

Polar Coding for Fading Channels: Binary and Exponential Channel Cases

Hongbo Si, O. Ozan Koyluoglu, *Member, IEEE*, and Sriram Vishwanath, *Senior Member, IEEE*

Abstract

This work presents a polar coding scheme for fading channels, focusing primarily on fading binary symmetric and additive exponential noise channels. For fading binary symmetric channels, a hierarchical coding scheme is presented, utilizing polar coding both over channel uses and over fading blocks. The receiver uses its channel state information (CSI) to distinguish states, thus constructing an overlay erasure channel over the underlying fading channels. By using this scheme, the capacity of a fading binary symmetric channel is achieved without CSI at the transmitter. Noting that a fading AWGN channel with BPSK modulation and demodulation corresponds to a fading binary symmetric channel, this result covers a fairly large set of practically relevant channel settings.

For fading additive exponential noise channels, expansion coding is used in conjunction to polar codes. Expansion coding transforms the continuous-valued channel to multiple (independent) discrete-valued ones. For each level after expansion, the approach described previously for fading binary symmetric channels is used. Both theoretical analysis and numerical results are presented, showing that the proposed coding scheme approaches the capacity in the high SNR regime. Overall, utilizing polar codes in this (hierarchical) fashion enables coding without CSI at the transmitter, while approaching the capacity with low complexity.

Index Terms

Binary symmetric channel, Channel coding, Expansion coding, Fading channels, Polar codes

H. Si and S. Vishwanath are with the Laboratory for Informatics, Networks, and Communications, Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX 78712. Email: sihongbo@mail.utexas.edu, sriram@austin.utexas.edu.

O. O. Koyluoglu is with the Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ 85721. Email: ozan@email.arizona.edu.

I. INTRODUCTION

Polar codes are the first family of provably capacity achieving codes for symmetric binary-input discrete memoryless channels (B-DMC) with low encoding and decoding complexity [1] [2]. These codes polarize the underlying channel in the sense that, via channel combining and channel splitting stages, multiple uses of the given channel are transformed into equivalent polarized ones: either purely noisy (referred to as “bad” channel instances) or noiseless (referred to as “good” channel instances). Then, information symbols are mapped to the good instances of polarized channels, whereas channel inputs corresponding to the bad instances are fixed and shared between the transmitter and receiver. It is shown in [1] that the fraction of the good channel instances approaches the symmetric capacity of the channel, which is equal to the capacity of the underlying channel if the channel is symmetric. That is, polar codes achieve the capacity of symmetric B-DMCs. This phenomenon of channel polarization has then been generalized to arbitrary discrete memoryless channels with a construction complexity to the same order and a similar 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 coding for larger source alphabets [6].

Recently, polar codes have been adapted to channels with non-discrete inputs as well. In [7], using polarization results for multiple access channels [8], a polar coding scheme for additive white Gaussian noise (AWGN) channels is presented. It is shown that the approach of using multiple access channel coding with a large number of binary-input users possesses much better complexity attributes compared to that of using single-user channels with large input cardinality. In a separate work [9], by adopting discrete polar codes as embedded codes, an expansion coding approach is presented, and the capacity of additive exponential noise channel is shown to be achievable in the high SNR regime.

The analysis of polar coding for fading channels, with either discrete-valued or continuous-valued noises, is still limited. Recent work [10] investigates binary input real number output AWGN fading channel, where the fading coefficient is assumed to be one of the two states with equal probabilities. These fading coefficients are assumed to follow arbitrary distributions with the requirement of satisfying some tail probability constraints. For this setup, the authors proposed polar coding schemes where symbols are multiplexed in a specific fashion at the

encoder. In particular, the paper analyzes diagonal, horizontal, and uniform multiplexers; and, the corresponding diversity and outage analysis have been performed. Another recent work [11] focuses on polar coding schemes for Rayleigh fading channel under two scenarios: block fading with known channel state information (CSI) at the transmitter and fast fading with fading distribution known at the transmitter. For the latter case, the channel is shown to be symmetric, and through quantization of the channel output, the polar coding scheme is shown to achieve a constant gap to the capacity.

