Polar Coding for Fading Channels

Hongbo Si, O. Ozan Koyluoglu, and Sriram Vishwanath
Laboratory for Informatics, Networks, and Communications
The University of Texas at Austin
1 University Station, C0806, Austin, TX 78712
Email: {sihongbo,ozan}@mail.utexas.edu, sriram@austin.utexas.edu

***Abstract*—A polar coding scheme for fading channels is proposed in this paper. More specifically, the focus is on the Gaussian fading channel with a BPSK modulation, where the equivalent channel is modeled as a binary symmetric channel with varying cross-over probabilities. To deal with variable channel states, a coding scheme of hierarchically utilizing polar codes is proposed. In particular, by observing the polarization of different binary symmetric channels over different fading blocks, each channel use corresponding to a different polarization is modeled as a binary erasure channel such that polar codes could be adopted to encode over blocks. It is shown that the proposed coding scheme, without instantaneous channel state information at the transmitter, achieves the capacity of the corresponding fading binary symmetric channel.**

I. Introduction

Polar codes are the first family of provably capacity achieving codes for arbitrary symmetric binary-input discrete memoryless channels (B-DMC) with low encoding and decoding complexity [1] [2]. Channel polarization has then been generalized to arbitrary discrete memoryless channels with the same order of construction complexity and error probability behavior [3]. Moreover, polar codes are proved to be optimal for lossy compression with respect to binary symmetric source [4][5], and then further extended to larger source alphabet [6].

Polar coding is useful for channels with non-discrete inputs as well. By adopting polar codes as embedded codes, [7] shows that the capacity of additive exponential noise channel is achievable in high SNR region. Besides, a polar coding scheme achieving capacity for additive Gaussian noise channel is proposed in [8], which utilizes the polarization result for multiple access channel [9]. It has been shown that the approach of using a multiple access channel with a large number of binary-input users has much better complexity attributes than the one of using a single-user channel with large input cardinality.

In this paper, we propose a polar coding scheme for AWGN block fading channel with BPSK modulation and demodulation. (The model is similar to the binary input AWGN model analyzed in [10][11], but here with channel coefficients that vary according to a block fading model.) By adopting BPSK modulation and demodulation technique, additive Gaussian noise fading channel is converted into a binary symmetric channel (BSC) with finite set of transition probabilities according to the channel quality. The key intuition of the proposed scheme is based on observing the polarization characteristics of different BSCs. By hierarchically using polar codes, where the transmitter encodes over blocks, it can be proved that the designed coding scheme achieves the capacity of converted channel (which we refer to as the fading BSC model).

The rest of paper is organized as follows. After introducing the preliminary results on polar codes and problem background in Section II and III, respectively, the polar coding scheme for fading channels is stated and illustrated in Section IV. The paper concludes with a discussion section.

II. Polar Coding

The construction of polar code is based on the observation of channel polarization. Consider a binary-input discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} = \{0, 1\}$. Define

$$F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Let B_N be the bit-reversal operator defined in [1], where $N = 2^n$. By applying the transform $G_N = B_N F^{\otimes n}$ ($F^{\otimes n}$ denotes the n^{th} Kronecker power of F) to $u_{1:N}$, consider transmitting the encoded output $x_{1:N}$ through N independent copies of W. Then N new binary-input coordinate channels $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}$ are constructed, where for each $i \in \{1, \ldots, N\}$ the transition probability is given by

$$W_N^{(i)}(y_{1:N}, u_{1:i-1}|u_i) \triangleq \sum_{u_{i+1:N}} \frac{1}{2^{N-1}} W^N(y_{1:N}|u_{1:N}G_N).$$

Then, as N tends to infinity, the channels $\{W_N^{(i)}\}$ polarize to either noiseless or pure-noisy, and the fraction of noiseless channels is close to $I(W)$, the symmetric mutual information of channel W [1].

To this end, polar codes can be considered as G_N-coset codes with parameter $(N, K, \mathcal{A}, u_{\mathcal{A}^c})$, where $u_{\mathcal{A}^c} \in \mathcal{X}^{N-K}$ is frozen vector (can be set to all-zeros for symmetric channels [1]), and the information set $\mathcal{A}$ is chosen as a K-element subset of $\{1, \ldots, N\}$ such that the Bhattacharyya parameters satisfy $Z(W_N^{(i)}) \leq Z(W_N^{(j)})$ for all $i \in \mathcal{A}$ and $j \in \mathcal{A}^c$.

