbf{q}_{f_\ell(i)} + \mathbf{q}_{f_\ell(j)}) + \text{wt}(\mathbf{q}_{f_\ell(i)} \mathbb{1}_1(b_\ell(i)) + \mathbf{q}_{f_\ell(j)}) \\ &= \text{wt}(\mathbf{g}_i + \mathbf{g}_j), \end{aligned}$$

where the inequality is due to (18)-(20), and the right-most equality is implied by (13)-(14). This completes the proof of cases 1 and 2, and we are now ready to tackle (11).

For any i and j , we have that $\text{wt}(\mathbf{g}_i + \mathbf{g}_j) = \text{wt}(\mathbf{g}_i) + \text{wt}(\mathbf{g}_j) - 2T_{i,j}$, where $T_{i,j}$ denotes the number of positions at which \mathbf{g}_i and \mathbf{g}_j are both one, i.e. $T_{i,j} := |\text{supp}(\mathbf{g}_i) \cap \text{supp}(\mathbf{g}_j)|$. Thus, to verify (11), it suffices to check that $T_{i,j} = 2^{t_{i,j}}$ for any i and j . We shall check this equality by induction on n .

For the base case $n = 1$, we have either $(i, j) \in \{(0, 0), (0, 1), (1, 0)\}$ with $t_{i,j} = 0$ and $T_{i,j} = 1$, or $(i, j) = (1, 1)$ with $t_{i,j} = 1$ and $T_{i,j} = 2$. Assume the statement holds for $n - 1$. Let us show that it is true for n . By (12), we have that

$$\mathbf{g}_i = \bar{\mathbf{g}}_{b_{n-1}(i)} \otimes \mathbf{q}_{f_{n-1}(i)}, \quad \mathbf{g}_j = \bar{\mathbf{g}}_{b_{n-1}(j)} \otimes \mathbf{q}_{f_{n-1}(j)},$$

where $\bar{\mathbf{g}}_0 = (1, 0)$ and $\bar{\mathbf{g}}_1 = (1, 1)$. This implies the property $|\text{supp}(\mathbf{g}_i) \cap \text{supp}(\mathbf{g}_j)| = 2|\text{supp}(\mathbf{q}_{f_{n-1}(i)}) \cap \text{supp}(\mathbf{q}_{f_{n-1}(j)})|$ if $b_{n-1}(i) = b_{n-1}(j) = 1$, and $|\text{supp}(\mathbf{g}_i) \cap \text{supp}(\mathbf{g}_j)| = |\text{supp}(\mathbf{q}_{f_{n-1}(i)}) \cap \text{supp}(\mathbf{q}_{f_{n-1}(j)})|$ otherwise. By the inductive assumption, we obtain that

$$\begin{aligned} T_{i,j} &= |\text{supp}(\mathbf{g}_i) \cap \text{supp}(\mathbf{g}_j)| \\ &= 2^{b_{n-1}(i)b_{n-1}(j)} |\text{supp}(\mathbf{q}_{f_{n-1}(i)}) \cap \text{supp}(\mathbf{q}_{f_{n-1}(j)})| \\ &= 2^{b_{n-1}(i)b_{n-1}(j)} 2^{t_{f_{n-1}(i), f_{n-1}(j)}} \\ &= 2^{\sum_{k=0}^{n-1} b_k(i)b_k(j)} \\ &= 2^{t_{i,j}}. \end{aligned}$$

This completes the proof. \square

V. CONCLUSION

In this paper, we discuss the exact weight distribution of the coset associated with each synthetic channel $W_N^{(i)}$. Also, we find the minimal distance between cosets associated with paths that differ in two positions for successive cancellation list decoding. This study represents initial steps towards understanding the performance of polar codes under successive cancellation list decoding.

The approximate union bound (4) takes into account only weight distributions of the one coset $C^{(n)}(\mathbf{0}_0^{i-1}, 1)$. The drawback of this approach is evident: for low and medium signal-to-noise ratio, the estimate cannot be tight. Based on Remark 1, the weight distributions of any zero coset $C^{(n)}(\mathbf{0}_0^{i-1}, 0)$ 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 word (and the number of minimal weight words) of a set $C^{(n)}(\mathbf{u}_0^{i-1})$ for an arbitrary \mathbf{u}_0^{i-1} . We believe that such an analysis can be helpful for constructing polar codes under successive cancellation list decoding and estimating the distance spectrum of polar codes.

VI. ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers and the Associate Editor Prof. Ido Tal for their careful reading of the manuscript and their many insightful comments and suggestions which improved both the exposition of the paper and the clarity of the proofs.

REFERENCES

- [1] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [2] G. J. Pottie and D. P. Taylor, "Multilevel codes based on partitioning," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 87–98, 1989.
- [3] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3221–3227, 2012.
- [4] E. Arkan, "A performance comparison of polar codes and reed-muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447–449, 2008.
- [5] H. Herzberg, "On the spectrum of distances of a multilevel code, decoded by a multistage decoder," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1736–1740, 1997.
- [6] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Lausanne, 2009.
- [7] Z. Liu, K. Chen, K. Niu, and Z. He, "Distance spectrum analysis of polar codes," in *2014 IEEE Wireless Commun. Netw. Conf. (WCNC)*. IEEE, 2014, pp. 490–495.
- [8] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.
- [9] M. Valipour and S. Yousefi, "On probabilistic weight distribution of polar codes," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2120–2123, 2013.
- [10] Q. Zhang, A. Liu, and X. Pan, "An enhanced probabilistic computation method for the weight distribution of polar codes," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2562–2565, 2017.
- [11] M. P. Fossorier and S. Lin, "Weight distribution for closest coset decoding of $|u|u + v|$ constructed codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 1028–1030, 1997.
- [12] C. Schürch, "A partial order for the synthesized channels of a polar code," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2016, pp. 220–224.
- [13] 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 (ISIT)*, 2016, pp. 230–234.

- [14] F. Hemmati, "Closest coset decoding of $|u|u + v|$ codes," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 982–988, 1989.
- [15] I. M. Jacobs and J. Wozencraft, "Principles of communication engineering," 1965.
- [16] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*. IEEE, 2009, pp. 1496–1500.
- [17] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, 2013.