On the Symmetry of Polar Codes for Symmetric Binary-Input Discrete Memoryless Channels

Qiming Wang  
School of Information Science and Technology  
University of Science and Technology of China  
Email: hugh1234@mail.ustc.edu.cn

Liping Li  
School of Electronics and Information Engineering  
Anhui University, China,  
Email: liping_li@ahu.edu.cn

Abstract—In this paper, we study the symmetry of polar codes on symmetric binary-input discrete memoryless channels (B-DMC). The symmetry property of polar codes is originally pointed out in Arikan’s work for general B-DMC channels. With the symmetry, the output vector $y_i^N$ ($N$ be the block length) can be divided into equivalence classes in terms of their transition probabilities. In this paper, we present a new frame of analysis on the symmetry of polar codes for B-DMC channels. Theorems are provided to characterize the symmetries among the received vectors. With this new perspective, we can fully utilize the property of the underlying channels and reduce the number of equivalence classes. The analysis is applied to binary symmetric channels and shows a great reduction of the number of equivalence classes compared with the original symmetry setting.

I. INTRODUCTION

Polar codes are invented by Arikan in [1]. Polar codes are the first class of codes that can achieve the capacity for symmetric binary-input discrete memoryless channels (B-DMC) with a low complexity. The encoding and decoding (with successive cancellation, SC) has a complexity of $O(N \log N)$ [1]. The binary input alphabet in Arikan seminal work [1] is later on extended to non-binary input alphabet [2]–[4]. The construction of polar codes are reported in [5]–[8] and different procedures are proposed assuming the original $2 \times 2$ kernel matrix. Polar codes based on the kernel matrices of size $l \times l$ are studied in [9]. In this paper, we study the original binary-input coding scheme.

Let $W$ be a binary discrete memoryless channel (B-DMC). Let $N$ be the code block length. The transition probability of the $i$th split channel is determined by the received vector (with a length $N$), and the previously decoded $i-1$ bits, resulting in a number of possible outputs of $|\mathcal{Y}|^{N+1}$ where $\mathcal{Y}$ is the output alphabet. In [1], Arikan showed that if $W$ is symmetric, then the split channel is also symmetric and the number of possible outputs for the $i$th split channel can be reduced to $|\mathcal{Y}|^i$. The received vectors with the same transition probability form an equivalence class. The construction of polar codes in [5]–[8] did not fully utilize the property of the symmetry of polar codes as we just describe simply because the number $|\mathcal{Y}|^i$ is still large for large block lengths.

In this paper, we study the further symmetry of polar codes for B-DMC channels. We show that finding the number of classes of received vectors with the same transition probability can be transformed into a problem of finding the number of solutions to a set of equations. Given the symmetry of the underlying channel $W$, the number of solutions to the set of linear equations can be greatly smaller than $|\mathcal{Y}|^i$ for the $i$th split channel. We provide theoretical analysis on how to find the equivalent received vectors. The exact number of equivalence classes for a set of split channels are also provided in closed form expressions. The symmetry properties of polar codes in this paper are applied to binary symmetric channels (BSC). The comparison between the number of equivalence classes obtained in this paper and the original number $2^N$ for the $i$th split channel shows that our frame of analysis can be conveniently applied when trying to explore the symmetries of polar codes for potential uses.

Following the notations in [1], in this paper, we use $v_i^N$ to represent a row vector with elements $(v_1, v_2, ..., v_N)$. We also use $v$ to represent the same vector for notational convenience. Given a vector $v_i^N$, the vector $v_j^i$ is a subvector $(v_i, ..., v_j)$ with $1 \leq i, j \leq N$. If there is a set $A \subseteq \{1, 2, ..., N\}$, then $v_A$ denotes a subvector with elements in $\{v_i, i \in A\}$.

The rest of the paper is organized as follows. In Section II the symmetry of polar coding is introduced. Section III introduces theorems used in this paper for characterizing the equivalence classes. Then theorems and procedures on how to find the exact number of equivalence classes are presented in Section IV. Concluding remarks are presented in Section VII.

II. SYMMETRY OF POLAR CODES

In this section, we briefly restate the symmetry of polar codes introduced in [1] and present some of our own notations used in sequel of the paper.