The decoder in polar coding scheme is successive cancelation (SC) decoder, which gives an estimate $\hat{u}_{1:N}$ of $u_{1:N}$ given knowledge of $\mathcal{A}$, $u_{\mathcal{A}^c}$, and $y_{1:N}$ by computing

$$\hat{u}_i \triangleq \begin{cases} 0, & \text{if } i \in \mathcal{A}^c, \\ d_i(y_{1:N}, \hat{u}_{1:i-1}), & \text{if } i \in \mathcal{A}, \end{cases}$$

in the order i from 1 to N, where

$$d_i(y_{1:N}, \hat{u}_{1:i-1}) \triangleq \begin{cases} 0, & \text{if } \frac{W_N^{(i)}(y_{1:N}, \hat{u}_{1:i-1}|0)}{W_N^{(i)}(y_{1:N}, \hat{u}_{1:i-1}|1)} \geq 1, \\ 1, & \text{otherwise.} \end{cases}$$

It has been proved that by adopting an SC decoder, polar coding achieves any rate $R < I(W)$ with a decoding error scaling as $O(2^{-N^\beta})$, where $\beta < 1/2$. Moreover, the encoding and decoding complexity of polar coding are both $O(N \log N)$.

III. System Model

Fading channels characterize the wireless communication channels, where the channel states vary over channel uses. Fading coefficients typically vary much slower than transmission symbol duration in practice. To this end, a block fading model is proposed, whereby the channel state is assumed to be constant over each coherence time interval, and follow a stationary ergodic process across fading blocks. In addition, we consider the practical scenario where the channel state information (CSI) is available at the decoder [12].

Consider the AWGN fading channel,

$$Y_{b,i} = H_{b,i} X_{b,i} + Z_{b,i}, \quad b = 1, \ldots, B, \ i = 1, \ldots, N, \quad (1)$$

where $Z_{b,i}$ is i.i.d. additive Gaussian noise with variance E_Z; $X_{b,i}$ is channel input with power constraint

$$\frac{1}{BN} \sum_{b=1}^{B} \sum_{i=1}^{N} x_{b,i}^2 \leq E_X;$$

$H_{b,i}$ is the channel gain; N is blocklength; and B is number of blocks. For this moment, $H_{b,i}$ are assumed to be constant within a block and follow an i.i.d. fading process over blocks. (That is, $H_{b,i} = H_b$, $\forall i$, and H_b is an i.i.d. random variable.) In particular, for the case of two channel states $\{h_1, h_2\}$, the distribution of H_b is given by $\Pr\{H_b = h_1\} \triangleq q_1$ and $\Pr\{H_b = h_2\} \triangleq q_2 = 1 - q_1$ for each fading block b.

Using BPSK modulation, any codeword produced by encoder is mapped to a signal with element in $\{-\sqrt{E_X}, +\sqrt{E_X}\}$. After utilizing a BPSK demodulation at the decoder, the equivalent channel can be formulated as a binary symmetric channel, with transition probability relating to channel states. More specifically, the converted channel is given by

$$\bar{Y}_{b,i} = \bar{X}_{b,i} \oplus \bar{Z}_{b,i}, \quad b = 1, \ldots, B, \ i = 1, \ldots, N, \quad (2)$$

where $\bar{X}_{b,i}$ and $\bar{Y}_{b,i}$ are both Bernoulli random variables representing channel input and output correspondingly; $\bar{Z}_{b,i}$ is i.i.d. channel noise, also distributed as Bernoulli random variable, but related to channel state. More precisely, if $H_{b,i} = h_s$, where $s \in \{1, 2\}$, then

$$\Pr\{\bar{Z}_{b,i} = 1\} = 1 - \Phi(h_s \sqrt{\text{SNR}}) \triangleq p_s, \quad (3)$$

where $\Phi(\cdot)$ is CDF of normal distribution and $\text{SNR} = E_X/E_Z$. In other words, the channel can be modeled as $W_s \triangleq \text{BSC}(p_s)$ with probability q_s, for $s \in \{1, 2\}$.

