All the Codeword Symbols in Polar Codes Have the Same SER Under the SC Decoder

Guodong Li*, Min Ye*, and Sihuang Hu*, Member, IEEE

Abstract—Let \( F_p \) be a prime field and let \( F_q \) be a larger finite field obtained from adjoining an element \( \alpha \) to \( F_p \), i.e., \( F_q = F_p(\alpha) \). We consider polar codes constructed from the 2 \( \times \) 2 kernel \( \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \) over \( F_q \). We prove that for any \( F_q \)-symmetric memoryless channel, any code length, and any code dimension, all the codeword symbols in such polar codes have the same symbol error rate (SER) under the successive cancellation (SC) decoder.

Index Terms—Polar code, successive cancellation decoder, symbol error rate, \( F_q \)-symmetric channel.

I. INTRODUCTION

POLAR codes were proposed by Arıkan in [1], and they have been extensively studied over the last decade. In his original paper [1], Arıkan introduced the successive cancellation (SC) decoder to decode polar codes, and he proved that polar codes achieve the capacity of any binary-input memoryless symmetric (BMS) channel under the SC decoder. Later, the successive cancellation list (SCL) decoder and the CRC-aided SCL decoder were proposed to further reduce the decoding error probability of polar codes [2], [3].

Although the CRC-aided SCL decoder provides state-of-the-art performance in terms of the decoding error probability, the SC decoder still receives a lot of research attention due to the following two reasons. First, it is amenable to theoretical analysis. In fact, a large part of the theoretical research on polar codes focuses on the performance of the SC decoder.

Manuscript received 3 November 2022; revised 4 April 2023; accepted 4 April 2023. Date of publication 10 April 2023; date of current version 17 July 2023. Research partially funded by National Key R&D Program of China under Grant No. 2021YFA1001000, National Natural Science Foundation of China under Grant No. 12001322 and 12231014, Shandong Provincial Natural Science Foundation under Grant No. ZR202010220025, Shenzhen Stable Support program under Grant No. WDZC20220811170401001, RISC-V International Open Source Laboratory, and a Taishan scholar program of Shandong Province. The associate editor coordinating the review of this article and approving it for publication was I. Tal. (Corresponding author: Sihuang Hu.)

Guodong Li is with the Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, and the School of Cyber Science and Technology, Shandong University, Qingdao, Shandong 266237, China (e-mail: guodongli@mail.sdu.edu.cn).

Min Ye is with the Tsinghua Shenzhen International Graduate School, Tsinghua-Berkeley Shenzhen Institute, Shenzhen 518055, China (e-mail: yemmni@gmail.com).

Sihuang Hu is with the Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, and the School of Cyber Science and Technology, Shandong University, Qingdao, Shandong 266237, China, and also with the Quan Cheng Laboratory, Jinan 250103, China (e-mail: hoshuang@sdu.edu.cn). Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2023.3265687.

Digital Object Identifier 10.1109/TCOMM.2023.3265687

Second, the running time of the SC decoder is much smaller than the (CRC-aided) SCL decoder. Therefore, SC decoder is the better choice when there is a stringent requirement on the delay of the communication system.

In an early paper [4], Arıkan observed an interesting experimental result regarding the decoding performance of the SC decoder for binary polar codes: The average bit error rate (BER) of a particular subset of codeword coordinates is much smaller than the average BER of the message bits; see Section IV for a detailed explanation of Arıkan’s result. Even till today, there is no rigorous analysis that can explain this phenomenon. As we repeated the experiments in [4], we found another interesting phenomenon—the BER of each codeword coordinates in polar codes is surprisingly stable under the SC decoder. This is not a coincidence. In fact, Theorem 2 of [5] implies that all the codeword coordinates in binary polar codes have the same BER under the SC decoder, although this conclusion was never explicitly mentioned in [5] or any other papers.

In this paper, we extend the above conclusion to polar codes over finite fields. Specifically, we consider polar codes constructed from the 2 \( \times \) 2 kernel \( \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \) over a finite field \( F_q \), where \( q = p^s \) is a power of a prime number \( p \), and \( \alpha \) satisfies \( F_p(\alpha) = F_q \). (\( F_p(\alpha) \) is a standard notation commonly used in abstract algebra and finite fields. Specifically, \( F_p(\alpha) \) is a field extension of the prime field \( F_p \) obtained from adjoining the element \( \alpha \) to \( F_p \).) In other words, \( F_p(\alpha) \) is the smallest finite field containing \( \alpha \) and all the elements in \( F_p \). According to [6, Corollary 15], \( F_p(\alpha) = F_q \) is a necessary and sufficient condition for the constructed codes to polarize. We prove that for any \( F_q \)-symmetric memoryless channel, any code length, and any code dimension, all the codeword symbols in polar codes have exactly the same symbol error rate (SER) under the SC decoder.

The rest of this paper is organized as follows. In Section II, we provide the background on polar codes and state the main result. In Section III, we prove the main result. In Section IV, we discuss the connection between our results and the observation in [4].

II. BACKGROUND AND THE MAIN RESULT

For a set \( \mathcal{A} = \{i_1, i_2, \ldots, i_s\} \) of size \( s \), we use \( x_{A} \) to denote the vector \( (x_{i_1}, x_{i_2}, \ldots, x_{i_s}) \), where we assume that \( i_1 < i_2 < \cdots < i_s \) are nonnegative integers. We use \( x_{[a:b]} \) to denote the vector \( (x_a, x_{a+1}, x_{a+2}, \ldots, x_b) \). Let \( q = p^s \), where \( p \) is a prime number and \( s \) is a positive integer. We say
that a memoryless channel $W : \mathbb{F}_q \rightarrow \mathcal{Y}$ is $\mathbb{F}_q$-symmetric if it satisfies the following two conditions: (1) there exist $q$ permutations $\{\sigma_a : a \in \mathbb{F}_q\}$ on the output alphabet $\mathcal{Y}$ such that $W(y|x) = W(\sigma_x)(x'|x')$ for all $y \in \mathcal{Y}$ and all $x, x' \in \mathbb{F}_q$; (2) there exist $q-1$ permutations $\{\pi_a : a \in \mathbb{F}_q\}$ on $\mathcal{Y}$ such that $W(y|x) = W(\pi_y)(x|x)$ for all $y \in \mathcal{Y}, x \in \mathbb{F}_q$, and $a \in \mathbb{F}_q$, where $\mathbb{F}_q = \mathbb{F}_q \setminus \{0\}$. Both $q$-ary Symmetric Channel (qSC) and $q$-ary Erasure Channel (qEC) are prominent examples that satisfy these two conditions. These two channels are natural generalizations of BSC and BEC, respectively; see [7, pages 117 and 129] for a detailed discussion. Below we use the shorthand notation

$$y + b = \sigma_b(y)$$

and $a \cdot y = \pi_a(y)$ for $a \in \mathbb{F}_q^*$, $b \in \mathbb{F}_q$, $y \in \mathcal{Y}$. \hfill (1)

When both operations appear in the same formula, the operation $\cdot$ has a higher priority than the operation $+$, i.e., $a \cdot y + b = (a \cdot y) + b$.

