Channel Combining for Nonstationary Polarization on Erasure Channels

Hanwen Yao*, Hessam Mahdavifar†, Arman Fazeli*, and Alexander Vardy*
*Department of Electrical & Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA
†Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
Emails: hwyao@ucsd.edu, hessam@umich.edu, afazelic@ucsd.edu, avardy@ucsd.edu

***Abstract*—The problem of channel polarization for an arbitrary sequence $\{W_i\}_{i=0}^{n-1}$ of n independent channels, referred to as a *non-stationary* sequence of channels, is considered. Also, each of the channels is used only once for communication. We consider a general framework for polarization of non-stationary channels and aim at optimizing the framework toward obtaining the *best* polarization. This framework includes permuting channels before Arıkan's pairwise channel combining operations are applied at each polarization level and skipping certain combining operations. We define an explicit optimization problem with the objective of finding the best permutation and indices of skipped operations in order to minimize a certain measure of polarization in one-level polarization. We then provide a complete solution to this optimization problem in the case of non-stationary binary erasure channels (BECs). We also propose a greedy method for polarizing non-stationary BECs, based on our solution for one-level polarization. Numerical results confirm the superiority of our method, in terms of various performance metrics, for constructing polar codes in certain non-stationary settings compared to prior work.**

I. Introduction

Polar codes were introduced by Arıkan in [1]. They are the first family of codes for the class of binary-input symmetric discrete memoryless channels that are provable to be capacity-achieving with low encoding and decoding complexity [1]. In the past decade, channel polarization and polar coding paradigm have been extended to more general scenarios including channels with non-binary discrete inputs [2]–[4], asymmetric channels [5], channels with memory [6], [7], and, more recently, non-stationary channels [8]–[10]. In this paper, we consider the problem of non-stationary channel polarization and design methods towards optimal channel combining strategies.

In a non-stationary communication setting, an arbitrary sequence of independent channels $\{W_i\}_{i=0}^{n-1}$ is given. A coherent communication over these channels is assumed, i.e., the sequence is assumed to be completely known to both the transmitter and the receiver. This sequence is referred to as a non-stationary sequence of channels or simply non-stationary channels [8], [9]. The non-stationary channel coding scenario is then described as follows. It is assumed that each channel W_i is used once, i.e., a code with block length n is to be constructed and then the i-th coded bit is transmitted through W_i.

Besides the theoretical motivation to study polar coding and channel polarization beyond the scenario originally considered by Arıkan [1], there are practical applications indicating non-stationary scenarios are emerging in wireless systems [9]. Also, in data storage systems such as resistive memories, channel coding over non-stationary channels becomes relevant [10].

The channel polarization problem for non-stationary channels was first considered in [8]. In particular, it is shown in [8] that polarization happens by applying Arıkan's polarization transform. However, the proposed proof method in [8] is not powerful enough to conclude anything about the *speed of polarization* and, consequently, an achievability scheme. In [9] several modifications to Arıkan's channel polarization transform are proposed leading to polarization of non-stationary channels while lower bounding the speed of polarization. Then, polar coding schemes are constructed that achieve the average capacity of non-stationary channels while bounding the finite-length scaling exponent [9]. The modifications to Arıkan's channel polarization transform introduced in [9] include permuting the channels and skipping certain channel combining operations. In another related work [10], the problem of finding the *best* permutation for non-stationary channels is studied. As opposed to the approach in [9], the permutation is only applied once in [10] before applying the polarization transform. Various numerical results are shown in [10] indicating that the bit-reversal permutation leads to a better performance for the constructed polar code compared to the ordered permutation, where the channels are ordered according to their capacity, and several randomly selected permutations.

In this paper we consider a general framework for channel polarization and polar coding in non-stationary settings. This framework includes permutations of channels at each polarization level and skipping channel combining operations. In this work, we aim at optimizing the framework toward obtaining the *best* polarization. Our paper can be regarded as a first step towards the optimal non-stationary polar code construction under the considered framework. In order to measure how polarized a non-stationary sequence of channels is, we suggest to use a certain metric which we call the *polarization measure*. Then we define an explicit optimization problem with the objective of finding the best permutation and indices of skipped operations in order to minimize the polarization measure in one level of polarization. In this paper, we provide a complete solution to this optimization problem in the case of non-stationary BECs. We also propose a greedy method for polarizing non-stationary BECs, based on our solution for one-level polarization. We then show that in certain settings, our greedy construction outperforms the construction based on the bit-reversal permutation, proposed in [10], in terms of various performance metrics.