The ergodic capacity of the converted channel (fading BSC) is given by [12]

$$C_{\text{SI-D}} = q_1[1 - H(p_1)] + q_2[1 - H(p_2)], \quad (4)$$

where $H(\cdot)$ is the binary entropy function, and SI-D refers to channel state information at the decoder. The capacity achieving input distribution is uniform over $\{0, 1\}$. In this paper, we show a polar coding scheme achieving the capacity of converted fading channel with low encoding and decoding complexity, without having instantaneous channel state information at the transmitter (only the statistical knowledge is assumed).

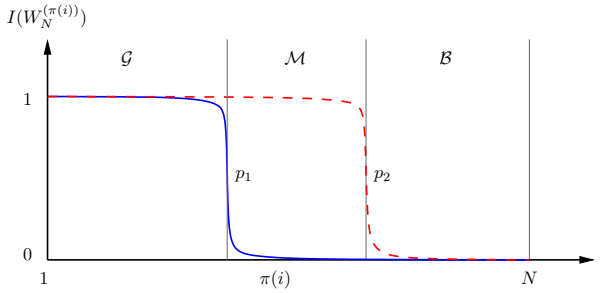

Fig. 1: **Illustration of polarizations for two BSCs.** The blue-solid line represents the channel with transition probability p_1, and the red-dashed one is for p_2 ($p_1 > p_2$). Values of $I(W_N^{(\pi(i))})$, the reordered mutual information, are shown for both channels.

IV. Polar Coding for Fading Channel

A. Intuition

In polar coding for a general B-DMC W, we have seen the channel can be polarized by transforming a set of independent copies of given channels into a new set of channels whose symmetric capacities tend to 0 or 1 for all but a vanishing fraction of indices. To this end, an information set $\mathcal{A}$ is constructed by picking the indices corresponding to K minimum values of $Z(W_N^{(i)})$, which is equivalent to picking those corresponding to K largest values of $I(W_N^{(i)})$. In this sense, the construction of $\mathcal{A}$ is deterministic. However, as indicated in [1], the indices in $\mathcal{A}$ are not adjacent. For this, we introduce a permutation $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$, which reorders all the indices by the value of $I(W_N^{(i)})$ ranging from high to low. Note that the construction of polar codes already implies the fact that for channels of the same type, their permutation mappings are the same.

Another fact about polar codes is that the polarization is uniform [13]. Consider polarizing two B-DMCs, for instance BSCs with parameters p_1 and p_2 respectively, then the information sets, denoted by $\mathcal{A}_1$ and $\mathcal{A}_2$, satisfy $\mathcal{A}_1 \subseteq \mathcal{A}_2$ if $p_1 \geq p_2$. In other words, if a particular channel index constructed from the worse channel (BSC with larger transition probability) polarizes to be noiseless, so does that of the better channel (BSC with smaller transition probability). Based on this observation, when polarizing W_1 and W_2 with transition probabilities defined by (3), the indices after permutation π can be divided into three categories (illustrated in Fig. 1, and without loss of generality, we assume $p_1 \geq p_2$.):

1) $\mathcal{G}$: both channels are good, i.e.