The construction of polar codes with code length $n = 2^m$ involves the $n \times n$ matrix $G_n = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}^\otimes m$, where $\alpha$ satisfies that $\mathbb{F}_q(\alpha) = \mathbb{F}_q$, and $\otimes$ is the Kronecker product. According to [6, Corollary 15], $G_n(\alpha) = \mathbb{F}_q$ is a necessary and sufficient condition for the polarization of the constructed codes. The matrix $G_n$ serves as a linear mapping between the message vector $u_{[0:n-1]}$ and the codeword vector $x_{[0:n-1]}$. The message vector $u_{[0:n-1]}$ consists of information symbols and frozen symbols. We use $\mathcal{A} \subseteq \{0, 1, \ldots, n - 1\}$ to denote the index set of information symbols and use $\mathcal{A}^c = \{0, 1, \ldots, n - 1\} \setminus \mathcal{A}$ to denote the index set of frozen symbols. A polar code has four parameters—the code length $n$, the code dimension $k$, the index set $\mathcal{A}$ of information symbols, and the vector $\bar{u}_{A^c} \in \mathbb{F}_q^{k-n}$ of frozen symbols.\footnote{We use $\bar{u}$ and its variations to denote the true value. We use $\hat{u}$ and its variations to denote the decoded value.} More precisely, we define

$$\text{Polar}(n, k, \mathcal{A}, \bar{u}_{A^c}) := \{u_{[0:n-1]} G_n : u_{\mathcal{A}^c} = \bar{u}_{A^c}, u_{\mathcal{A}} \in \mathbb{F}_q^k\}.$$

Next let us recall how the SC decoder works. For a symmetric memoryless channel $W : \mathbb{F}_q \rightarrow \mathcal{Y}$, we define $W_n : \mathbb{F}_q^n \rightarrow \mathcal{Y}^n$ as $W_n(y_{[0:n-1]}|x_{[0:n-1]}) = \prod_{i=0}^{n-1} W_i(y_i|x_i)$ for $x_{[0:n-1]} \in \mathbb{F}_q^n$ and $y_{[0:n-1]} \in \mathcal{Y}^n$. For $0 \leq i \leq n - 1$, we further define the symmetric channel $W_i^{(n)} : \mathbb{F}_q \rightarrow \mathcal{Y}^n \times \mathbb{F}_q^n$ as

$$W_i^{(n)}(y_{[0:n-1]}, u_{[0:i-1]}|u_i) = \frac{1}{q^{n-i}} \sum_{u_{[i+1:n-1]} \in \mathbb{F}_q^{n-i-1}} W_n(y_{[0:n-1]}|u_{[0:n-1]} G_n).$$

The SC decoder decodes one by one from $u_0$ to $u_{n-1}$. If $i \in \mathcal{A}^c$, then the decoding result of the $i$th symbol is $\hat{u}_i = \bar{u}_i$. If $i \in \mathcal{A}$, then the decoding result of the $i$th symbol is

$$\hat{u}_i = \hat{u}_i(y_{[0:n-1]}, u_{[0:i-1]}) = \arg\max_{u_i \in \mathbb{F}_q} W_i^{(n)}(y_{[0:n-1]}, u_{[0:i-1]}|u_i).$$ \hfill (2)

We write the elements in $\mathbb{F}_q$ as $\{a_0, a_1, a_2, \ldots, a_{q-1}\}$. If there is a tie, i.e., if

$$W_i^{(n)}(y_{[0:n-1]}, u_{[0:i-1]}|a_{i_0}) = \cdots = W_i^{(n)}(y_{[0:n-1]}, u_{[0:i-1]}|a_{i_{t}}),$$

then $W_i^{(n)}(y_{[0:n-1]}, u_{[0:i-1]}|a_{i_0}) = \cdots = W_i^{(n)}(y_{[0:n-1]}, u_{[0:i-1]}|a_{i_{t}})$ for all $y_{[0:n-1]}$, $u_{[0:i-1]} \in \mathbb{F}_q^n$, and $a_i \in \mathbb{F}_q^*$, where $i_0, i_1, \ldots, i_q$ is a permutation of $\{0, 1, \ldots, q - 1\}$, then the SC decoder outputs a random decoding result with probability \(\mathbb{P}(\hat{u}_i = a_{i_0}) = \cdots = \mathbb{P}(\hat{u}_i = a_{i_{t}}) = 1/s\). The decoding results of the SC decoder depend not only on the channel output vector $y_{[0:n-1]}$, but also on the set $\mathcal{A}$ and the values of the frozen symbols $\bar{u}_{A^c}$. We write the SC decoding result of the message vector as $\hat{x}_{[0:n-1]}(y_{[0:n-1]}, \mathcal{A}, \bar{u}_{A^c})$. The SC decoding result of the codeword vector is

$$\hat{x}_{[0:n-1]}(y_{[0:n-1]}, \mathcal{A}, \bar{u}_{A^c}) = u_{[0:n-1]}(y_{[0:n-1]}, \mathcal{A}, \bar{u}_{A^c}) G_n.$$

For a polar code $\mathcal{C} = \text{Polar}(n, k, \mathcal{A}, \bar{u}_{A^c})$, a symmetric channel $W$, and a specific choice of information vector $\bar{u}_{A} \in \mathbb{F}_q^k$, we write the transmitted codeword as $x_{[0:n-1]} = \bar{x}_{[0:n-1]} G_n$. In this case, the SER of the $j$th codeword symbol under the SC decoder is

$$\text{SER}_j(C, W, \bar{u}_{A}) = \frac{1}{2^k} \sum_{\bar{u}_{A} \in \mathbb{F}_q^k} \text{SER}_j(C, W, \bar{u}_{A}).$$

For an integer $0 \leq i \leq 2^m - 1$, we write its binary expansion as

$$i = \underbrace{2^{m-1}b_{m-1}(i) + 2^{m-2}b_{m-2}(i) + \cdots + 2b_1(i)}_{m-1} + b_0(i),$$

where $b_{m-1}(i), b_{m-2}(i), \ldots, b_1(i), b_0(i) \in \{0, 1\}$. For two integers $0 \leq i, j \leq 2^m - 1$, we say that $i \geq j$ if $b_{m-1}(i) \geq b_{m-1}(j)$ for all $0 \leq r \leq m - 1$. Now we are ready to state our main result.