II. Construction of Non-Stationary Polar Codes

In conventional polar coding [1], the polar transform is a channel combining operation that takes two independent copies of a binary memoryless symmetric (BMS) channel W, and generates two new BMS channels $\{W^-, W^+\}$. This operation can be also applied to a pair of two different channels as well. Let $W_1 : \{0,1\} \to \mathcal{Y}_1$ and $W_2 : \{0,1\} \to \mathcal{Y}_2$ be two independent BMS channels. The polar transform operation can take W_1 and W_2 as a pair of inputs, and generate two new BMS channels

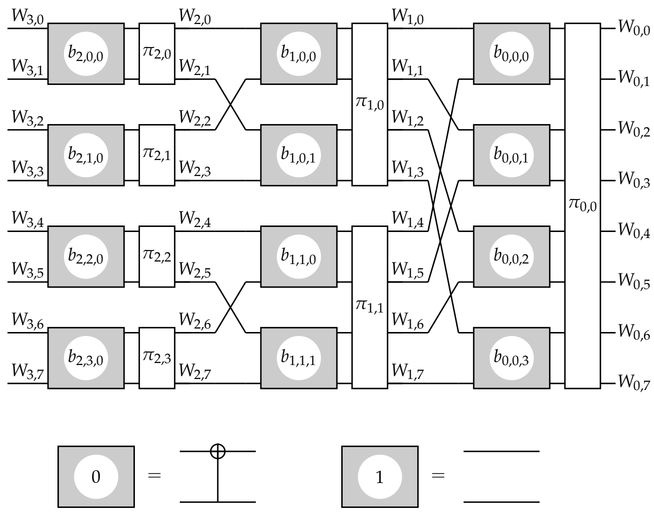

Figure 1. Construction framework for a length-8 non-stationary polar code.

$\langle W_1, W_2\rangle^-$ and $\langle W_1, W_2\rangle^+$ where

$$\langle W_1, W_2\rangle^-(y_1, y_2|u_1) = \frac{1}{2}\sum_{u_2\in\{0,1\}} W_1(y_1|u_1\oplus u_2)W_2(y_2|u_2),$$

$$\langle W_1, W_2\rangle^+(y_1, y_2, u_1|u_2) = \frac{1}{2}W_1(y_1|u_1\oplus u_2)W_2(y_2|u_2),$$

for all $y_1 \in \mathcal{Y}_1$, $y_2 \in \mathcal{Y}_2$, and $u_1, u_2 \in \{0,1\}$.

Now, suppose that an arbitrary sequence of non-stationary BMS channels $\{W_i\}_{i=0}^{n-1}$ is given. We refer to this sequence as a non-stationary sequence of BMS channels or simply non-stationary channels, keeping in mind that the channels are BMS. As mentioned earlier, it is shown in [8] that polarization happens by applying Arıkan's polarization transform recursively to this sequence. A modified scheme for non-stationary polarization is proposed in [9] to construct capacity-achieving codes. Later, similar modifications, e.g., permuting channels, is explored in [10] with a different objective of *optimizing* the code performance. To summarize these construction methods, we next describe a general framework that we follow in this paper for non-stationary channel polarization and polar coding.

Let $\{W_i\}_{i=0}^{n-1}$ be a sequence of n independent BMS channels where n is a power of 2. For $0 \leqslant m \leqslant \log_2(n)$, let $\{W_{m,i}\}_{i=0}^{n-1}$ denote the sequence of channels after m levels of polarization. The polarization is initialized at level 0 by setting $W_{0,i} = W_i$ for $0 \leqslant i \leqslant n-1$. Then a permutation $\pi_{0,0}$ is applied to $\{W_{0,i}\}_{i=0}^{n-1}$ followed by applying the polar transform channel combining operation or skipping it for all the pairs of consecutively indexed channels in order to obtain channels at level 1. Note that not all the pairs of channels are necessarily combined. A binary sequence $\{b_{0,0,i}\}_{i=0}^{n/2-1}$ of length $n/2$ is used to indicate the indices where the channel combining operation is skipped. The three sub-indices of b denote the level of polarization, the index of the sub-block, explained later, and the index of the channel pair within this sub-block. For $0 \leqslant k < n/2$, if $b_{0,0,k} = 0$, then the polar transform is applied to the pair of channels $W_{0,\pi_{0,0}(2k)}$ and $W_{0,\pi_{0,0}(2k+1)}$, and the resulting channels are indexed as $W_{1,k}$ and $W_{1,n/2+k}$, i.e.,

$$W_{1,k} = \langle W_{0,\pi_{0,0}(2k)}, W_{0,\pi_{0,0}(2k+1)}\rangle^-,$$
$$W_{1,n/2+k} = \langle W_{0,\pi_{0,0}(2k)}, W_{0,\pi_{0,0}(2k+1)}\rangle^+$$