$$I(W_{1,N}^{(\pi(i))}) \to 1, \quad I(W_{2,N}^{(\pi(i))}) \to 1.$$

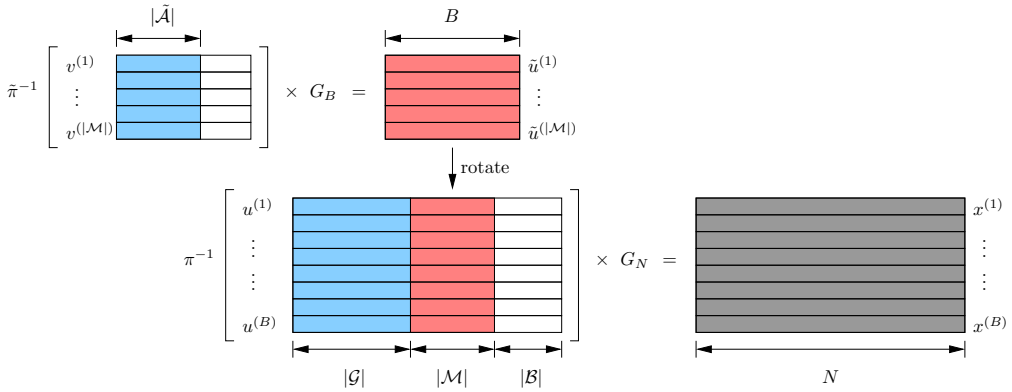

Fig. 2: **Illustration of polar encoder for a fading channel with two states.** Bits in blue are information bits, and those in white are frozen as zeros. The codewords from Phase 1 are used in Phase 2 to generate the final codeword.

2) $\mathcal{M}$: only channel 2 is good, while channel 1 is bad, i.e.

$$I(W_{1,N}^{(\pi(i))}) \to 0, \quad I(W_{2,N}^{(\pi(i))}) \to 1.$$

3) $\mathcal{B}$: both channels are bad, i.e.

$$I(W_{1,N}^{(\pi(i))}) \to 0, \quad I(W_{2,N}^{(\pi(i))}) \to 0.$$

Denote the information sets for two channels as $\mathcal{A}_1$ and $\mathcal{A}_2$ correspondingly, then obviously $\mathcal{A}_1 = \mathcal{G}$, and $\mathcal{A}_2 = \mathcal{G} \cup \mathcal{M}$. Moreover, we have:

$$|\mathcal{G}| = |\mathcal{A}_1| = N[1 - H(p_1) - \epsilon], \tag{5}$$
$$|\mathcal{M}| = |\mathcal{A}_2| - |\mathcal{A}_1| = N[H(p_1) - H(p_2)], \tag{6}$$
$$|\mathcal{B}| = N - |\mathcal{A}_2| = N[H(p_2) + \epsilon], \tag{7}$$

where ϵ is a arbitrary small positive number.

For the fading channel, we consider the transmitter has no prior knowledge of channel states before transmitting, hence, coding over channels with indices in $\mathcal{M}$ is challenging. Observe that for those channels, with probability q_2 they are nearly noiseless, and with probability q_1 they are purely noisy. To this end, each channel can be modeled as a binary erasure channel (BEC) from the viewpoint of blocks, and we denote this channel as $\tilde{W}$. This intuition inspires our design of encoder and decoder for fading channels.

B. Encoder

The encoding process of polar coding for fading channel has two phases, hierarchically using polar codes to generate NB-length codewords, where N is blocklength and B is the number of blocks.

1) Phase 1: Consider a set of B-length block messages $v^{(k)}$ with $k \in \{1, \ldots, |\mathcal{M}|\}$. For every $v^{(k)}$, construct polar codeword $\tilde{u}^{(k)}$, which is formed by the G_B-coset code with parameter $(B, |\tilde{\mathcal{A}}|, \tilde{\mathcal{A}}, 0)$, where $\tilde{\mathcal{A}}$ is the information set for $\tilde{W} \triangleq \text{BEC}(q_1)$, and we choose

$$|\tilde{\mathcal{A}}| = (1 - q_1 - \epsilon)B. \tag{8}$$

In other words, we construct a set of polar codes, where each code corresponds to an index in set $\mathcal{M}$, with the same rate $1 - q_1 - \epsilon$, the same information set $\tilde{\mathcal{A}}$, and the same frozen values 0 as well. Mathematically, if denote the reordering permutation for $\tilde{W}$ as $\tilde{\pi}$, then

$$\tilde{\pi}(v^{(k)}) = [v_1^{(k)}, \ldots, v_{|\tilde{\mathcal{A}}|}^{(k)}, 0, \ldots, 0], \tag{9}$$
$$\tilde{u}^{(k)} = v^{(k)} G_B. \tag{10}$$