**Theorem 1:** Let $n = 2^m$ and $k$ be two positive integers satisfying $k \leq n$. Let $\mathcal{A} \subseteq \{0, 1, \ldots, n - 1\}$ be a set of size $|\mathcal{A}| = k$ satisfying the following condition:

$$j \in \mathcal{A} \text{ and } i \geq j, \text{ then } i \in \mathcal{A}. \hfill (5)$$

Let $\bar{u}_{A^c} \in \mathbb{F}_q^{n-k}$ be a $q$-ary vector of length $n - k$. We write $\mathcal{C} = \text{Polar}(n, k, \mathcal{A}, \bar{u}_{A^c})$. Then for any $\mathbb{F}_q$-symmetric memoryless channel $W$, we have

$$\text{SER}_0(C, W) = \text{SER}_1(C, W) = \cdots = \text{SER}_{n-1}(C, W).$$
We need the following lemma to apply Theorem 1 to polar codes.

**Lemma 1:** The set $A$ in polar code construction always satisfies the condition (5).

This lemma was proved in [8], [9], and [10] for binary polar codes, and the extension to polar codes over finite fields is trivial. In fact, [8], [9], [10] proved an even stronger result, showing that binary polar codes are decreasing monomial codes. The properties of decreasing monomial codes were used to reduce the complexity of encoding, decoding, and code construction for polar codes [11], [12].

Note that the condition (5) is necessary for Theorem 1 to hold. Let us consider the simplest case of $n = 2$, $k = 1$, and $q = 2$. The choice of $A = \{1\}$ satisfies the condition (5). Under this choice, $t_0$ is the frozen symbol, so its decoding result $\hat{u}_0$ is equal to the true value $\hat{u}_0$. The decoding results of the codeword symbols are $\hat{x}_0 = \hat{u}_0 + \alpha \hat{u}_1$ and $\hat{x}_1 = \hat{u}_1$. Therefore, the three events $\{\hat{x}_0 \neq \hat{x}_1\}$, $\{\hat{x}_1 \neq \hat{x}_1\}$ and $\{\hat{u}_1 \neq \hat{u}_1\}$ are the same, so the two codeword symbols have the same SER. On the other hand, the choice of $A = \{0\}$ does not satisfy the condition (5). Under this choice, $u_1$ is the frozen symbol, so its decoding result $\hat{u}_1$ is equal to the true value $\hat{u}_1$, which implies that $\hat{x}_1 = \hat{x}_1$. However, the event $\{\hat{x}_0 \neq \hat{x}_1\}$ happens with positive probability unless the channel $W$ is noiseless. Therefore, the two codeword symbols do not have the same SER when the condition (5) is not satisfied.

### III. PROOF OF THE MAIN RESULT

#### A. Restricting to the All-Zero Codeword

By (1), we have $W(y|x) = W(a \cdot y + b \cdot a \cdot x + b)$ for all $a \in \mathbb{F}_q$, $b \in \mathbb{F}_q$, $x \in \mathbb{F}_q$, $y \in \mathcal{Y}$. For $a \in \mathbb{F}_q$ and two vectors $b_{[0:n-1]} \in \mathbb{F}_q$, $y_{[0:n-1]} \in \mathcal{Y}^n$, we define $a \cdot y_{[0:n-1]} + b_{[0:n-1]} = (a \cdot y_0 + b_0, a \cdot y_1 + b_1, \ldots, a \cdot y_{n-1} + b_{n-1})$. Then we have $W^n(y_{[0:n-1]} | x_{[0:n-1]}) = W^n(a \cdot y_{[0:n-1]} + b_{[0:n-1]} | a \cdot x_{[0:n-1]} + b_{[0:n-1]})$ for all $a \in \mathbb{F}_q$ and $b_{[0:n-1]}, x_{[0:n-1]} \in \mathbb{F}_q^n$.

We first prove that the SER of each codeword symbol is independent of the true value $u_{[0:n-1]}$ of the message vector. Note that the proof technique of Lemma 2 below is quite standard in polar coding literature. In fact, this proof technique was already used in Arıkan’s original paper; see Section VI of [1]. The only difference is that the proof in [1] focused on the word error rate while we focus on the SER.

**Lemma 2:** For any $\tilde{u}_{[0:n-1]} \in \mathbb{F}_q^n$, any $\tilde{u}_{[0:n-1]} \in \mathbb{F}_q^n$, and any $0 \leq j \leq n - 1$, we have

$$\text{SER}_j(\text{Polar}(n, k, \tilde{u}, \tilde{u}_A; W, \tilde{u}_A)) = \text{SER}_j(\text{Polar}(n, k, \tilde{u}, \tilde{u}_A; W, \tilde{u}_A)).$$

**Proof:** For any $\tilde{u}_{[0:n-1]} \in \mathbb{F}_q^n$ and $\tilde{u}_{[0:n-1]} \in \mathbb{F}_q^n$, we can always find a pair $a \in \mathbb{F}_q$ and $b_{[0:n-1]} \in \mathbb{F}_q^n$ such that

$$\tilde{u}_{[0:n-1]} = a \cdot \tilde{u}_{[0:n-1]} + b_{[0:n-1]}.$$

In fact, for each choice of $a \in \mathbb{F}_q$, we can always find a vector $b_{[0:n-1]} \in \mathbb{F}_q^n$ satisfying the above equation, so there are $q - 1$ choices of the pair $(a, b_{[0:n-1]})$ in total. We may pick an arbitrary one of them in this proof.\footnote{For the proof of Lemma 2, we can simply choose $a = 1$. Other choices of $a$ are needed for the proof of Lemma 3.}

Define $\bar{x}_{[0:n-1]} = \tilde{u}_{[0:n-1]} G_n$, $\bar{x}_{[0:n-1]} = \tilde{u}_{[0:n-1]} G_n$, and $x_{[0:n-1]} = b_{[0:n-1]} G_n$. Then we have $x_{[0:n-1]} = a \cdot x_{[0:n-1]} + x_{[0:n-1]}$, so

$$W^n(y_{[0:n-1]} | x_{[0:n-1]}) = W^n(a \cdot y_{[0:n-1]} + x_{[0:n-1]} | x_{[0:n-1]})$$

for all $y_{[0:n-1]} \in \mathcal{Y}^n$.

$$1[z \neq x_j] = [a \cdot z + x^b \neq x^j]$$

for all $z \in \mathbb{F}_q$ and all $0 \leq j \leq n - 1$.

In this proof, we use the shorthand notation

$$\tilde{u}_{[0:n-1]}(y_{[0:n-1]}) = \tilde{u}_{[0:n-1]}(y_{[0:n-1]}, A, \tilde{u}_A),$$

$$\bar{u}_{[0:n-1]}(y_{[0:n-1]}) = \bar{u}_{[0:n-1]}(y_{[0:n-1]}, A, \tilde{u}_A).$$

We first prove that

$$P(\tilde{u}_{[0:n-1]}(y_{[0:n-1]}) = u_{[0:n-1]}),$$

$$P(\tilde{u}_{[0:n-1]}(a \cdot y_{[0:n-1]} + x^b_{[0:n-1]}) = \bar{u}_{[0:n-1]}(y_{[0:n-1]}, A, \tilde{u}_A).$$

The randomness here comes from the random decision of the SC decoder when there is a tie in (2).