If $b_{0,0,k} = 1$, then this combining operation is skipped and we have

$$W_{1,2k} = W_{0,\pi_{0,0}(2k)}, \quad W_{1,n/2+k} = W_{0,\pi_{0,0}(2k+1)}.$$

In the second level of polarization, the same procedure is applied to the sub-blocks $\{W_i\}_{i=0}^{n/2-1}$ and $\{W_i\}_{i=n/2}^{n-1}$ separately. In general, at level m, the sequence $\{W_{m,i}\}_{i=0}^{n-1}$ is split into 2^m sub-blocks of consecutively indexed channels, each of length $n/2^m$. The sub-blocks are indexed by numbers from 0 to (2^m-1). For instance, consider the j-th sub-block $\{W_{m,i}\}_{i=(n/2^m)j}^{(n/2^m)(j+1)-1}$ at the m-th level of polarization. The channels in this sub-block are permuted using the permutation $\pi_{m,j}$. Note that $\pi_{m,j}$ is applied only on channels indexed from $(n/2^m)j$ to $((n/2^m)(j+1)-1)$. Then for $0 \leqslant k < (n/2^{m+1})$, if $b_{m,j,k} = 0$, we have

$$W_{m+1,(n/2^m)j+k} = \langle W_{m,\pi_{m,j}((n/2^m)j+2k)}, W_{m,\pi_{m,j}((n/2^m)j+2k+1)}\rangle^-,$$
$$W_{m+1,(n/2^m)j+(n/2^{m+1})+k} = \langle W_{m,\pi_{m,j}((n/2^m)j+2k)}, W_{m,\pi_{m,j}((n/2^m)j+2k+1)}\rangle^+.$$

Otherwise, $b_{m,k} = 1$ and this combining operation is skipped, i.e.,

$$W_{m+1,(n/2^m)j+k} = W_{m,\pi_{m,j}((n/2^m)j+2k)},$$
$$W_{m+1,(n/2^m)j+(n/2^{m+1})+k} = W_{m,\pi_{m,j}((n/2^m)j+2k+1)}.$$

This general construction framework is illustrated for length-8 non-stationary polar coding in Figure 1.

Now, given the aforementioned general framework for non-stationary polarization, the main question is how to *optimize* it for a given sequence of non-stationary channels, i.e., what are *the best* choices for the permutations $\pi_{m,j}$'s and the skipped operations indicated by $b_{m,j,k}$'s? Before proceeding further, we need to define a metric that can be used to measure how polarized a sequence of channels is. To this end, we introduce a polarization measure for a set of BMS channels in the next section. That measure will serve as our primary metric for designing *optimal* schemes for non-stationary polarization in this paper.

III. A Measure of Polarization

In this section, we introduce a polarization measure for a set of BMS channels. Polarization is, in general, an asymptotic notion. However, in a non-asymptotic regime, a measure to quantify the level of polarization in a finite set of channels is needed. One such a measure, that we use throughout this paper, is defined next.

Definition 1. *For a set of BMS channels $\{W_i\}_{i=0}^{n-1}$, and $\alpha, \beta \in (0,1]$, the* (α,β)-polarization measure *of this set, denoted by $M_{\alpha,\beta}(\{W_i\}_{i=0}^{n-1})$, is defined as*

$$M_{\alpha,\beta}(\{W_i\}_{i=0}^{n-1}) \overset{\text{def}}{=} \frac{1}{n}\sum_{i=0}^{n-1} Z(W_i)^\alpha(1-Z(W_i))^\beta,$$

where $Z(W_i)$ is the Bhattacharyya parameter of the channel W_i.

Accordingly, for a fixed α and β, we say a set of channels is *more polarized* than another set, if the former has a smaller (α,β)-polarization measure than the latter.