In this work, we focus on a block fading model without the CSI at the transmitter, and propose a hierarchical polar coding scheme for such channels. First, we focus on fading binary symmetric channel (BSC), which is an important model as it is closely related to an AWGN block fading channel with BPSK modulation and demodulation. Such binary input AWGN models are previously analyzed in [12][13] to evaluate the performance of polar codes over AWGN channels. Here, we focus on communication channel models that involve fading, where the channel coefficients vary according to a block fading model. This scenario of fading AWGN with BPSK modulation resembles a fading binary symmetric channel model, where each fading block has a cross-over probability depending on the corresponding channel state realization. Specifically, AWGN channel states with higher SNRs map to binary symmetric channels with lower crossover probabilities. For this binary symmetric fading model, we propose a novel polar coding approach that utilizes polarization in a *hierarchical* manner *without* channel state information (CSI) at the transmitter (with channel state statistics assumed to be known at the transmitter). The key factor enabling our coding scheme is the hierarchical utilization of polar coding. More precisely, polar codes are not only designed over channel uses for each fading state, but also utilized over fading blocks. By taking advantage of the degradedness property of channel polarization between different BSCs, an erasure model (over fading blocks) is constructed for every channel instance that polarizes differently depending on the channel states. It is shown that this proposed coding scheme, without instantaneous CSI at the transmitter, achieves the capacity of the fading binary symmetric channel.

As an additional analog fading model (in addition to the AWGN fading scenario with BPSK modulation), we consider additive exponential noise channels. In analog channels, the additive exponential noise (AEN) channel is of particular interest as it models worst-case noise given mean and non-negativity constraints [14]. In addition, the AEN model naturally arises in non-

coherent communication settings, and in optical communication scenarios [15][16]. In [9], an expansion scheme is proposed to achieve the capacity of AEN channels. Here, we adopt a similar approach for the fading AEN channels. In particular, due to the decomposition property of the exponential distribution, we show that a fading AEN channel can be transformed into a set of fading BSCs. Then, by employing the aforementioned polar coding scheme for fading BSCs at each level, we show that the proposed method approaches to the capacity of fading AEN channels in the high SNR regime.

In both cases considered in this paper, utilizing polar codes in such a novel (hierarchical) way enables coding without CSI at the transmitter, a practically important scenario in wireless systems. In addition, the low encoding and decoding complexity of polar codes are inherited in the proposed schemes. (The scaling of complexity with respect to the system parameters are detailed in the later parts of the sequel.) Therefore, the proposed approach, by having both low complexity and realistic CSI assumption properties, is suitable for practical utilization of polar codes over fading channels (especially for those channels with long fading coherence intervals).

The rest of paper is organized as follows. After a brief introduction of the preliminary results on polar codes in Section II, the polar coding scheme for fading binary symmetric channels is detailed in Section III. Section IV is devoted to the study of fading additive exponential noise channels. Finally, concluding remarks are provided in Section V.

II. INTRODUCTION TO POLAR CODES

The construction of polar code is based on the observation of channel polarization. Consider a binary-input discrete memoryless channel $W : \mathcal{X} \rightarrow \mathcal{Y}$, 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 the message $u_{1:N}$, the encoded $x_{1:N} = u_{1:N} G_N$ is transmitted through N independent copies of channel W . Then, consider N binary-input coordinate channels $W_N^{(i)} : \mathcal{X} \rightarrow \mathcal{Y}^N \times \mathcal{X}^{i-1}$, where for each $i \in \{1, \dots, N\}$ the transition probability is given by

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

Here, as N tends to infinity, the channels $\{W_N^{(i)}\}_{1:N}$ polarize to either noiseless or purely noisy ones, and the fraction of noiseless channels is close to $I(W)$, the symmetric capacity of the 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 the frozen vector (can be set to all-zeros for symmetric channels [1], which is the focus of this paper), and the information set \mathcal{A} is chosen as a K -element subset of $\{1, \dots, N\}$ 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 indices in \mathcal{A} are “good” channel indices, whereas those in \mathcal{A}^c corresponds to “bad” channel indices.)

The decoder in the polar coding scheme is a successive cancellation (SC) decoder, which gives an estimate $\hat{u}_{1:N}$ of $u_{1:N}$ given the 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

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

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

To summarize, polar codes have excellent rate and complexity, but the code design is sensitive to the channel state information at the transmitter. For example, the choices of “good” indices depend on the crossover probability p of a BSC, i.e., the set $\mathcal{A}(p)$ is a function of the value of p . Thus, it is not straight-forward to design capacity-achieving polar coding schemes if there is uncertainty in the channel parameters. In particular, for a block fading scenario, if the channel states are not known a priori at the transmitter, one may not know which indices should correspond to “good” channel instances. As detailed in the next section, the proposed scheme solves this problem via a hierarchical code design.