Note that

$$P(\tilde{u}_{[0:n-1]}(y_{[0:n-1]}) = u_{[0:n-1]}),$$

$$\prod_{i=0}^{n-1} P(\tilde{u}_i(y_{[0:i-1]}, u_{[i:i-1]})) = u_i,$$

$$P(\tilde{u}_{[0:n-1]}(a \cdot y_{[0:n-1]} + x^b_{[0:n-1]}) = a \cdot u_{[0:n-1]} + b_{[0:n-1]}).$$

Therefore, in order to prove (8), we only need to show that

$$P(\tilde{u}_i(y_{[0:i-1]}, u_{[i:i-1]})) = u_i,$$

$$P(\tilde{u}_i(a \cdot y_{[0:i-1]} + x^b_{[0:i-1]}, a \cdot u_{[i:i-1]} + b_{[i:i-1]})) = a \cdot u_i + b_i).$$

for all $0 \leq i \leq n - 1$ and all $u_{[i:i]} \in \mathbb{F}_q^{i+1}$. If $i \in A^c$, then $\tilde{u}_i = \tilde{u}_i$, and $\tilde{u}_i = \tilde{u}_i$. Since $\tilde{u}_i = a \cdot \tilde{u}_i + b_i$, the equality (9) clearly holds. If $i \in A$, then we need to analyze the synthetic channel in (2). More specifically, we have

$$W_i^n(y_{[0:n-1]}, u_{[i:i-1]} | X_i) = 1 q^n - 1 \sum_{u_{[i:i-1]} \in \mathcal{Y}^{i-1}} W^n(y_{[0:n-1]} | X_i | G_n),$$

$$= 1 q^n - 1 \sum_{u_{[i:i-1]} \in \mathcal{Y}^{i-1}} W^n(a \cdot y_{[0:n-1]} + x^b_{[0:n-1]} | G_n),$$

$$= W_i^n(a \cdot y_{[0:n-1]} + x^b_{[0:n-1]}, a \cdot u_{[i:i-1]} + b_{[i:i-1]} | a \cdot u_i + b_i).$$

The last equality holds because when $u_{[i:i-1]}$ ranges over all values in $\mathbb{F}_q^{i-1}$, the sum $a \cdot u_{[i:i-1]} + b_{[i:i-1]}$ also
ranges over all values in $\mathbb{F}_q^{n-1}$. This equality together with the decoding rule (2) immediately implies that (9) holds for $i \in \mathcal{A}$. This completes the proof of (8) and (9).

Recall the notation for the decoding result of the codeword vector in (3). Similarly to (7), we use the shorthand notation

$$\hat{x}_{[0:n-1]}(y_{[0:n-1]}),\hat{x}'_{[0:n-1]}(y_{[0:n-1]}),\hat{x}_{[0:n-1]}(y_{[0:n-1]}),\hat{x}'_{[0:n-1]}(y_{[0:n-1]}).$$

With this notation, Equation (8) is equivalent to

$$P(\hat{x}_{[0:n-1]}(y_{[0:n-1]})) = P(\hat{x}'_{[0:n-1]}(a \cdot y_{[0:n-1]} + x_{[0:n-1]}^b)) = a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b) \quad (10)$$

for all $x_{[0:n-1]} \in \mathbb{F}_q^n$. Combining this with (6), we have

$$\text{SER}_j(\text{Polar}(n, k, \mathcal{A}, \bar{u}_d), W, \bar{u}_d)$$

$$= \sum_{y_{[0:n-1]} \in \mathcal{Y}^n} \sum_{x_{[0:n-1]} \in \mathbb{F}_q^n} \left( W^n(y_{[0:n-1]} | \hat{x}_{[0:n-1]}(y_{[0:n-1]})) \cdot P(\hat{x}_{[0:n-1]}(y_{[0:n-1]})) = x_{[0:n-1]}(1)[x_j \neq \hat{x}_j) \right)$$

$$= \sum_{y_{[0:n-1]} \in \mathcal{Y}^n} \sum_{x_{[0:n-1]} \in \mathbb{F}_q^n} \left( W^n(a \cdot y_{[0:n-1]} + x_{[0:n-1]}^b) \cdot x_{[0:n-1]}^b) \cdot P(\hat{x}'_{[0:n-1]}(a \cdot y_{[0:n-1]} + x_{[0:n-1]}^b)) = a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b) \cdot 1(a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b \neq \hat{x}_j) \right)$$

$$= \sum_{y_{[0:n-1]} \in \mathcal{Y}^n} \sum_{x_{[0:n-1]} \in \mathbb{F}_q^n} \left( W^n(y_{[0:n-1]} | \hat{x}'_{[0:n-1]}(y_{[0:n-1]})) = x_{[0:n-1]}(1)[x_j \neq \hat{x}_j) \right)$$

$$= \text{SER}_j(\text{Polar}(n, k, \mathcal{A}, \bar{u}_d), W, \bar{u}_d).$$

The equality (1) is obtained from replacing $a \cdot y_{[0:n-1]} + x_{[0:n-1]}^b$ with $y_{[0:n-1]}$ and replacing $a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b$ with $x_{[0:n-1]}$. These two replacements are eligible because when $y_{[0:n-1]}$ ranges over $\mathcal{Y}^n$, $a \cdot y_{[0:n-1]} + x_{[0:n-1]}^b$ also ranges over all values in $\mathcal{Y}^n$; similarly, when $x_{[0:n-1]}$ ranges over $\mathbb{F}_q^n$, $a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b$ also ranges over all values in $\mathbb{F}_q^n$. Moreover, since $a \cdot x_{[0:n-1]}^b$ is the $j$-th coordinate of $a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b$, we also replace $a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b$ with $x_{[0:n-1]}$ when we replace $a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b$ with $x_{[0:n-1]}$. This completes the proof of the lemma.

This lemma implies that in the analysis of SER of the SC decoder, we can always assume that the true value $\bar{u}_{[n-1]}$ of the message vector (including both information symbols and frozen symbols) is all-zero, or equivalently, the transmitted codeword is all-zero. More specifically, we have the following expression for the SER:

$$\text{SER}_j(\text{Polar}(n, k, \mathcal{A}, \bar{u}_d), W)$$

$$= \text{SER}_j(\text{Polar}(n, k, \mathcal{A}, 0^n, k), W)$$

$$= \text{SER}_j(\text{Polar}(n, k, \mathcal{A}, 0^n, k), W, 0^k)$$

$$= \sum_{y_{[0:n-1]} \in \mathcal{Y}^n} \sum_{x_{[0:n-1]} \in \mathbb{F}_q^n} \left( W^n(y_{[0:n-1]} | 0^n) \cdot P(\hat{x}_{[0:n-1]}(y_{[0:n-1]}), \mathcal{A}, 0^n) = x_{[0:n-1]}(1)[x_j \neq 0) \right) \quad (11)$$

where $0^i$ is the all-zero vector of length $i$.