In this paper, we focus only on the case with $\alpha = \beta = 1$ and adopt $M_{1,1}$ as the measure of polarization. Therefore, for the sake of notation convenience, we simply use M to denote $M_{1,1}$ through the rest of this paper. Note that the expression $Z(W)(1 - Z(W))$ first appeared in [1] in the proof of channel polarization theorem. The aforementioned polarization measure also appears in several works studying finite-length scaling exponent of polar codes, see, e.g., [9], [11]. In the remainder of this section, we justify our choice of this polarization measure from another perspective.

It can be also observed that for a non-stationary sequence of channels, this polarization measure gives an upper bound on the number of *unpolarized* channels in the sequence, specified as follows. Let $\epsilon > 0$ be fixed. Then any channel W_i with $\epsilon < Z(W_i) < 1 - \epsilon$, is referred to as an unpolarized channel with respect to ϵ. The following lemma relates the polarization measure with the number of unpolarized channels corresponding to a parameter ϵ.

Lemma 1. *For a set of BMS channels $\{W_i\}_{i=0}^{n-1}$, we have*

$$\frac{|\{i : \epsilon < Z(W_i) < 1 - \epsilon\}|}{n} < \frac{M(\{W_i\}_{i=0}^{n-1})}{\epsilon(1-\epsilon)}.$$

Proof. Note that for each i with $\epsilon < Z(W_i) < 1 - \epsilon$, we have

$$M(W_i) > \epsilon(1-\epsilon).$$

The proof then follows by this together with the definition of $M(\{W_i\}_{i=0}^{n-1})$. □

Roughly speaking, Lemma 1 implies that if the polarization measure for a collection of channels is close to 0, the fraction of unpolarized channels with respect to a fixed ϵ will also be close to 0.

IV. An Optimal Solution for One-Level Polarization of BECs

Given the polarization measure in Section III and motivated by the problem of *optimizing* non-stationary polarization discussed in Section II, we pose the following optimization problem: Given an arbitrary sequence of non-stationary channels, how to pick the collection of permutations π's and skipping operations specified by b's to minimize the polarization measure for the resulting channels? Answering this question in general seems to be a formidable task. In this section, we provide a complete solution to this problem for the case with one-level polarization applied to a sequence of non-stationary BECs. We first give a precise description of the optimization problem for one-level polarization on BECs, and then provide the complete solution to it.

We start by providing a precise description of one-level polarization. For a sequence of channels $\{W_i\}_{i=0}^{n-1}$ with even n, one-level polarization refers to the operation that we first permute W_i's according to a permutation π and then combine a certain subset of the set of all consecutive channel pairs, i.e., the pairs $\{W_{\pi(2k)}, W_{\pi(2k+1)}\}$, for $0 \leqslant k < n/2$, and apply the channel combining operation to them resulting in a new sequence of non-stationary channels. Note that the operation that results in obtaining $\{W_{1,i}\}_{i=0}^{n-1}$ from $\{W_{0,i}\}_{i=0}^{n-1}$, as specified in Section II, is one-level polarization.

For non-stationary channels, different choices of π and skipping operations specified by $\{b_i\}_{i=0}^{n/2-1}$ may result in different outcomes of one-level polarization. We aim at finding the *best* choices for them in the sense that they minimize the polarization measure of $\{W_{1,i}\}_{i=0}^{n-1}$. Next, this optimization problem is stated, which we aim to solve:

Optimization Problem A. *Let $\{W_i\}_{i=0}^{n-1}$ be an arbitrary sequence of non-stationary BECs, where n is even. Let $\{W_{1,i}\}_{i=0}^{n-1}$ be the sequence of channels after one-level polarization. What is the permutation π and the binary sequence $\{b_i\}_{i=0}^{n/2-1}$ that minimize the polarization measure $M(\{W_{1,i}\}_{i=0}^{n-1})$?*

We first show that in Optimization Problem A, it's always better to apply the polar transform to all the channel pairs. In other words, in the solution to Optimization Problem A, we have $b_i = 0$ for all $0 \leqslant i < n/2$. This is shown in the following proposition.