Denote the transition probability of the underlying channel $W$ as $W(y|x)$. Define

$$W_N(y_1^N|u_1^N) = W^N(y_1^N|u_1^NG_N)$$

where $W^N(y_1^N|u_1^NG_N)$ is the $N$-time uses of the underlying channel $W$, the block length $N = 2^n$, and $G_N = BF^{\otimes n} = F^{\otimes n}B$. The matrix $B$ is the bit reverse permutation matrix, and $F = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix}$ [1]. In this paper we use $G_N = BF^{\otimes n}$ directly instead of $G_N = BF^{\otimes n}$. The only difference $G_N = BF^{\otimes n}$ makes, compared with $G_N = F^{\otimes n}$, is to perform a permutation on the received vector $y_1^N$, which doesn’t affect our analysis. Note that throughout this paper, the input vector $u_1^N$,
the matrix \( F = [1 0] \), and the generator matrix \( G_N \) are all defined on the binary field \( \mathbb{F}_2 \). The output \( y_1^n \) is defined on the set \( \mathcal{Y} \).

The operation in (11) is the channel combining stage. In the channel split stage, the \( i \)th split channel has a transition probability of

\[
W_N^{(i)}(y_1^N, u_{i-1}^1 | u_i) = \sum_{u_{i-1}^N \in \mathbb{A}^{N-i}} \frac{1}{2^{N-i}} W_N(y_1^N | u_1^N) \tag{2}
\]

To facilitate the analysis, the operator \( \cdot \) is defined as:

\[
0 \cdot y = y, \quad y \in \mathcal{Y} \tag{3}
\]

\[
1 \cdot y = y, \quad y, v \in \mathcal{Y} \tag{4}
\]

and \( v \) satisfies

\[
W(v(0)) = W(y(1)) \tag{5}
\]

\[
W(v(1)) = W(y(0)) \tag{6}
\]

In other words, \( v \) is the corresponding symbol that is symmetric to \( y \). To divide the output alphabet into two groups, we also define two sets:

\[
\text{SELF(} \mathcal{Y} \text{)} = \{ y \in \mathcal{Y} | 1 \cdot y = y \} \tag{7}
\]

\[
\text{SYMM(} \mathcal{Y} \text{)} = \{ y \in \mathcal{Y} | 1 \cdot y \neq y \} \tag{8}
\]

Set \( \text{SELF(} \mathcal{Y} \text{)} \) includes the output \( y \) which has a symmetric symbol of itself while set \( \text{SYMM(} \mathcal{Y} \text{)} \) has elements with a different symmetric symbol. Obviously, the cardinality of the set \( \text{SYMM}(\mathcal{Y}) \), \( |\text{SYMM}(\mathcal{Y})| \), is an even number. As shown in (11), for symmetric B-DMC channels, it is sufficient to discuss

\[
W_N^{(i)}(y_1^N, 0_{i-1}^1 | 0) = \sum_{u_{i-1}^N \in \mathbb{A}^{N-i}, u_1^N = 0_1} \frac{1}{2^{N-i}} W_N(y_1^N | u_1^N) \tag{9}
\]

as every \( W_N^{(i)}(y_1^N, u_{i-1}^1 | u_i) \) is equivalent to a \( W_N^{(i)}(y_1^N, 0_{i-1}^1 | 0) \) where \( y_1^N \) is selected from the following set:

\[
\tilde{A}_{i+1} = \{ y_1^N : a_1^N G_N \cdot y_1^N, \quad a_1^N = u_1^N \} \tag{10}
\]

In (10), the operation \( a_1^N G_N \cdot y_1^N \) is the element-wise operation of the \( \cdot \) operation defined in (3) and (4). In this paper, we will simply use \( W_N^{(i)}(y, y \in \mathcal{Y}) \) to replace \( W_N^{(i)}(y_1^N, 0_{i-1}^1 | 0) \) as the previous bits and the current bit are all zeros. And if two received vectors \( y \) and \( v \) have the same transition probability

\[
W_N^{(i)}(y) = W_N^{(i)}(v) \tag{11}
\]

then we call \( y \) and \( v \) probability equivalent on bit channel \( i \). See from (10), if \( y \) and \( v \) are probability equivalent on bit channel \( i \), they are also probability equivalent on the row space of the submatrix \( G_N^{i+1} \), where \( G_N^{i+1} \) means the rows of \( G_N \) from the \( (i + 1) \)th row to the \( N \)th row.