Note that (10) immediately implies the following lemma, which will be used later to prove the main theorem.

Lemma 3: Let $x_{[0:n-1]} = y_{[0:n-1]}G_0$ be a codeword of $\text{Polar}(n, k, \mathcal{A}, 0^n, k)$, i.e., $b_i = 0$ for all $i \in \mathcal{A}$. Then

$$P(\hat{x}_{[0:n-1]}(y_{[0:n-1]}), \mathcal{A}, 0^n) = x_{[0:n-1]}(1)$$

$$= P(\hat{x}_{[0:n-1]}(a \cdot y_{[0:n-1]} + x_{[0:n-1]}^b), \mathcal{A}, 0^n)$$

$$= a \cdot x_{[0:n-1]} + x_{[0:n-1]}^b$$

for all $a \in \mathbb{F}_q^n, x_{[0:n-1]} \in \mathbb{F}_q^n$ and all $y_{[0:n-1]} \in \mathcal{Y}^n$.

As a final remark, we note that we do not need condition (5) for the set $\mathcal{A}$ in Lemma 2 and Lemma 3, but we will need this condition later in the proof of the main theorem.

B. Recursive Implementation of the SC Decoder

In practice, the SC decoder is usually implemented recursively based on the transition probability. Below we recap this recursive structure since it is needed in the next step of our proof. Recall that $W : \mathbb{F}_q \rightarrow \mathcal{Y}$ is a symmetric channel. We denote the set of transition probabilities for an output symbol $y \in \mathcal{Y}$ as

$$T(y) = \{ W(y | u) : u \in \mathbb{F}_q \}.$$ 

For a channel output vector $y_{[0:n-1]} \in \mathcal{Y}^n$, we define an $n$-tuple

$$T_{[0:n-1]} = T_{[0:n-1]}(y_{[0:n-1]}) = (T_0, T_1, \ldots, T_{n-1}),$$

where $T_i = T(y_i)$ for $0 \leq i \leq n - 1$. Note that each coordinate $T_i = T(y_i)$ is the set of transition probabilities for the output symbol $y_i$. We write $T_{[0:n-1]}(y_{[0:n-1]})$ when we want to emphasize its dependence on $y_{[0:n-1]}$. In most scenarios this dependence is clear from the context, and we will simply write $T_{[0:n-1]}$.

It is well known that the decoding result of the SC decoder only depends on the transition probabilities of the channel outputs. In other words, if two channel output vectors have the same vector of transition probability sets, then their decoding results have the same probability distribution under the SC decoder. Below we will write $\bar{u}_i(T_{[0:n-1]}, u_{[i-1]}), \text{ and } \bar{u}_i(T_{[0:n-1]}, u_{[i-1]}), \text{ interchangeably. We also write } \bar{u}_{[0:n-1]}(y_{[0:n-1]}, \mathcal{A}, \bar{u}_d), \text{ and } \bar{u}_{[0:n-1]}(T_{[0:n-1]}, \mathcal{A}, \bar{u}_d), \text{ interchangeably.}$

For $a \in \mathbb{F}_q^n, b_{[0:n-1]} \in \mathbb{F}_q^n$, and $T_{[0:n-1]} = T_{[0:n-1]}(y_{[0:n-1]})$, we define

$$a \cdot a \cdot T_{[0:n-1]} + b_{[0:n-1]}$$

$$= (a \cdot T_0 + b_0, a \cdot T_1 + b_1, \ldots, a \cdot T_{n-1} + b_{n-1}),$$

where $a \cdot T_i + b_i = T(a \cdot y_i + b_i)$ for $0 \leq i \leq n - 1$. Each coordinate $a \cdot T_i + b_i = T(a \cdot y_i + b_i)$ is the set of transition probabilities for the output symbol $a \cdot y_i + b_i$. With the new notation, we restate Lemma 3 as follows.

3Recall that the SC decoder produces a random decoding result when there is a tie in (2).
Lemma 4: Let \( x_{[0:n-1]}^b \) be a codeword of Polar \((n, k, A, 0^{n-k})\), i.e., \( b_i = 0 \) for all \( i \in A \). Then
\[
\mathbb{P}(\hat{x}_{[0:n-1]}(T_{[0:n-1]}), A, 0^{n-k}) = \mathbb{P}(\hat{x}_{[0:n-1]}(\alpha \cdot T_{[0:n-1]} + x_{[0:n-1]}^b), A, 0^{n-k})
\]
for all \( a \in \mathbb{F}_q^*, x_{[0:n-1]}^b \in \mathbb{F}_q^n \) and all \( y_{[0:n-1]} \in \mathcal{Y}^n \).

We need some more notation to describe the recursive structure of the SC decoder. For \( y_0, y_1 \in \mathcal{Y} \) and \( u_0 \in \mathbb{F}_q \), we define two sets \( T^-(y_0, y_1) = \{ W^-(y_0, y_1, u) : u \in \mathbb{F}_q \} \) and \( T^+(y_0, y_1, u_0) = \{ W^+(y_0, y_1, u_0) : u \in \mathbb{F}_q \} \) of size \( q \), where
\[
W^-(y_0, y_1, u) = \frac{1}{q} \sum_{u_0 \in \mathbb{F}_q} W(y_0|u + \alpha u_1)W(y_1|u_1),
\]
\[
W^+(y_0, y_1, u_0) = \frac{1}{q} W(y_0|u_0 + \alpha u)W(y_1|u).
\]
(12)

For \( y_{[0:n-1]} \in \mathcal{Y}^n \) and \( z_{[0:n-2]/1} = \mathbb{F}_q^n/2 \), we define
\[
T^-_{[0:n-2/1]}(y_{[0:n-1]}) = (T_0, T_1, \ldots, T_{n/2-1}),
\]
where \( T_i^- = T^-(y_i, y_{i+n/2}) \) for \( 0 \leq i \leq n/2 - 1 \),
\[
T^+_{[0:n-2/1]}(y_{[0:n-1]}, z_{[0:n-2/1]}) = (T_0, T_1, \ldots, T_{n-2/1}^+),
\]
where \( T_i^+ = T^+(y_i, y_{i+n/2}, z_i) \) for \( 0 \leq i \leq n/2 - 1 \).
(13)