Proposition 1. *For any two BECs W_1 and W_2, we have*

$$M(W_1, W_2) \geqslant M(\langle W_1, W_2\rangle^-, \langle W_1, W_2\rangle^+).$$

Proof. Let the erasure probabilities of W_1 and W_2 be denoted by z_1 and z_2, respectively. Then we have

$$\begin{aligned} M(W_1, W_2) - M(\langle W_1, W_2\rangle^-, \langle W_1, W_2\rangle^+) \\ = z_1 z_2 (1 - z_1)(1 - z_2) \geqslant 0, \end{aligned}$$

which completes the proof. □

By Proposition 1, one can observe that for any two BECs, applying the polar transform to them is always better than not applying it when the goal is to minimize the polarization measure. Therefore, to minimize the polarization measure after one-level polarization, we should always include all the BECs for channel combining operations. Therefore, in order to find the best permutation, i.e., the solution to Optimization Problem A, we assume all the BECs are paired for channel combination.

Note that there are many permutations that are equivalent to each other in one-level polarization. For instance, when we split a sequence of BECs $\{W_i\}_{i=0}^{n-1}$ into $n/2$ channel pairs, the order of channels within each pair doesn't change the resulting channels. Separately, one can swap the channel pair $\{W_1, W_2\}$ with the channel pair $\{W_3, W_4\}$, and yet obtain the same set of polarized channels after one-level polarization. In that sense, we call two permutations *equivalent*, if we can obtain one of them from the other one, by flipping the channel indices that are paired, and/or changing the order of the channel index pairs. It can be observed that this is a well-defined equivalence relation. Hence, one can partition S_n, the set of all permutations on $\{0, 1, \cdots, n-1\}$, into equivalence classes of size $\left(\frac{n}{2}\right)! 2^{n/2}$ according to this equivalence relation. When searching for the optimal permutation as the solution to Optimization Problem A, it is sufficient to only search over the equivalence classes. In the prior work [10], there are similar discussions regarding the equivalence classes of permutations. However, the authors in [10] consider this for multiple levels of polarization, while we focus on one-level polarization in this section.

We are now ready to define the *proximity-to-half permutation*, that are crucial in solving Optimization Problem A, as follows.

Definition 2. *Given an arbitrary sequence of BECs* $\{W_i\}_{i=0}^{n-1}$ *with erasure probabilities* $z_0, z_1, \cdots, z_{n-1}$*, we call a permutation* $\pi \in S_n$ *a proximity-to-half permutation for this sequence, if for the permuted sequence of channels* $\{W_{\pi(i)}\}_{i=0}^{n-1}$*, the values*

$$\left|z_{\pi(0)} - 1/2\right|, \left|z_{\pi(1)} - 1/2\right|, \cdots, \left|z_{\pi(n-1)} - 1/2\right|$$

are in an ascending order.

Notice that for a set of BECs, if $z_i = z_j$ or $z_i = 1 - z_j$ for some $0 \leqslant i < j < n$, then there will be more than one permutations in S_n that satisfy the condition of being *proximity-to-half*. We will prove that all such permutations will give us the same polarization measure after one-level polarization, and that measure is also the minimum among all possible permutations. Consequently, this shows that the equivalence classes containing *proximity-to-half* permutations are the optimal ones, thereby providing a complete solution to Optimization Problem A.

Theorem 1. *Given an arbitrary sequence of BECs* $\{W_i\}_{i=0}^{n-1}$*, with n being even, a permutation* $\pi \in S_n$ *together with setting* $b_i = 0$*, for all* $0 \leqslant i < n/2$*, is an optimal solution for Optimization Problem A if and only if* π *is equivalent to a proximity-to-half permutation.*

The rest of this Section is devoted to the proof of Theorem 1, which is the main result of this paper. We first prove that Theorem 1 holds for the case with four BECs. Then we generalize our result into n BECs to complete the proof.

A. Proof for four BECs

Lemma 2. *Theorem 1 holds when* $n = 4$.

Proof. Suppose that we have four BECs W_0, W_1, W_2 and W_3 with erasure probabilities z_0, z_1, z_2 and z_3, respectively. Without loss of generality assume that

$$\left|z_0 - 1/2\right| \leqslant \left|z_1 - 1/2\right| \leqslant \left|z_2 - 1/2\right| \leqslant \left|z_3 - 1/2\right|. \quad (1)$$