### III. Characterization of the Symmetry

For the convenience of the description, we denote the submatrix \( G_N^{i+1} \) of the matrix \( G_N \) as \( A(N, i) \). Then \( A(2N, i) \) is the submatrix formed by the rows from \( (i + 1) \) to \( 2N \) of \( G_{2N} = F \otimes n+1 \). For a given block length \( N \), we abbreviate \( A(N, i) = G_N^{i+1} \) as \( A = G_N^{i+1} \) since the size of \( A \) is clear from the context. Denote \( R(A) \) as the row space of \( A \). For the received vector \( y \in \mathcal{Y}^N \), define

\[
R(A) \cdot y = \{ u \cdot y | u \in R(A) \} \tag{12}
\]

In the following, we use a simpler notation to refer to the set defined in (12): \( R(A, y) = R(A) \cdot y \). Seen from (9), the transition probability for the \( i \)th channel is exactly summed over the row space of \( A \). Using the notation \( W_N^{(i)}(y) = W_N^{(i)}(y, 0_{i-1}^1 | 0) \), we can rewrite (9) as

\[
W_N^{(i)}(y) = \sum_{u \in R(A)} \frac{1}{2^{N-i}} W_N(y | u) \tag{13}
\]

\[
= \sum_{u \in R(A)} \frac{1}{2^{N-i}} W_N(u \cdot y | 0_1^N) \tag{14}
\]

\[
= \sum_{v \in R(A,y)} \frac{1}{2^{N-i}} W_N(v | 0) \tag{15}
\]

The equality in (14) is obtained from the application of the symmetry property (3) and (4) of the underlying channel \( W \).

In the following, we present theorems to characterize the properties of the received vectors which belong to the same equivalence class. With our frame of analysis, the original symmetry property presented in (11) is also provided.

**Theorem 1:** Let \( P \) be a permutation matrix. Denote \( T_P(\cdot) \) as the permutation transformation. If there exists an invertible matrix \( H \) so that \( AP = HA \), then \( v = T_P(y) \) and \( y \) are probability equivalent on \( A \), where \( A = A(N, i) \).

**Proof:** Since the matrix \( P \) is a permutation matrix, the inverse matrix of it is \( P^T \). Thus the transformation \( T_P \) is a bijection. Let’s consider the two spaces \( R(A, T_P(y)) \) and \( R(AP, T_P(y)) \). Obviously the row space of \( A \) and \( AP \) is the same, so the row space of \( A \) and \( AP \) must also be the same, since \( AP = HA \) and \( H \) is invertible. For each \( z \in R(A, y) \), there exists a \( T_P(z) \in R(AP, T_P(y)) = R(A, T_P(y)) \) and vice versa. Since

\[
W_N(z | 0) = W_N(T_P(z) | 0) \tag{16}
\]

\( y \) and \( T_P(y) \) must be probability equivalent on \( A \) for the \( i \)th channel.

The following corollary can be easily obtained from Theorem 1 if we assume the matrix \( A \) is an invertible matrix.

**Corollary 1:** The vectors \( y, T_P(y) \in \mathcal{Y}^N \) are always probability equivalent on an invertible matrix, where \( P \) is a permutation matrix.

Note that the matrix \( A = G_N^{i+1} \) can’t be an invertible matrix for any bit channel \( i \). But if \( A \) can be decomposed into several small matrices and each of these small matrices is row equivalent to an identity matrix, then this Corollary 1 can...
be used to characterize the symmetry of the received vectors operated on those small matrices.

**Theorem 2:** If there exists a vector \( \mathbf{u} \in \mathcal{V}^{N-i} \) so that \( \mathbf{k} = \mathbf{u} \mathbf{A} \), then \( \mathbf{v} = \mathbf{k} \cdot \mathbf{y} \) and \( \mathbf{y} \) are probability equivalent on bit channel \( i \).

**Corollary 2:** For BSC channels, the number of equivalence classes for the \( i \)th bit channel can be calculated from \( \mathbf{y} \in \mathbb{F}_2^N \) and \( y^{N+1}_{i+1} = 0 \).