For a set \( A \subseteq \{0, 1, \ldots, n-1\} \), we define
\[
A^\perp = A \cap \{0, 1, \ldots, n/2 - 1\}
\]
and
\[
A^\perp = \{i : i + n/2 \in A \cap \{n/2, n/2 + 1, \ldots, n\}\}.
\]

The encoding procedure \( x_{[0:n-1]} = u_{[0:n-1]}G_n \) for polar codes can be decomposed in the following way: Let \( z_{[0:n-2]} = u_{[0:n-2]}G_n/2 \) and \( z_{[n/2-1]} = u_{[n/2-1]}G_n/2 \), i.e., we first encode the two halves separately. Then \( x_{[0:n-2]} = z_{[0:n-2]} + \alpha \cdot z_{[n/2-1]} \) and \( x_{[n/2-1]} = z_{[n/2-1]} \). This recursive structure can be summarized as
\[
u_{[0:n-1]}G_n = (u_{[0:n-2]} + \alpha \cdot u_{[n/2-2]})G_n/2, u_{[n/2-2]}G_n/2.
\]
(14)

From this point on, we assume that all the frozen symbols take value 0, which is justified by Lemma 2. We use the shorthand notation
\[
\hat{u}_{[0:n-1]}(T_{[0:n-1]}), A) = \hat{u}_{[0:n-1]}(T_{[0:n-1]}), A, 0^{n-k}),
\]
\[
\hat{x}_{[0:n-1]}(T_{[0:n-1]}), A) = \hat{x}_{[0:n-1]}(T_{[0:n-1]}), A, 0^{n-k}).
\]

Given a channel output vector \( y_{[0:n-1]} \), the SC decoder first decodes \( u_{[0:n/2-1]} \) as
\[
\hat{u}_{[0:n/2-1]}(T_{[0:n/2-1]}(y_{[0:n-1]}), A^\perp).
\]
Then it calculates \( \hat{z}_{[0:n/2-1]} = \hat{u}_{[0:n/2-1]}(T_{[0:n/2-1]}(y_{[0:n-1]}), A^\perp)G_n/2 \). In the next step, the SC decoder decodes \( u_{[n/2:n-1]} \) as
\[
\hat{u}_{[n/2:n-1]}(T_{[n/2:n-1]}(y_{[0:n-1]}), A^\perp).
\]
Proof: We will prove another equation that is equivalent to (17). Specifically, we will prove that
\[
\mathbb{P}(\hat{x}_{0:n-1}(y_{0:n-1}), A) = (z_{0:n/2-1} + \alpha \cdot z_{n/2:n-1}, z_{n/2:n-1})
\]
and observing that
\[
\hat{P} - P_0 \quad \text{with} \quad q, y
\]
for all \(y_{0:n-1} \in \mathbb{Y}^n\) and all \(z_{0:n-1} \in \mathbb{F}_q^n\). These two equations are equivalent because (18) is obtained from replacing \(x_{0:n-1}\) with \((z_{0:n/2-1} + \alpha \cdot z_{n/2:n-1}, z_{n/2:n-1})\) in (17).
By (12), we have
\[
W^-(\alpha \cdot y_1, -\alpha^{-1} \cdot y_0|u_0) = -q \sum_{u_1 \in \mathbb{F}_q} W(\alpha \cdot y_1|u_0 + \alpha u_1) W(\alpha^{-1} \cdot y_0|u_1)
\]
and
\[
W^+(\alpha \cdot y_1, -\alpha^{-1} \cdot y_0|u_0) = -q \sum_{u_1 \in \mathbb{F}_q} W(y_1|\alpha^{-1} u_0 - u_1) W(y_0|\alpha u_1)
\]
where equality (a) is obtained by setting \(v = -\alpha^{-1} u_0 - u_1\) and observing that \(-\alpha u_1 = u_0 + \alpha v\). Therefore, \(T^-(y_0, y_1) = T^-(\alpha \cdot y_1, -\alpha^{-1} \cdot y_0)\). Defining (13) further implies that \(T^+_0(y_0, y_1) = T^+_0(y_{0:n-1})\) for all \(y_{0:n:n-1} \in \mathbb{Y}^n\). Therefore,
\[
\mathbb{P}(\hat{x}_{0:n/2-1}(T^+_0(y_{0:n-1}), A) = z_{0:n/2-1})
\]
\[
= \mathbb{P}(\hat{x}_{0:n/2-1}(T^+_0(y_{0:n-1}), A) = z_{0:n/2-1})
\]
for all \(y_{0:n-1} \in \mathbb{Y}^n\).

By (12),
\[
W^+(\alpha \cdot y_1, -\alpha^{-1} \cdot y_0|u_0) = -q W(\alpha \cdot y_1|u_0 + \alpha u_1) W(\alpha^{-1} \cdot y_0|u_1)
\]
and
\[
W^+(\alpha \cdot y_1, -\alpha^{-1} \cdot y_0|u_0) = -q W(y_1|\alpha^{-1} u_0 - u_1) W(y_0|\alpha u_1)
\]
for all \(y_0, y_1 \in \mathbb{Y}\) and \(u_0, u_1 \in \mathbb{F}_q\), where \(\langle y_0, y_1, u_0, u_1 \rangle\) denotes an output symbol of the synthetic channel \(W^+ : \mathbb{F}_q \rightarrow \mathbb{Y}^2 \times \mathbb{F}_q\). In the Appendix, we prove that \(W^+\) is also \(\mathbb{F}_q\)-symmetric memoryless.

To see this, we consider the \(i\)th \(0 \leq i \leq n/2 - 1\) coordinate of \(T^+_0(y_{0:n-1})\) and the \(i\)th coordinate of \(A\).

Observe that
\[
T^+_0(y_{0:n-1}) = \xi_{m-r}(x_{0:n-1})
\]
for all \(y_{0:n-1} \in \mathbb{Y}^n\), and all \(x_{0:n-1} \in \mathbb{F}_q^n\), and all \(1 \leq r \leq m\).