Note that for one-level polarization of four BECs, we can partition S_4 into three equivalence classes, and only compare three representatives, one from each class. Denote the three representative permutations as π_1, π_2, and π_3. We can pick π_1 to be the identity permutation, π_2 that permutes the channels to be W_0, W_2, W_1, W_3, and π_3 that permutes the channels to be W_0, W_3, W_1, W_2. Also denote the polarization measures after one-level polarization following those three permutations as M_1, M_2 and M_3 respectively:

$$M_1 = M(\{\langle W_0, W_1\rangle^-, \langle W_0, W_1\rangle^+, \langle W_2, W_3\rangle^-, \langle W_2, W_3\rangle^+\})$$
$$M_2 = M(\{\langle W_0, W_2\rangle^-, \langle W_0, W_2\rangle^+, \langle W_1, W_3\rangle^-, \langle W_1, W_3\rangle^+\})$$
$$M_3 = M(\{\langle W_0, W_3\rangle^-, \langle W_0, W_3\rangle^+, \langle W_1, W_2\rangle^-, \langle W_1, W_2\rangle^+\})$$

To compare M_1 with M_2, notice that $2(M_1 - M_2)$ is factored as:

$$-\left[\left(z_1 - 1/2\right)^2 - \left(z_2 - 1/2\right)^2\right]\left[\left(z_0 - 1/2\right)^2 - \left(z_3 - 1/2\right)^2\right].$$

Hence $M_1 \leqslant M_2$, and the equality holds if and only if

$$\left|z_1 - 1/2\right| = \left|z_2 - 1/2\right|. \quad (2)$$

Similarly, $2(M_2 - M_3)$ is factored as:

$$-\left[\left(z_0 - 1/2\right)^2 - \left(z_1 - 1/2\right)^2\right]\left[\left(z_2 - 1/2\right)^2 - \left(z_3 - 1/2\right)^2\right].$$

So we have $M_2 \leqslant M_3$, and the equality holds if and only if

$$\left|z_0 - 1/2\right| = \left|z_1 - 1/2\right| \quad \text{or} \quad \left|z_2 - 1/2\right| = \left|z_3 - 1/2\right|. \quad (3)$$

Next, in order to complete the proof, we consider three different cases depending on whether (2) and/or (3) hold or not:

1) **Case 1: (2) doesn't hold.**
In this case, we have
$$\left|z_0 - 1/2\right| \leqslant \left|z_1 - 1/2\right| < \left|z_2 - 1/2\right| \leqslant \left|z_3 - 1/2\right|.$$
Hence, π_1 is the only proximity-to-half permutation. Then Theorem 1 holds here since we have $M_1 < M_2 \leqslant M_3$.
2) **Case 2: (2) holds but (3) doesn't hold.**
In this case, we have
$$\left|z_0 - 1/2\right| < \left|z_1 - 1/2\right| = \left|z_2 - 1/2\right| < \left|z_3 - 1/2\right|.$$
Hence, both π_1 and π_2 are proximity-to-half permutations, while π_3 is not. In this case, Theorem also 1 holds since we have $M_1 = M_2 < M_3$.
3) **Case 3, both (2) and (3) hold.**
In this case, we have
$$\left|z_0 - 1/2\right| = \left|z_1 - 1/2\right| = \left|z_2 - 1/2\right|$$
or
$$\left|z_1 - 1/2\right| = \left|z_2 - 1/2\right| = \left|z_3 - 1/2\right|.$$
Hence, π_1, π_2, and π_3 will all be proximity-to-half permutations. Then we have $M_1 = M_2 = M_3$, implying that Theorem 1 holds in this case as well.

□

B. Proof for n BECs

This section is devoted to the proof of Theorem 1 for the general even number n. We first prove that if a permutation $\pi \in S_n$ is not equivalent to any proximity-to-half permutation, then it cannot be an optimal solution to Optimization Problem A.

Suppose that π is not equivalent to any proximity-to-half permutation for one-level polarization. Then there should exist four channels W_i, W_j, W_k and W_l with erasure probabilities z_i, z_j, z_k and z_l, respectively, where

$$\left|z_i - 1/2\right| \leqslant \left|z_j - 1/2\right| < \left|z_k - 1/2\right| \leqslant \left|z_l - 1/2\right|,$$

and, also, W_i is paired with W_k and W_j is paired with W_l, after the permutation π is applied. We refer to such a couple of channel pairs as a *cross pair*, as illustrated in Figure 2. When we order the channels such that the values

$$\left|z_0 - 1/2\right|, \left|z_1 - 1/2\right|, \cdots, \left|z_{n-1} - 1/2\right|$$

are in an ascending order, there must be at least one cross pair, as described above, with $i < j < k < l$ in the channel pairing following π. Next, by invoking Lemma 2, we show that π cannot be optimal. Note that the permutation π can be altered such that the resulting channel pairing is altered only for those four channels from $\{\{W_i, W_k\}, \{W_j, W_l\}\}$ to $\{\{W_i, W_j\}, \{W_k, W_l\}\}$.