2) Phase 2: Consider another set of N-length messages $u^{(b)}$ with $b \in \{1, \ldots, B\}$. For every $u^{(b)}$, construct polar code $x^{(b)}$, which is G_N-coset code with parameter $(N, |\mathcal{G}|, \mathcal{G}, u_{\mathcal{G}^c}^{(b)})$, where $\mathcal{G}$ is BSC information set with size given by (5). Remarkably, we do not set all non-information bits to be 0, but embed the blockwise codewords from Phase 1. More precisely, if denote the permutation operator of BSC as π, then

$$\pi(u^{(b)}) = [u_1^{(b)}, \ldots, u_{|\mathcal{G}|}^{(b)}, \tilde{u}_b^{(1)}, \ldots, \tilde{u}_b^{(|\mathcal{M}|)}, 0, \ldots, 0], \tag{11}$$
$$x^{(b)} = u^{(b)} G_N. \tag{12}$$

By collecting all $\{x^{(b)}\}_{1:B}$ together, the encoder generates and outputs a codeword with length NB. The proposed coding method is illustrated in Fig. 2.

C. Decoder

After receiving the sequence $y_{1:NB}$ from the channel, the decoder's task is to make estimates $\{\hat{v}^{(k)}\}_{1:|\mathcal{M}|}$ and $\{\hat{u}^{(b)}\}_{1:B}$, such that the information bits in both sets of messages match the ones at the transmitter with high probability. Rewrite channel output $y_{1:NB}$ as a $B \times N$ matrix, with row vectors $\{y^{(b)}\}_{1:B}$. As that of the encoding process, the decoding process also has two phases:

1) Phase 1: For every $b \in \{1, \ldots, B\}$, decode $\hat{u}^{(b)}$ from $y^{(b)}$ using SC decoder. Here, as the channel state is available at the receiver, one can modify the SC decoder based on the channel state observed. In this coding scheme, we consider declaring an erasure, denoted as "e", for bad channel states for the blocks with index in $\mathcal{M}$. To this end, polar decoder is given by: if the channel state is h_1, then use *Decoder 1*, otherwise use *Decoder 2*, where the two decoders are expressed follows:

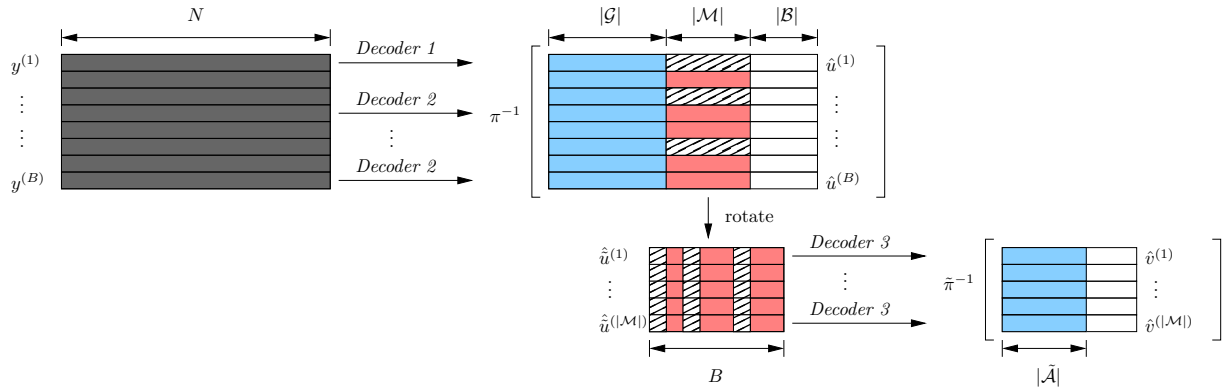

Fig. 3: **Illustration of polar decoder for a fading channel with two states.** After Phase 1, decoder outputs all estimates $\{\hat{u}^{(l)}\}_{1:B}$ using BSC SC decoder based on channel states. Selected columns are delivered as inputs to Phase 2, where the decoder uses BEC SC decoder to decode $\hat{v}^{(k)}$ from $\hat{\tilde{u}}^{(k)}$ for every $k \in \{1, \ldots, |\mathcal{M}|\}$. Bits in shade represent for erasures.