**Proof:** For a BSC channel, the operation in (5) and (8) becomes \( 0 \cdot \mathbf{y} = \mathbf{y} \) and \( 1 \cdot \mathbf{y} = \mathbf{y} + 1 \). Also note that columns from \( i + 1 \) to \( N \) of the matrix \( \mathbf{A} \) form an independent set as the matrix \( G_N \) is lower triangular. Therefore, for any \( y^{N+1}_{i+1} \) defined in the binary field \( \mathbb{F}_2 \), we can find a unique vector \( \mathbf{u} \in \mathcal{V}^{N-i} \) such that the vector \( \mathbf{k} = \mathbf{u} \mathbf{A} \) and \( k^{N+1}_{i+1} = y^{N+1}_{i+1} \). This makes the vector \( \mathbf{v} = \mathbf{u} \mathbf{A} \cdot \mathbf{y} \) have \( v^{N+1}_{i+1} = 0^{N+1} - i - 1 \). From Theorem 2, it’s seen that \( \mathbf{v} \) and \( \mathbf{y} \) are probability equivalent on the \( i \)th bit channel. Thus the equivalence classes for BSC channels can be investigated based on the received vector \( v^{N+1}_{i+1} \) with \( v^{N+1}_{i+1} = 0^{N+1} - i - 1 \), which results in \( 2^i \) as the largest possible number of equivalence classes.

The following Theorem 3 states that the equivalence of two vectors for the \( i \)th bit channels is preserved when the block length increases from \( N \) to \( 2N \).

**Theorem 3:** If vectors \( \mathbf{y} \) and \( \mathbf{v} \) are probability equivalent on \( A(N, i) = A \), then \( \mathbf{u} = (1, 0) \otimes \mathbf{y} \) and \( \mathbf{t} = (1, 0) \otimes \mathbf{v} \) are also probability equivalent on \( A(2N, i) \).

**Proof:** Define \( B = (1, 0) \otimes A \), \( C = A(2N, i) \), and \( D = A(2N, N) \). The matrix \( \mathbf{D} \) is row equivalent to \( (1, 1) \otimes I_N \) from the definition of the generator matrix \( G_N \). It’s clear that the row space of the matrix \( \mathbf{B} \), \( \mathbf{D} \) are subsets of the row space of matrix \( \mathcal{C} = R(B) \subseteq R(C), R(D) \subseteq R(C) \). If \( \mathbf{y} \) and \( \mathbf{v} \) are probability equivalent on \( A \), there exists a bijection \( T_{\mathbf{p}} \) such that

\[
W^{2N}(\mathbf{h}|0) = W^N(T_{\mathbf{p}}(\mathbf{h})|0)
\]

where \( \mathbf{h} \in R(A, \mathbf{y}) \), \( T_{\mathbf{p}}(\mathbf{h}) \in R(A, \mathbf{v}) \), and \( P_{\mathbf{h}} \) is a permutation matrix that may vary with \( \mathbf{h} \). Then \( \mathbf{u} \) and \( \mathbf{t} \) must also be probability equivalent on \( B \), since there exists a bijection \( f \) such that

\[
f : \mathbf{h} \in R(B, \mathbf{u}) \to \mathbf{h}L_{\mathbf{h}} \in R(B, \mathbf{t}) \quad L_{\mathbf{h}} = (1, 1) \otimes P_{\mathbf{h}}
\]

Now, there’s a partition of \( R(C, \mathbf{t}) \),

\[
\text{Par}(\mathbf{t}) = \{ R(C) \cdot \mathbf{r} | \mathbf{r} \in R(B, \mathbf{t}) \}
\]

The same for \( R(C, \mathbf{u}) \),

\[
\text{Par}(\mathbf{u}) = \{ R(C) \cdot \mathbf{r} | \mathbf{r} \in R(B, \mathbf{u}) \}
\]

For any \( \mathbf{e} \in R(C, \mathbf{u}) \), it must belong to some \( R(D) \cdot \mathbf{r} \in \text{Par}(\mathbf{u}) \). According to Theorem 1, \( \mathbf{r} \in R(B, \mathbf{u}) \) and \( f(\mathbf{r}) \in R(B, \mathbf{t}) \) are probability equivalent on \( R(D) \), since \( D \) and \( DL_{\mathbf{h}} \) are row equivalent. It follows that there exists a bijection \( e^{*}_{\mathbf{e}}(\mathbf{e}) \in R(D) \cdot f(\mathbf{r}) \subseteq R(C, \mathbf{t}) \) that satisfies \( e^{*}_{\mathbf{e}}(\mathbf{e}) = eP_{\mathbf{e}} \).