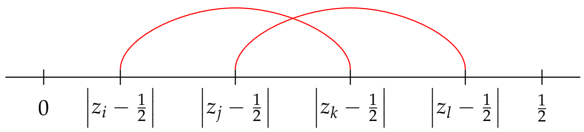

Figure 2. A cross pair in the channel pairing for a permutation that is not equivalent to any proximity-to-half permutation.

In that case, by Lemma 2, the altered permutation has a smaller polarization measure establishing that π cannot be optimal.

Next, we prove that if two permutations are both proximity-to-half, then their corresponding polarization measures are equal. Consider two proximity-to-half permutations π_1 and π_2. Then we have

$$\left|z_{\pi_1(i)} - 1/2\right| = \left|z_{\pi_2(i)} - 1/2\right|, \quad \text{for all } i = 0, 1, \cdots n-1$$

Let M_{ijkl} denote the polarization measure after one-level polarization for two channel pairings $\{\{W_i, W_j\}, \{W_k, W_l\}\}$, i.e.,

$$M_{ijkl} = M(\langle W_i, W_j\rangle^-, \langle W_i, W_j\rangle^+, \langle W_k, W_l\rangle^-, \langle W_k, W_l\rangle^+)$$

Then, as shown in the proof of Lemma 2, we have

$$M_{\pi_1(i)\pi_1(i+1)\pi_1(i+2)\pi_1(i+3)} = M_{\pi_2(i)\pi_2(i+1)\pi_2(i+2)\pi_2(i+3)}$$

for $i = 0, 2, \ldots, n-2$, where the indices are considered modulo n. Now, let M_{π_1} and M_{π_2} be the polarization measures after one-level polarization corresponding to permutations π_1 and π_2, respectively. Then we have

$$\begin{aligned} 2nM_{\pi_1} &= 4M_{\pi_1(0)\pi_1(1)\pi_1(2)\pi_1(3)} + 4M_{\pi_1(2)\pi_1(3)\pi_1(4)\pi_1(5)} \\ &\quad + \cdots + 4M_{\pi_1(n-2)\pi_1(n-1)\pi_1(0)\pi_1(1)} \\ &= 4M_{\pi_2(0)\pi_2(1)\pi_2(2)\pi_2(3)} + 4M_{\pi_2(2)\pi_2(3)\pi_2(4)\pi_2(5)} \\ &\quad + \cdots + 4M_{\pi_2(n-2)\pi_2(n-1)\pi_2(0)\pi_2(1)} \\ &= 2nM_{\pi_2}. \end{aligned}$$

Hence, $M_{\pi_1} = M_{\pi_2}$ and we conclude that all proximity-to-half permutations, and permutations in their equivalence classes, result in equal polarization measures after one-level polarization. Furthermore, this is the minimum value of polarization measure that can be obtained by one-level polarization. Hence, they are all optimal. This together with Proposition 1 provide a complete solution to Optimization Problem A.

V. A Greedy Construction

In Section IV, we provided a solution to Optimization Problem A, which concerns with one-level polarization, assuming all channels in the non-stationary sequence are BECs. However, finding the optimal solution for permutations and indices of skipped operations under the general construction framework in Section II still remains open. In this Section, we propose a greedy method toward optimizing non-stationary polarization, where we repeatedly use proximity-to-half permutations for every block in all polarization levels. We compare the resulting construction with the one proposed in [10] and show that our greedy construction is better in certain numerical settings.

Next, our proposed greedy method is described. Recalling the discussion in Section II, we first set $b_{i,j,k} = 0$ for all $0 \leqslant i < \log_2 n, 0 \leqslant j < 2^i$, and $0 \leqslant k < n/2^{i+1}$. Then we set $\pi_{i,j}$'s to be one of the proximity-to-half permutations for their corresponding blocks, for each $0 \leqslant i < \log_2 n$ and $0 \leqslant j < 2^i$. It is shown in Section IV that a proximity-to-half permutation minimizes the polarization measure for the underlying one-level polarization. Roughly speaking, our greedy method is based on the idea that we recursively use the proximity-to-half permutations that minimize the polarization measure at the level $i+1$ given prior polarization levels up to level i, for i between 0 and $\log_2 n$. Note that when there is more than one choices for the proximity-to-half permutation, we pick one of them arbitrarily.