III. POLAR CODING FOR FADING BINARY SYMMETRIC CHANNEL

A. 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. For such cases, a block fading model is considered, wherein the channel state is assumed to be a constant over each coherence time interval, and follows a stationary ergodic process across fading blocks. For such a block fading model, we consider the practical scenario where the channel state information is available only at the decoder (CSI-D) [17, pages 186-187], while the transmitter is assumed to know only the statistics of the channel states.

Binary symmetric channel (BSC) is a channel with binary input X , binary noise Z , and a binary output $Y = X \oplus Z$. Here, for the fading BSC, the channel noise is a Bernoulli distributed random variable, where its statistics depend on the channel states. For the block fading BSC considered in this work, the channel is modeled as follows.

$$Y_{b,i} = X_{b,i} \oplus Z_{b,i}, \quad b = 1, \dots, B, \quad i = 1, \dots, N, \quad (1)$$

where N is the block length, and B is the number of fading blocks. Here, $Z_{b,i}$ are assumed to be identically distributed within a block and follow an i.i.d. fading process over blocks. That is, if we consider fading BSC with S states, with probability q_s the parameter p_s is chosen for the fading block b , where the channel noise $Z_{b,i}$ is sampled from a Bernoulli random variable with parameter p_s for all $i \in \{1, \dots, N\}$. Here, $1 \leq s \leq S$ and $\sum_{s=1}^S q_s = 1$.

In wireless communications, the fading binary symmetric channel is utilized to model a fading AWGN channel with BPSK modulation and demodulation. In particular, for a fading AWGN channel with input power constraint P_X , the channel noise is distributed as i.i.d. Gaussian with variance P_Z , and the channel gain (the factor h in the AWGN channel $Y = hX + Z$) remains the same statistic within a fading block, and follows an ergodic process over different blocks. After utilizing the BPSK modulation and demodulation at the encoder and decoder, respectively, the equivalent channel is a binary input and binary output channel, with transition probability relating to AWGN channel state. More precisely, if the channel gain for a particular fading block b , $h_{b,i} \forall i$, is equal to h_s with probability q_s for some $s \in \{1, 2, \dots, S\}$, then the corresponding binary noise in the equivalent fading BSC has the statistics of

$$p_s \triangleq \Pr\{Z_{b,i} = 1\} = 1 - \Phi(h_s \sqrt{\text{SNR}}), \quad (2)$$

where $\Phi(\cdot)$ is the CDF of normal distribution and SNR is the signal-to-noise ratio, i.e. $\text{SNR} = P_X/P_Z$. In other words, the channel at each fading block can be modeled as $W_s \triangleq \text{BSC}(p_s)$ with probability q_s .

The ergodic capacity of a fading binary symmetric channel is given by [18, pages 584-586]

$$C_{\text{CSI-D}} = \sum_{s=1}^S q_s [1 - H(p_s)], \quad (3)$$

where $H(\cdot)$ is the binary entropy function, and CSI-D refers to channel state information at the decoder. Note that, the ergodic capacity of fading BSC is an average over the capacities of all possible channels corresponding to different channel states. In this section, we propose a polar coding scheme that achieves the capacity of this fading BSC with low encoding and decoding complexity, without having instantaneous channel state information at the transmitter. Towards this end, we first focus on a fading BSC with two channel states, and then generalize our results to arbitrary finite number of channel states.

B. Intuition

In polar coding for a BSC, we see that 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). Towards applying such a polarization phenomenon to fading BSC, we first focus on how two binary symmetric channels polarize at the same time. We summarize a result given in [19] regarding the polarization of degraded channels.

Lemma 1 ([19]). *For two binary symmetric channels $W_1 \triangleq \text{BSC}(p_1)$ and $W_2 \triangleq \text{BSC}(p_2)$, if W_1 is degraded with respect to W_2 , i.e. $p_1 \geq p_2$, then for any channel index $i \in \{1, \dots, N\}$, the reconstructed channels after polarization have the relationship that $W_{1,N}^{(i)}$ is degraded with respect to $W_{2,N}^{(i)}$, and hence $I(W_{1,N}^{(i)}) \leq I(W_{2,N}^{(i)})$.*