– *Decoder 1 (for blocks b with a bad channel state):*

$$\hat{u}_i^{(b)} \triangleq \begin{cases} 0, & \text{if } b \in \mathcal{B}, \\ \text{e}, & \text{if } b \in \mathcal{M}, \\ d_{1,i}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}), & \text{if } b \in \mathcal{G}, \end{cases}$$

in the order i from 1 to N, where

$$d_{1,i}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}) \triangleq \begin{cases} 0, & \text{if } \frac{W_{1,N}^{(i)}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}|0)}{W_{1,N}^{(i)}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}|1)} \geq 1, \\ 1, & \text{otherwise.} \end{cases}$$

– *Decoder 2 (for blocks b with a good channel state):*

$$\hat{u}_i^{(b)} \triangleq \begin{cases} 0, & \text{if } b \in \mathcal{B}, \\ d_{2,i}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}), & \text{if } b \in \mathcal{G} \cup \mathcal{M}, \end{cases}$$

in the order i from 1 to N, where

$$d_{2,i}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}) \triangleq \begin{cases} 0, & \text{if } \frac{W_{2,N}^{(i)}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}|0)}{W_{2,N}^{(i)}(y^{(b)}, \hat{u}_{1:i-1}^{(b)}|1)} \geq 1, \\ 1, & \text{otherwise.} \end{cases}$$

After decoding from $y^{(b)}$ block by block, the decoder output a $B \times N$ matrix $\hat{\mathbf{U}}$ with rows $\{\hat{u}^{(b)}\}_{1:B}$.

2) Phase 2: Select columns of $\hat{\mathbf{U}}$ with indices in $\mathcal{M}$ after permutation π to construct a $B \times |\mathcal{M}|$ matrix $\hat{\tilde{\mathbf{U}}}$. Consider each column of $\hat{\tilde{\mathbf{U}}}$, denoted by $\hat{\tilde{u}}^{(k)}$ for $k \in \{1, \ldots, |\mathcal{M}|\}$, as the input to decoder in Phase 2. Then, receiver aims to decode $\hat{v}^{(k)}$ from $\hat{\tilde{u}}^{(k)}$ using SC decoder with respect to $\tilde{W}$ =BEC(q_1). More formally, the decoder in Phase 2 is expressed as follow:

– *Decoder 3:*

$$\hat{v}_j^{(k)} \triangleq \begin{cases} 0, & \text{if } k \in \tilde{\mathcal{A}}^c, \\ \tilde{d}_j(\hat{\tilde{u}}^{(k)}, \hat{v}_{1:j-1}^{(k)}), & \text{if } k \in \tilde{\mathcal{A}}, \end{cases}$$

in the order j from 1 to B, where

$$\tilde{d}_j(\hat{\tilde{u}}^{(k)}, \hat{v}_{1:j-1}^{(k)}) \triangleq \begin{cases} 0, & \text{if } \frac{\tilde{W}_N^{(j)}(\hat{\tilde{u}}^{(k)}, \hat{v}_{1:j-1}^{(k)}|0)}{\tilde{W}_N^{(j)}(\hat{\tilde{u}}^{(k)}, \hat{v}_{1:j-1}^{(k)}|1)} \geq 1, \\ 1, & \text{otherwise.} \end{cases}$$

After Phase 2, the decoder output a $|\mathcal{M}| \times B$ matrix $\hat{\mathbf{V}}$ with rows $\{\hat{v}^{(k)}\}_{1:|\mathcal{M}|}$. Decoding process is illustrated in Fig. 3.

D. Achievable Rate

We want to show the rate in proposed polar coding scheme achieves the capacity of converted fading channel given by (4). Intuitively, by using BSC SC decoders corresponding to channel states, the output from Phase 1 successfully recovers all information bits in $\{u^{(b)}\}_{1:B}$. Moreover, for those with indices corresponding to $\mathcal{M}$, the decoder could decode correctly if channel state is h_2, and set to erasure otherwise. Thus, for Phase 2, vector $\hat{\tilde{u}}^{(k)}$ can be considered as an output of BEC(q_1), hence BEC SC decoder could decode all information bits in $v^{(k)}$ correctly for any $k \in \{1, \ldots, |\mathcal{M}|\}$.