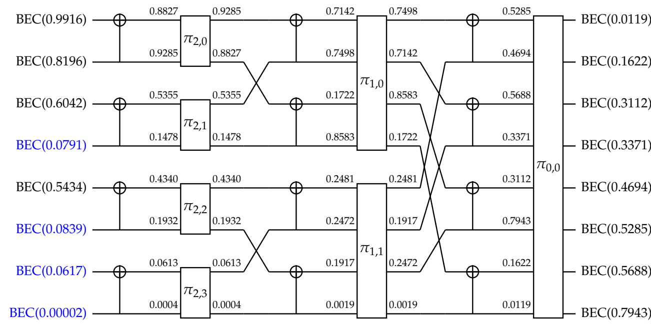

Figure 3. Greedy construction example.

An example of our greedy method is shown in Figure 3. In this example, a sequence of 8 non-stationary BECs with erasure probabilities shown on the right of Figure 3 is assumed. The greedy method with 3 levels of polarization results in a sequence of 8 channels with erasure probabilities shown on the left of Figure 3. In this example, the polarization measure of the resulting 8 channels is 0.1064. Now, suppose that a polar code of rate 1/2 needs to be constructed. Then, among these 8 channels, we choose 4 of them with the smallest erasure probabilities (colored blue in Figure 3) to obtain a rate 1/2 non-stationary polar code.

In [10], an alternative metric is used to *optimize* the construction of non-stationary polar codes. The optimization criterion in [10] is motivated from the code construction perspective. More specifically, the authors of [10] propose to maximize $\prod_{i\in\mathcal{I}} I(W_i)$, where $\mathcal{I}$ is the set of selected channel indices for code construction according to a certain given rate, and $I(W_i)$ is the symmetric capacity of bit channel W_i. Note that $I(W_i)$ is equal to the capacity of W_i when W_i is a BMS channel. We refer to this metric as the *sum-capacity* metric. Note that when the non-stationary channels are BECs, we can accurately compute the frame error rate (FER) of Arıkan's successive cancellation decoding, denoted by FER_{SC}, as $\text{FER}_{\text{SC}} = 1 - \prod_{i\in\mathcal{I}}(1 - Z(W_i))$. For the example shown in Figure 3, we compare our numerical results with the construction proposed in [10] that is based on the bit-reversal permutation. The comparison between the two methods, both used to construct codes of rate 1/2, is shown in Table I.

	polarization measure	sum-capacity metric	FER_{SC}
greedy	0.1064	3.7753	0.1082
bit-reversal	0.1072	3.7413	0.1254

Table I. Comparison between two constructions for rate 1/2 code given the channels shown in Figure 3.

It can be observed that in this example, our greedy construction outperforms the scheme suggested in [10] based on the bit-reversal permutation in terms of various metrics, i.e., our suggested polarization measure, the sum-capacity metric suggested in [10], and the FER_{SC}.

REFERENCES

[1] E. Arıkan, "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. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," *Proceedings of Information Theory Workshop (ITW)*, pp. 144–148, 2009.

[3] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7839–7857, 2013.

[4] T. C. Gulcu, M. Ye, and A. Barg, "Construction of polar codes for arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 309–321, 2018.

[5] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, 2013.

[6] E. Şaşoğlu and I. Tal, "Polar coding for processes with memory," *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pp. 225–229, 2016.

[7] R. Wang, J. Honda, H. Yamamoto, R. Liu, and Y. Hou, "Construction of polar codes for channels with memory," *Proceedings of Information Theory Workshop (ITW)*, pp. 187–191, 2015.

[8] M. Alsan and E. Telatar, "A simple proof of polarization and polarization for non-stationary channels," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4873–4878, 2016.

[9] H. Mahdavifar, "Polar coding for non-stationary channels," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6920–6938, 2020.

[10] M. Zorgui, M. E. Fouda, Z. Wang, A. M. Eltawil, and F. Kurdahi, "Non-stationary polar codes for resistive memories," *arXiv preprint arXiv:1904.08966*, 2019.

[11] S. H. Hassani, K. Alishahi, and R. L. Urbanke, "Finite-length scaling for polar codes," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 5875–5898, 2014.