That is, when polarizing two binary symmetric channels, the reconstructed channels of the degraded channel have lower symmetric rate compared to that of the other channel. This statement also implies that

$$\mathcal{A}_1 \subseteq \mathcal{A}_2,$$

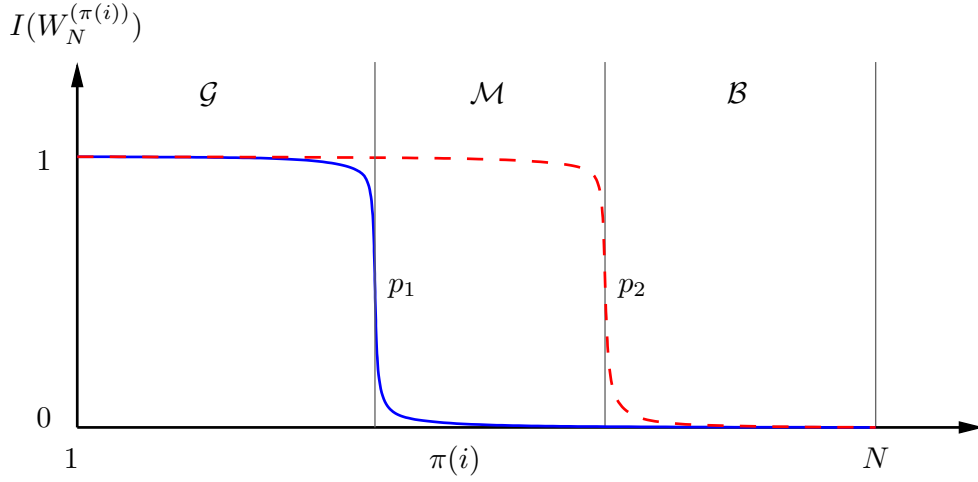


Fig. 1: **Illustration of polarizations for two binary symmetric channels.** The blue-solid line represents the degraded channel with transition probability p_1 , and the red-dashed one represents the superior channel with p_2 ($p_1 \geq p_2$). Values of $I(W_N^{(\pi(i))})$, the reordered symmetric mutual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

where \mathcal{A}_1 and \mathcal{A}_2 denote the information sets of the degraded and superior channels, respectively. This relationship is illustrated in Fig. 1. Based on this observation, when polarizing two BSCs, the channel indices after reordering permutation π can be divided into three categories (we assume that channel 1 is degraded, i.e., $p_1 \geq p_2$):

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

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

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

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

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

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

We have the following relationships between these sets. First, information sets for two channels

are given by $\mathcal{A}_1 = \mathcal{G}$, and $\mathcal{A}_2 = \mathcal{G} \cup \mathcal{M}$. Moreover, considering the sizes of these sets, we have:

$$|\mathcal{G}| = |\mathcal{A}_1| = N[1 - H(p_1) - \epsilon], \quad (4)$$

$$|\mathcal{M}| = |\mathcal{A}_2| - |\mathcal{A}_1| = N[H(p_1) - H(p_2)], \quad (5)$$

$$|\mathcal{B}| = N - |\mathcal{A}_2| = N[H(p_2) + \epsilon], \quad (6)$$

where ϵ is an arbitrarily small positive number (that vanishes as $N \rightarrow \infty$).

For a fading binary symmetric channel, we again utilize Fig. 1 to illustrate our coding scheme. Here, consider a fading BSC with only two fading states, the degraded state and the superior one (denoted as state 1 and 2, respectively). If channel is in state 1, which happens with probability q_1 , the fading channel polarizes to the blue-solid curve, and otherwise the channel is in state 2, which happens with the probability $q_2 = 1 - q_1$, and the fading channel polarizes to the red-dashed curve. Hence, the reconstructed channel with index in set \mathcal{G} always polarizes to a good one, i.e., its symmetric mutual information is close to 1 no matter what the fading state is. And, the reconstructed channel with index in set \mathcal{B} always polarizes to a bad one, i.e., its symmetric mutual information is close to 0 no matter what the fading state is. Therefore, one can reliably transmit information for channel instances belonging to \mathcal{G} , whereas one may not transmit any information for channel instances belonging to \mathcal{B} . The novel part of the proposed coding scheme is for the middle region, i.e., coding over the set \mathcal{M} , where reconstructed channels polarize differently depending on the channel states. Since we consider the transmitter has no prior knowledge of channel states before transmitting, coding over channels with indices in \mathcal{M} is challenging. At this point, we observe that for these channels, with probability q_2 they are nearly noiseless, and with probability q_1 they are purely noisy. Thus, each channel can be modeled as a binary erasure channel (BEC) from the viewpoint of blocks, where the erasure probability is equal to q_1 . Here, we denote this channel as

$$\tilde{W} \triangleq \text{BEC}(q_1).$$

This observation motivates our design of hierarchical encoder and decoder for fading BSCs.