Theorem 1. *The proposed polar coding scheme achieves any rate $R < C_{SI\text{-}D}$. (For sufficiently large N and B, the error probability scales as $O(B2^{-N^\beta}) + O(N2^{-B^\beta})$ with $\beta < 1/2$; and it vanishes with a choice of $B = o(2^{N^\beta})$ and $N = o(2^{B^\beta})$.)*

Proof: The proof is straightforward by utilizing error bound from polar coding. In Phase 1 of decoding, the error probability of recovering $u^{(b)}$ correctly for each $b \in \{1, \ldots, B\}$ is given by $P_{1,e}^{(b)} = O(2^{-N^\beta})$. Similarly, in decoding Phase 2, the error probability of recovering $v^{(k)}$ correctly for each $k \in \{1, \ldots, M\}$ is given by $P_{2,e}^{(k)} = O(2^{-B^\beta})$. Hence, by union bound, the total decoding error probability is upper bounded by

$$P_e \leq \sum_{b=1}^{B} P_{1,e}^{(b)} + \sum_{k=1}^{|\mathcal{M}|} P_{2,e}^{(k)} = O(B2^{-N^\beta}) + O(N2^{-B^\beta}),$$

as N and B tend to infinity. Therefore, P_e vanishes if $B = o(2^{N^\beta})$ and $N = o(2^{B^\beta})$.

The achievable rate (corresponding to the transmission of messages bits in $v^{(k)}$ and $u^{(b)}$) is given by

$$\begin{aligned} R &= \frac{1}{NB}\Big\{|\mathcal{M}||\tilde{\mathcal{A}}| + B|\mathcal{G}|\Big\} \\ &= [H(p_1) - H(p_2)][1 - q_1 - \epsilon] + [1 - H(p_1) - \epsilon] \\ &= q_1[1 - H(p_1)] + q_2[1 - H(p_2)] - \delta(\epsilon), \end{aligned}$$

where we have used (5), (6) and (8), and

$$\delta(\epsilon) \triangleq \epsilon[1 + H(p_1) - H(p_2)] \to 0, \text{ as } \epsilon \to 0.$$

Thus, any rate $R < C_{\text{SI-D}}$ is achievable. ■

E. Complexity Analysis

Polar coding schemes for both BSC and BEC have relatively low complexity. Since the proposed polar coding scheme for fading channel hierarchically utilizes these polar codes, the low complexity is inherited. More precisely, $|\mathcal{M}|$ number of B-length polar codes as well as B number of N-length polar codes are utilized. Thus, the overall complexity of the coding scheme for both encoding and decoding is given by

$$|\mathcal{M}| \cdot O(B \log B) + B \cdot O(N \log N) = O(NB \log(NB)).$$

V. Discussion

In this section, we generalize the polar coding scheme to fading channels with arbitrary finite number of states. Consider that the channel gain H_b has S states $\{h_1, \ldots, h_S\}$, where $\Pr\{H_b = h_s\} \triangleq q_s$, $s \in \{1, \ldots, S\}$, and $\sum_s q_s = 1$. Then, the BPSK modulated channel, defined in (2), is still a BSC, where the transition probability, with probability q_s, is given by

$$\Pr\{\bar{Z} = 1\} = 1 - \Phi(h_s\sqrt{\text{SNR}}) \triangleq p_s. \tag{13}$$

Denote the converted BSC corresponding to state h_s as W_s, then the capacity of converted channel is given by

$$C_{\text{SI-D}} = \sum_{s=1}^{S} q_s[1 - H(p_s)], \tag{14}$$

where $1 - H(p_s)$ is the capacity of W_s.

Observe that when polarizing S BSCs with different transition probabilities, the indices could be divided into $S+1$ sets after permutation π. $S-1$ mixture sets $\mathcal{M}_1, \ldots, \mathcal{M}_{S-1}$ are considered for this case. Without loss of generality, we assume $p_1 \geq p_2 \geq \cdots \geq p_S$. Then, $|\mathcal{M}_s|/N = H(p_s) - H(p_{s+1})$, and for index in set $\mathcal{M}_s$, $W_1, \ldots, W_s$ are polarized to be purely noisy and all others to be noiseless. To this end, we consider a BEC with erasure probability $e_s = \sum_{t=1}^{s} q_t$ to characterize the polarization result for an index in $\mathcal{M}$. (See Fig. 4.)