Therefore, \( \mathbf{u} \) and \( \mathbf{t} \) are probability equivalent on \( A(2N,i) \).

**Corollary 3:** For BSC channels (which means \( \mathcal{Y} = \mathbb{F}_2 \)), the number of different transition probabilities for bit channel \( i \) doesn’t change with the block length \( N \). Though the value of the transition probability \( W^N_N(\mathbf{y}) \) and \( W^{2N}_N((1, 0) \otimes \mathbf{y}) \) may not be the same for different block lengths.

### IV. Evaluation of the Number of Equivalence Classes

In Section (3) the symmetry property is introduced which can be used to determine if any two outputs have the same transition probability for bit channel \( i \). In this section, we analyze the exact number of different transition probabilities for bit channel \( i \). In (1), this is the problem to find the number of equivalence classes.

#### A. Transformation of the Output Alphabet

In this section, we transform the study of the symmetry based on the output \( \mathbf{y} \in \mathcal{Y}^N \) to another output alphabet, which is more convenient to manipulate.

Before Theorem 2 is introduced, some new notations are needed. We use \( (\mathcal{Y}^N, A, \cdot) \) to denote the problem of studying the symmetry based on the transition probability in equation (15). Denote \( S_1 = |\text{SELF}(\mathcal{Y})| \) and \( S_2 = |\text{SYMM}(\mathcal{Y})| \). Remember that \( \text{SELF}(\mathcal{Y}) \) and \( \text{SYMM}(\mathcal{Y}) \) are defined in (7) and (8), respectively. Consider the received vector \( \mathbf{y} \in \mathcal{Y}^N \) and we count the number of occurrences of all possible symbols within this received vector. Suppose \( q_1 \) is the number of occurrences of the \( t \)th symbol of \( \text{SELF}(\mathcal{Y}) \) and \( e_t \) is the number of occurrences of the \( t \)th symbol of \( \text{SYMM}(\mathcal{Y}) \). Without loss of generality, assume the corresponding symmetric symbol of the \( t \)th symbol is its \((S_2 + 1 - t)\)th symbol.

\[
W(s_t|0) = W(s_{S_2+1-t}|1), s_t \in \text{SYMM}(\mathcal{Y}), 1 \leq t \leq S_2
\]

Define a new mapping as

\[
y' = (q_1, ..., q_{S_1}, e_1, ..., e_{S_2})
\]

It’s seen that \( y'(\mathbf{y}) \) is well defined in the sense that for any \( \mathbf{y} \), the mapping \( y'(\mathbf{y}) \) is unique. Define a new set corresponding to \( \mathcal{Y}^N \) as

\[
\text{SELF}(y') = \{ \mathbf{y}' | \mathbf{y} \in \mathcal{Y}^N \}
\]

\[
\text{SYMM}(y') = \mathcal{Y}' \setminus \text{SELF}(y')
\]

If \( \mathbf{z} = (q_1, ..., q_{S_1}, e_1, ..., e_{S_2}) \in \mathcal{Y}' \), define * as

\[
0 * \mathbf{z} = \mathbf{z}
\]

\[
1 * \mathbf{z} = (q_1, ..., q_{S_1}, e_{S_2}, ..., e_1)
\]

Obviously

\[
\text{SELF}(\mathcal{Y}') = \{ \mathbf{z} \in \mathcal{Y}' | \mathbf{z} = 1 * \mathbf{z} \}
\]
Corollary 4: From the definition of the set $\mathcal{Y}$, the cardinality of it, $|\mathcal{Y}|$, is the number of solutions to the following equations:

$$\sum_{1 \leq t \leq s_2} e_t + \sum_{1 \leq t \leq s_1} q_t = N$$

(29)

The number of solutions to this equation is $|\mathcal{Y}| = \binom{N + s_i + s_j - 1}{s_i + s_j - 1}$, which is proved in the Appendix. The size of the set $\text{SELF}(\mathcal{Y})$, $|\text{SELF}(\mathcal{Y})|$, is the number of solutions to the following equations:

$$\sum_{1 \leq t \leq s_2} e_t + \sum_{1 \leq t \leq s_1} q_t = N$$

$$e_t = e_{s_1+1-t}, 1 \leq t \leq s_2$$

$$e_t \geq 0, q_t \geq 0$$

(30)