Proof: We prove by induction on \(r\). The base case \(r = 1\) is already proved in Lemma 5. Now we assume that (21) holds for \(r - 1\) and all values of \(m\), and we prove it for \(r\).
vector on the right-hand side is $T^-(y_{\xi_{(m-1)}(i)} y_{\xi_{(m-1)}(i)+n/2})$. Since $i < n/2$, we have $\xi_{m-r}(i) = \xi_{(m-1)}(i)$ and $\xi_{m-r}(i + n/2) = \xi_{(m-1)}(i) + n/2$, implying that the $i$th coordinate of the two vectors are the same. Since $A$ satisfies the condition (5), both $A^+$ and $A^-$ also satisfy the condition (5). By the induction hypothesis, (21) holds for $r - 1$ and $m - 1$, so

$$
\mathbb{P}(\hat{x}_{[0:n/2-1]}(T^-[y_{[0:n/2-1]}(y_{[0:n-1]}), A^-) = z_{[0:n/2-1]}) = \mathbb{P}(\hat{x}_{[0:n/2-1]}(\xi_m^{-1}(T^-[y_{[0:n/2-1]}(y_{[0:n-1]}), A^-) = \xi_m^{-1}(z_{[0:n/2-1]})) = \mathbb{P}(\hat{x}_{[0:n/2-1]}(T^-[y_{[0:n/2-1]}(y_{[0:n-1]}), A^-) = \xi_m^{-1}(z_{[0:n/2-1]})) (22)
$$

for all $z_{[0:n/2-1]} \in \mathbb{F}_q^{n/2}$. It is easy to verify that

$$
T^+[y_{\xi_{(m-1)}(i)} y_{\xi_{(m-1)}(i)+n/2}] = \xi_{m-r}(z_{[0:n/2-1]}). (23)
$$

To see this, the $i$th ($0 \leq i \leq n/2 - 1$) coordinate of the vector on the left-hand side is $T^+[y_{\xi_{(m-1)}(i)} y_{\xi_{(m-1)}(i)+n/2}]$. Since $i < n/2$, we have $\xi_{(m-1)}(i) = \xi_{m-r}(i)$ and $\xi_{(m-1)}(i + n/2) = \xi_{m-r}(i) + n/2$, implying that the $i$th coordinate of the two vectors are the same. Again by the induction hypothesis,

$$
\mathbb{P}(\hat{x}_{[0:n/2-1]}(T^+[y_{[0:n/2-1]}(y_{[0:n-1]}), A^+) = z_{[0:n/2-n/1]}) = \mathbb{P}(\hat{x}_{[0:n/2-1]}(\xi_m^{-1}(T^+[y_{[0:n/2-1]}(y_{[0:n-1]}), A^+) = \xi_m^{-1}(z_{[0:n/2-n/1]})) = \mathbb{P}(\hat{x}_{[0:n/2-1]}(T^+[y_{[0:n/2-1]}(y_{[0:n-1]}), A^+) = \xi_m^{-1}(z_{[0:n/2-n/1]}))
$$

for all $z_{[0:n/2-n/1]} \in \mathbb{F}_q^n$. Combining this with (16) and (22), we obtain that

$$
\mathbb{P}(\hat{x}_{[0:n-1]}(y_{[0:n-1]}, A) = (z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}, z_{[n/2-n/1]})) = \mathbb{P}(\hat{x}_{[0:n-1]}(y_{[0:n-1]}, A) = (\xi_m^{-1}(z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}, \xi_m^{-1}(z_{[n/2-n/1]})))
$$

for all $z_{[0:n-1]} \in \mathbb{F}_q^n$. Since

$$
(\xi_m^{-1}(z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}), \xi_m^{-1}(z_{[n/2-n/1]})) = (\xi_m^{-1}(z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}), \xi_{m-r}(z_{[n/2-n/1]})),
$$

we further obtain that

$$
\mathbb{P}(\hat{x}_{[0:n-1]}(y_{[0:n-1]}, A) = (z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}, z_{[n/2-n/1]}) = \mathbb{P}(\hat{x}_{[0:n-1]}(y_{[0:n-1]}, A) = (\xi_m^{-1}(z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}, \xi_{m-r}(z_{[n/2-n/1]})))
$$

for all $z_{[0:n-1]} \in \mathbb{F}_q^n$. Finally, (21) follows from replacing $(z_{[0:n/2-1]} + \alpha \cdot z_{[n/2-n/1]}, z_{[n/2-n/1]})$ with $x_{[0:n-1]}$. This completes the proof of the lemma.

**Lemma 7:** Let $n = 2^m$. Let $W$ be a $\mathbb{F}_q$-symmetric memoryless channel. Suppose that the index set $A$ of information symbols satisfies the condition (5). Then

$$
\text{SER}_j(\text{Polar}(n, k, A, \bar{u}_A), W) = \text{SER}_{j-r}(\text{Polar}(n, k, A, \bar{u}_A), W)
$$

for all $0 \leq j \leq n-1$, all $0 \leq r \leq m-1$, and all $\bar{u}_A \in \mathbb{F}_q^{n-k}$.

**Proof:** By (11), we have

$$
\text{SER}_j(\text{Polar}(n, k, A, \bar{u}_A), W) = \sum_{y_{[0:n-1]} \in \mathbb{F}_q^n} \sum_{x_{[0:n-1]} \in \mathbb{F}_q^n} \left( W^n(y_{[0:n-1]}|0^n) \cdot \mathbb{P}(\hat{x}_{[0:n-1]}(y_{[0:n-1]}, A) = x_{[0:n-1]} \cdot 1[x_j \neq 0]) \right)
$$

Fig. 1. Simulation results of the $(n = 256, k = 128)$ binary polar code over binary-input AWGN channel with $E_b/N_0 = 2$ dB.
of u with the observation in [4], we are still not able to explain to the previous bits. The BER of the last message bit is an outlier because it drops abruptly compared to the previous bits. We can see that there is an increasing trend in the BERs of message bits (increasing with the indices). The BER of the last message bit because the last codeword bit replacing y = 0 from replacing ξ(m) with y = 1 and replacing ξ(m) with y = 0. This means that SER({Pol(n, k, A, u_A), W}) = SER0({Pol(n, k, A, u_A), W}) for all j ∈ {0, 1, ..., n – 1} and completes the proof of Theorem 1.

IV. THE CONNECTION TO [4]

In this section, we restrict ourselves to binary polar codes. As mentioned in Introduction, we observed the simulation results in Fig. 1 as we repeated the experiments in [4]. From Fig. 1 we can see that the variance of the BERs of message bits is very large. In contrast, the BERs of codeword bits are extremely stable, to the extent that they form a straight line. At this point, it makes sense to discuss the connection between our results and [4].

We still use u[0:n–1] to denote the message vector and use x[0:n–1] to denote the codeword vector. The set A is still the index set of information bits. The main observation in [4] is that the average BER of x_A is much smaller than the average BER of u_A under the SC decoder. This is somewhat counter-intuitive because the SC decoder directly decides u_A, and x_A is obtained from multiplying the decoding result of u[0:n–1] with the encoding matrix G_n. Even till today, no rigorous analysis is available to explain this phenomenon.

The results in this paper imply that the BERs of all the codeword bits are equal to each other, and they equal to the BER of the last message bit because the last codeword bit is the same as the last message bit. On the bright side, the last message bit is the best-protected message bit if we assume that all the previous message bits are decoded correctly. However, we also have an argument in the opposite direction: The decoding error in the SC decoder accumulates, and the last message bit takes the most damage from decoding errors in previous message bits. In fact, if we look at Fig. 1 closely, we can see that there is an increasing trend in the BERs of message bits (increasing with the indices). The BER of the last message bit is an outlier because it drops abruptly compared to the previous bits.