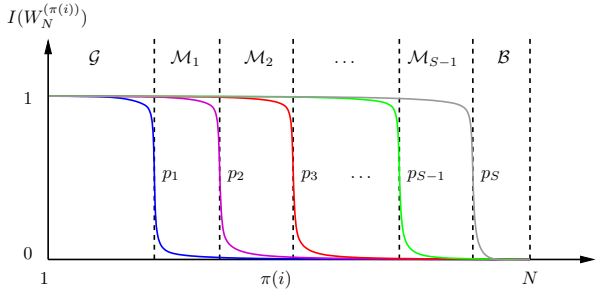

Fig. 4: **Illustration of polarizations for S BSCs.** There are $S-1$ mixture sets, denoted as $\mathcal{M}_1, \ldots, \mathcal{M}_{S-1}$.

Polar coding scheme designed for this channel is similar. In Phase 1 of encoding, transmitter needs to generate $S-1$ sets of polar codes, where each one is G_B-coset codes with parameter $(B, |\tilde{\mathcal{A}}_s|, \tilde{\mathcal{A}}_s, 0)$ with respect to BEC(e_s), and all the encoded codewords are embed into messages for Phase 2. At the receiver end, Phase 1 should use one of S SC decoders for BSC to decode $\hat{u}_{(b)}$, based on observation of channel states. Then, in Phase 2, $S-1$ BEC SC decoders are implemented in parallel to recover the information bits. By adopting this polar coding scheme, the achievable rate is given by

$$\begin{aligned} R &= \frac{1}{NB}\left\{B|\mathcal{G}| + \sum_{s=1}^{S-1} |\mathcal{M}_s||\tilde{\mathcal{A}}_s|\right\} \\ &= [1 - H(p_1) - \epsilon] + \sum_{s=1}^{S-1}[H(p_s) - H(p_{s+1})](1 - e_s - \epsilon) \\ &= \sum_{s=1}^{S} q_s[1 - H(p_s)] - \delta'(\epsilon), \end{aligned}$$

where $\delta'(\epsilon) = \epsilon[1 + H(p_1) - H(p_S)]$. Thus, the proposed polar coding scheme achieves the capacity of channel, and the encoding and decoding complexities are both given by

$$\sum_{s=1}^{S-1} |\mathcal{M}_s| \cdot O(B \log B) + B \cdot O(N \log N) = O(NB \log(NB)),$$

which is independent to the value of S as $\sum_{s=1}^{S} |\mathcal{M}_s| \leq N$.

References

[1] E. Arıkan, "Channel polarization: A Method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. on Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[2] E. Arıkan, and E. Telatar, "On the rate of channel polarization," in *Proc. 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, Seoul, Korea, Jun. 2009.

[3] E. Şaşoğlu, E. Arıkan, and E. Telatar, "Polarization for arbitrary discrete memoryless channels," in *Proc. 2009 IEEE Information Theory Workshop (ITW 2009)*, Taormina, Sicily, Italy, Oct. 2009.

[4] E. Arıkan, "Source polarization,"in *Proc. 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, Austin, Texas, U.S.A., Jun. 2010.

[5] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. on Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[6] M. Karzand, and E. Telatar, "Polar Codes for Q-ary Source Coding," in *Proc. 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, Austin, Texas, U.S.A., Jun. 2010.

[7] O. O. Koyluoglu, K. Appaiah, H. Si, and S. Vishwanath, "Expansion coding: Achieving the capacity of an AEN channel," in *Proc. 2012 IEEE International Symposium on Information Theory (ISIT 2012)*, Boston, Massachusetts, U.S.A., Jul. 2012.

[8] E. Abbe, and A. Barron, "Polar coding schemes for the AWGN channel," in *Proc. 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, Saint Petersburg, Russia, Jul. 2011.

[9] E. Abbe, and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. on Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012.

[10] E. Arıkan, "Systematic polar coding," *IEEE Communications Letters*, vol. 15, no. 8, pp. 860–862, Aug. 2011.

[11] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check", arXiv:1208.3091.

[12] D. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[13] S. B. Korada, "Polar codes for channel and source coding," *PhD Thesis*, EPFL, 2009.