The number of solutions of (30), which we will denote as $N\text{F}(S_1, S_2, N)$ is: when $S_1 \neq 0$

$$|\text{SELF}(\mathcal{Y})| = \sum_{0 \leq r \leq N/2} \binom{r + s_2/2 - 1}{s_2/2 - 1} \binom{N - 2r + s_1 - 1}{s_1 - 1}$$

(31)

and when $S_1 = 0$

$$|\text{SELF}(\mathcal{Y})| = \binom{N/2 + s_2/2 - 1}{s_2/2 - 1}$$

(32)

Note that, when $N = 1$, $|\text{SELF}(\mathcal{Y})| = 0$. The proof of the solution is omitted in this paper due to the space limit. And the size of the set $\text{SYM}(\mathcal{Y})$ is

$$|\text{SYM}(\mathcal{Y})| = |\mathcal{Y}| - |\text{SELF}(\mathcal{Y})|$$

(33)

which will be denoted as $N\text{M}(S_1, S_2, N)$.

Applying Corollary 4 to the BSC channel, we can obtain the following corollary.

Corollary 5: If $S_1 = 0$, $S_2 = 2$, then $|\text{SELF}(\mathcal{Y})| = 1$, $|\text{SYM}(\mathcal{Y})| = N$.

B. Evaluation of the Number of Equivalence Classes

In this section, we use the transformed output alphabet in Section IV-A to evaluate the exact number of equivalence classes. We first provide the following theorem.

Theorem 4: For bit channel $i$, we denote the number of the remaining rows of the generator matrix $G_N$ as $a = N - i$ ($N = 2^k$). Let $a'$ be any number in the range $N \geq a' > a$ and $a'' = 2^k$ for some number $k'$. Denote the number of different transition probabilities of bit channel $i$ as $\text{NC}(N, i, S_1, S_2)$, where $S_1 = |\text{SELF}(\mathcal{Y})|$ and $S_2 = |\text{SYM}(\mathcal{Y})|$. Then the number of different transition probabilities for bit channel $i$ is equivalent to $\text{NC}(N, i, S_1, S_2) = \text{NC}(a', i - (N - a'), S_3, S_4)$, where $S_3 = N\text{F}(S_1, S_2, N/a')$, $S_4 = N\text{M}(S_1, S_2, N/a')$.

Proof: We know

$$F^\otimes n = F^\otimes n-k' \otimes F^\otimes k'$$

(34)

The last $a'$ rows are $(1, 1, ... 1) \otimes F^\otimes k'$. Define vector $z(y) \in \mathcal{Y}^{a''}$ as

$$z(y)\{j\} = y'(y\{j, j+a', j+2a'..j+N-a'\}), 0 < j < a'$$

(35)

where $y$ is the output vector. Here $y\{j\}$ means the $j$th entry of $y$. Look at the definition of $\mathcal{Y}$, it’s seen that $z$ is surjective. For any given $y$ we can get a unique $z(y)$. Also, from the definition of $y'$, $W_N(\mathcal{Y})$ can be calculated from $z(y)$. Observe

$$z([(1,1,...,1) \otimes v] \cdot y) = z(y) \ast v$$

(36)

where $v \in \mathcal{R}(A(a', i - (N - a'), i))$. Then $\mathcal{Y}^{a''}, A(N, i, \cdot)$ is equivalent to $\mathcal{Y}^{a''}, A(a', i - (N - a'), \cdot)$. So we get $\text{NC}(N, i, S_1, S_2) = \text{NC}(a', i - (N - a'), S_3, S_4)$.

Theorem 5:

$$\text{NC}(N, 0, S_1, S_2) \leq \binom{N + S_2/2 + S_1 - 1}{S_2/2 + S_1 - 1}$$

(37)

Obviously, this is not the number of transition probabilities for any bit channel $i$. However, it can be used to deduce the number of transition probabilities for some bit channels.

Proof: Since $\text{rank}(A(N, 0)) = N$, obviously:

$$R(A(N, 0)) = F_2^N$$

(38)

Define $S \in \text{SYM}(\mathcal{Y})$ as a set that satisfies $|S| = S_2/2$, and $\forall y_1, y_2 \in S, 1 \cdot y_1 \neq y_2$. Then for any $y \in \mathcal{Y}^{a''}$, we can have a $v = u \cdot y, u \in F_2^N, v \in \mathcal{Y}^{a''}$, so that elements of $v$ will only choose from $S \cup \text{SELF}(\mathcal{Y})$. Since any element has $S_2/2 + S_1$ choices and the order doesn’t matter, so there’s total $\binom{S_2/2 + S_1}{S_2/2 + S_1}$ cases at most.