To conclude, although our result is somewhat correlated with the observation in [4], we are still not able to explain the phenomenon in [4]. That calls for future research effort.

APPENDIX

Lemma 8: If W is an Fq-symmetric memoryless channel, then both W^+ and W^- defined in (12) are Fq-symmetric memoryless channels.

Proof: We first prove the claim for W^+. By definition, we need to show that there exist q permutations {σ_b : b ∈ F_q} on the output alphabet Y^2 × F_q such that W^+(y_0, y_1, u_0|u) = W^+(σ_{u’-u}(y_0, y_1, u_0)|u’) for all y_0, y_1 ∈ Y and all u, u’, u_0 ∈ F_q. (2) there exist q – 1 permutations {π_a : a ∈ F_q^∗} on Y^2 × F_q such that W^+(y_0, y_1, u_0|u) = W^+(π_a((y_0, y_1, u_0))|au) for all y_0, y_1 ∈ Y, u, u_0 ∈ F_q, and a ∈ F_q^∗.

We may choose σ_b((y_0, y_1, u_0)) = (y_0 + ab, y_1 + b, u_0) and π_a((y_0, y_1, u_0)) = (a · y_0, a · y_1, a · u_0). The following calculations allow us to verify this choice.

W^+(σ_b((y_0, y_1, u_0)))|u + b) = W^+(y_0 + ab, y_1 + b, u_0|u + b) = 1 - qW(y_0 + ab|u_0 + a(u + b))W(y_1 + b|u + b) = 1 - qW(y_0|u_0 + au)W(y_1|u) = W^+(y_0, y_1, u_0|u),

W^+(π_a((y_0, y_1, u_0))|au) = W^+(a · y_0, a · y_1, a · u_0|au) = 1 - qW(a · y_0|au_0 + au)W(a · y_1|au) = 1 - qW(y_0|u_0 + au)W(y_1|u) = W^+(y_0, y_1, u_0|u).

As for W^-, we need to show that (1) there exist q permutations {σ_b : b ∈ F_q} on the output alphabet Y^2 such that W^–(y_0, y_1|u) = W^–(σ_{u’-u}(y_0, y_1)|u’) for all y_0, y_1 ∈ Y and all u, u’ ∈ F_q; (2) there exist q – 1 permutations {π_a : a ∈ F_q^∗} on Y^2 such that W^–(y_0, y_1|u) = W^–(π_a((y_0, y_1, u_0))|au) for all y_0, y_1 ∈ Y, u ∈ F_q, and a ∈ F_q^∗.

We may choose σ_b((y_0, y_1)) = (y_0 + b, y_1) and π_a((y_0, y_1)) = (a · y_0, a · y_1). The following calculations allow us to verify this choice.

W^–(σ_b((y_0, y_1, u_0)))|u + b) = W^–(y_0 + b, y_1|u + b) = 1 - q W(y_0 + b|u + b + au_1)W(y_1|u_1) = 1 - q W(y_0|u + au_1)W(y_1|u_1) = W^–(y_0, y_1, u_0|u),

W^–(π_a((y_0, y_1, u_0))|au) = W^–(a · y_0, a · y_1, a u_0|au) = 1 - q W(a · y_0|au_0 + au)W(a · y_1|au) = 1 - q W(y_0|u_0 + au)W(y_1|u) = W^–(y_0, y_1, u_0|u).
ACKNOWLEDGMENT

The authors would like to thank Ling Liu and Henry Pfister for providing valuable feedback on an earlier version of this article.

REFERENCES

[1] E. Arikan, “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, Jul. 2008.
[2] I. Tal and A. Vardy, “List decoding of polar codes,” IEEE Trans. Inf. Theory, vol. 61, no. 5, pp. 2213–2226, May 2015.
[3] K. Niu and K. Chen, “CRC-aided decoding of polar codes,” IEEE Commun. Lett., vol. 16, no. 10, pp. 1668–1671, Oct. 2012.
[4] E. Arikan, “Systematic polar coding,” IEEE Commun. Lett., vol. 15, no. 8, pp. 860–862, Aug. 2011.
[5] M. Geiselhart, A. Elkelesh, M. Ebada, S. Cammerer, and S. T. Brink, “Automorphism ensemble decoding of Reed–Müller codes,” IEEE Trans. Commun., vol. 69, no. 10, pp. 6424–6438, Jul. 2021.
[6] R. Mori and T. Tanaka, “Source and channel polarization over finite fields and Reed-Solomon matrices,” IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2720–2736, May 2014.
[7] V. Guruswami, A. Rudra, and M. Sudan. (2012). Essential Coding Theory. Draft. [Online]. Available: http://www.cse.buffalo.edu/tratti/courses/coding-theory/book
[8] M. Bardet, V. Dragoi, A. Otmani, and J.-P. Tillich, “Algebraic properties of polar codes from a new polynomial formalism,” 2016, arXiv:1601.06215.
[9] 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), Jul. 2016, pp. 230–234.
[10] C. Schurch, “A partial order for the synthesized channels of a polar code,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2016, pp. 220–224.
[11] G. Sarkis, I. Tal, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, “Flexible and low-complexity encoding and decoding of systematic polar codes,” IEEE Trans. Commun., vol. 64, no. 7, pp. 2732–2745, Jul. 2016.
[12] M. Mondelli, S. H. Hassani, and R. L. Urbanke, “Construction of polar codes with sublinear complexity,” IEEE Trans. Inf. Theory, vol. 65, no. 5, pp. 2782–2791, May 2019.

Guodong Li is currently pursuing the Ph.D. degree with the School of Cyber Science and Technology, Shandong University, Qingdao, China. His research interests include information theory and coding theory.

Min Ye received the B.S. degree in electrical engineering from Peking University, Beijing, China, in 2012, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA, in 2017. He spent two years as a Post-Doctoral Researcher with Princeton University. Since 2019, he has been an Assistant Professor with the Tsinghua Shenzhen International Graduate School, Tsinghua-Berkeley Shenzhen Institute, Shenzhen, China. His research interests include coding theory, information theory, differential privacy, and machine learning. He received the 2017 IEEE Data Storage Best Paper Award.

Sihuang Hu (Member, IEEE) received the B.Sc. degree in applied mathematics from Beihang University, Beijing, China, in 2008, and the Ph.D. degree in applied mathematics from Zhejiang University, Hangzhou, China, in 2014. He is currently a Professor with Shandong University, Qingdao, China. Before that, he was a Post-Doctoral Researcher with RWTH Aachen University, Germany, from 2017 to 2019 and from 2014 to 2015, and a Post-Doctoral Researcher with Tel Aviv University, Israel, from 2015 to 2017. His research interests include lattices, combinatorics, and coding theory for communication and storage systems. He was a recipient of the Humboldt Research Fellowship in 2017.