For some cases, we can have more accurate results than Theorem 6. Define $0 \cdot w_i = \log W(y_i|0), 1 \cdot w_i = \log W(1 \cdot y_i|0)$, and $D_i = 1 \cdot w_i - 0 \cdot w_i$, then we have

Theorem 6: If $D_i$s are all different for all $i$, then

$$\text{NC}(N, 0, S_1, S_2) = \binom{N + S_2/2 + S_1 - 1}{S_2/2 + S_1 - 1}$$

(39)

Proof:

$$\log(W^N(y|a)) = \sum_{1 \leq i \leq N} a_i \cdot w_i, a \in F_2^N$$

(40)

$$= \log(W^N(y|0^N)) + \sum_{1 \leq i \leq N} a_i D_i$$

(41)

If $y$ and $v$ are probability equivalent, there exists $u \in F_2^N$ so that $W^N(u \cdot y|0) = W^N(v|0)$. Define $v_2 = u \cdot y$. We can see that $v_2$ and $v$ must have the same set of $\{D_i\}$, which means $v_2$ must be a permutation of $v$. This shows that the symmetric properties have been discovered (Theorem 1 and Theorem 2) are enough to find all equivalent output vectors. So the less equal sign in Theorem 5 can be replaced by an equal sign.

Also, if bit channel $i$ can be converted to a channel with $i = 0$, then we can have an exact evaluation.

Corollary 6: If $S_1 = 0$, $S_2 = 2$, then for $i = N - a$, where $a$ is power of 2,

$$\text{NC}(N, i, S_1, S_2) = \text{NC}(a, 0, 1, N/a) = \binom{a + N}{N/2}$$

(42)

This is a numeric result on BSC channels. It provides the exact number of different transition probabilities for bit channel $i$. 
For $N = 2^n$, we choose $a = 2^{(n/2)}$ to calculate the number of equivalence classes. When $n$ is even $\text{NC}(N, i, 0, 2) =$ $\left(\frac{N + 2a}{N}\right) = \left(\frac{2\sqrt{N}}{N}\right) \left(\frac{N}{\sqrt{2}}\right) \approx \left(\frac{3}{2}\right)\sqrt{N}$ $2^{2N}$ $\approx 2^{2N}$ (43) and when $n$ is odd $\text{NC}(N, i, 0, 2) =$ $\left(\frac{N + 2a}{N}\right) = \left(\frac{2\sqrt{N}}{N}\right) \left(\frac{N}{\sqrt{2}}\right) \approx \left(\frac{2\sqrt{N}}{N}\right) \left(\frac{N}{\sqrt{2}}\right) = 2\sqrt{N}$ 2 $\approx 2\sqrt{N}$ (44) Although these are quite large numbers, they are still much smaller than $2^i = 2^{N-a} = 2^{N-2(n/2)}$ obtained using the symmetry in (1).

V. NUMERICAL RESULTS

To test Corollary 6 we numerically calculated the number of different transition probabilities $W_N^{(i)}(y)$ for all $y \in \mathbb{F}_2^N$. The numerically calculated results match the numbers calculated with Corollary 6 up to $i = 16$. For larger $i$, the time consumed to calculate the number of different transition probabilities is too long to be practical. From the table, we see for the BSC channels, the number of equivalence classes can be greatly smaller than $2^i$ for the $i$th channel.

| $i$ | Result | $i$ | Result |
|-----|--------|-----|--------|
| 2   | 3      | 128 | 129    |
| 3   | 3      | 192 | 2,145  |
| 4   | 5      | 224 | 58,905 |
| 6   | 6      | 240 | 7.35 \cdot 10^5 |
| 7   | 5      | 248 | 7.35 \cdot 10^5 |
| 8   | 9      | 252 | 58,905 |
| 12  | 15     | 254 | 2,145  |
| 14  | 15     | 255 | 129    |
| 15  | 9      | 256 | 257    |
| 16  | 17     | 384 | 8,385  |
| 24  | 45     | 448 | 8.14 \cdot 10^5 |
| 28  | 70     | 480 | 7.69 \cdot 10^7 |
| 30  | 45     | 496 | 6.71 \cdot 10^5 |
| 31  | 17     | 504 | 7.69 \cdot 10^7 |
| 32  | 33     | 508 | 8.14 \cdot 10^5 |
| 48  | 153    | 510 | 8,385  |
| 56  | 495    | 511 | 257    |
| 60  | 495    | 512 | 513    |
| 62  | 153    | 768 | 33,153 |
| 63  | 33     | 896 | 1.21 \cdot 10^7 |
| 64  | 65     | 960 | 1.2 \cdot 10^{10} |
| 96  | 561    | 992 | 7.68 \cdot 10^5 |
| 112 | 4,845  | 1,008 | 7.68 \cdot 10^5 |
| 120 | 12,870 | 1,016 | 1.2 \cdot 10^{10} |
| 124 | 4,845  | 1,020 | 1.21 \cdot 10^7 |
| 126 | 561    | 1,022 | 33,153 |
| 127 | 65     | 1,023 | 513    |

VI. CONCLUSION

In this paper, we present a frame of analysis to fully explore the symmetries of polar codes, which is used to analyze the smallest number of different transition probabilities for each bit channel. For some of the bit channels, an exact evaluation of the number of different transition probabilities (or the number of equivalence classes) is provided. The analysis for any given bit channel is still under investigation.

APPENDIX

Lemma 1: The number of solutions to this equation
\[
\sum_{1 \leq t \leq S_2} e_t + \sum_{1 \leq t \leq S_1} q_t = N
\]
\[e_t \geq 0, q_t \geq 0\]
is \(N + S_1 + S_2 - 1\).

Proof: The equation can be transformed into
\[
\sum_{1 \leq t \leq S_1 + S_2} w_t = N + S_1 + S_2
\]
\[w_t \geq 1\]
which is equivalent to
\[
1 \leq p_1 \leq N + S_1 + S_2 - 1
\]
\[1 \leq t \leq S_1 + S_2 - 1\]
\[p_{t+1} \neq p_{t+2}\]
\[p_1 < p_2 < \ldots < p_{S_1+S_2-1}\]
Then $w_t$ can be seen as
\[
w_1 = p_1
\]
\[w_2 = p_2 - p_1\]
\[
\ldots
\]
\[
w_{S_1 + S_2 - 1} = p_{S_1 + S_2 - 1} - p_{S_1 + S_2 - 2}
\]
\[w_{S_1 + S_2} = N + S_1 + S_2 - p_{S_1 + S_2 - 1}\]
The number of solutions to equation (47) is equivalent to taking $S_1 + S_2$ balls out of $N + S_1 + S_2 - 1$ balls without considering the order. Thus the lemma is proven.

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] E. Sasoglu, E. Telatar, and E. Arikan, “Polarization for Arbitrary Discrete Memoryless Channels,” Online: [http://arxiv.org/pdf/0908.0302v1.pdf](http://arxiv.org/pdf/0908.0302v1.pdf)
[3] R. Mori and T. Tanaka, “Non-Binary Polar Codes using Reed-Solomon Codes and Algebraic Geometry Codes,” in IEEE Information Theory Workshop (ITW), 2010, pp. 1–5.
[4] A. G. Saheli and S. S. Pradhan, “Multilevel Polarization of Polar Codes Over Arbitrary Discrete Memoryless Channels,” in 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), September 2011, pp. 1718–1725.
[5] R. Mori and T. Tanaka, “Performance and Construction of Polar codes on Symmetric Binary-Input Memoryless Channels,” in IEEE International Symposium on Information Theory, June 2009, pp. 1496–1500.
[6] R. Pedarsani, S. Hassani, I. Tal, and I. Telatar, “On the Construction of Polar Codes,” in IEEE International Symposium on Information Theory Proceedings (ISIT), 2011, pp. 11–15.
[7] P. Trifonov, “Efficient Design and Decoding of Polar Codes,” IEEE Transactions on Communications, vol. 60, no. 11, pp. 3221–3227, November 2012.
[8] I. Tal and A. Vardy, “How to Construct Polar Codes,” Online: [http://arxiv.org/abs/1304.3850](http://arxiv.org/abs/1304.3850)
[9] S. B. Korada, E. Sasoglu, and R. Urbanke, “Polar Codes: Characterization of Exponent, Bounds, and Constructions,” IEEE Transactions on Information Theory, vol. 56, no. 12, pp. 6253–6264, December 2010.