C. Encoder

The encoding process 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 (BEC Encoding)*: In this phase, we generate $|\mathcal{M}|$ number of BEC polar codes, each with length B . Consider a set of blockwise messages $v^{(k)}$ with $k \in \{1, \dots, |\mathcal{M}|\}$. 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 $\tilde{W} = \text{BEC}(q_1)$, and we choose the rate to be optimal, i.e.

$$|\tilde{\mathcal{A}}| = (1 - q_1 - \epsilon)B. \quad (7)$$

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. More precisely, if we denote the reordering permutation for \tilde{W} as $\tilde{\pi}$, then

$$\tilde{\pi}(v^{(k)}) = [v_1^{(k)}, \dots, v_{|\tilde{\mathcal{A}}|}^{(k)}, 0, \dots, 0], \quad (8)$$

$$\tilde{u}^{(k)} = v^{(k)}G_B. \quad (9)$$

2) *Phase 2 (BSC Encoding)*: In this phase, we generate B number of BSC polar codes, each with length N . Consider a set of messages $u^{(b)}$ with $b \in \{1, \dots, B\}$. For every $u^{(b)}$, construct polar codeword $x^{(b)}$, 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 (4). Remarkably, we do not set all non-information bits to be 0, but transpose the blockwise codewords generated from Phase 1 and embed them into the messages of this phase. More precisely, if denote the reordering permutation operator of BSC as π , then

$$\pi(u^{(b)}) = [u_1^{(b)}, \dots, u_{|\mathcal{G}|}^{(b)}, \tilde{u}_b^{(1)}, \dots, \tilde{u}_b^{(|\mathcal{M}|)}, 0, \dots, 0], \quad (10)$$

$$x^{(b)} = u^{(b)}G_N. \quad (11)$$

By collecting all $\{x^{(b)}\}_{1:B}$ together, the encoder outputs a codeword with length NB . We equivalently express these codewords by a $B \times N$ matrix. The proposed encoder for fading binary symmetric channel is illustrated in Fig. 2.

D. 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,

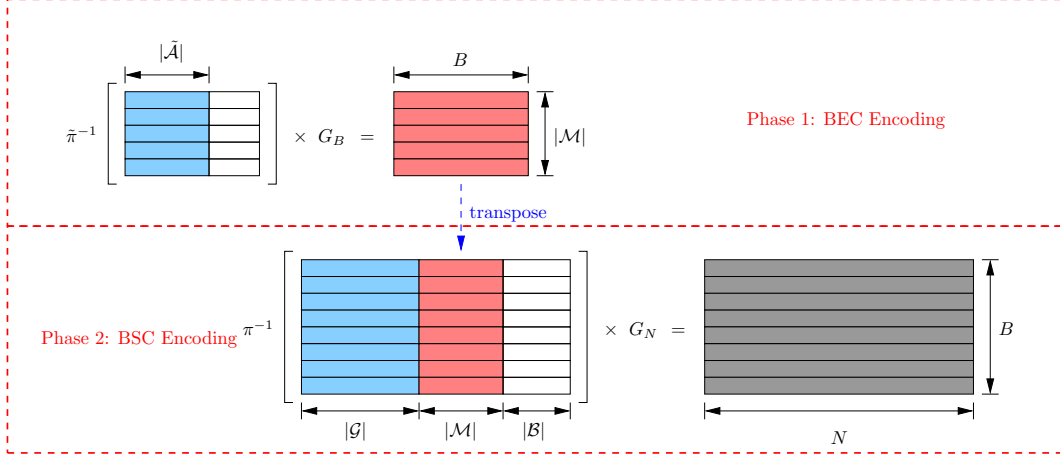


Fig. 2: **Illustration of proposed polar encoder for a fading binary symmetric channel with two states.** Bits in blue are information bits, and those in white are frozen as zeros. The codewords generated from Phase 1 are transposed and embedded into the messages of Phase 2 to generate the final codeword of length NB . $\tilde{\pi}$ and π are column reordering permutations with respect to BEC and BSC, correspondingly.

with row vectors $\{y^{(b)}\}_{1:B}$. As that of the encoding process, the decoding process also works in phases:

1) *Phase 1 (BSC Decoding I)*: In this phase, we decode part of the output blocks using the BSC SC decoder with respect to the superior channel state. More precisely, since the receiver knows channel states, it can adopt the correct SC decoder (BSC(p_2) SC decoder in this case) to obtain $\hat{u}^{(b)}$ from $y^{(b)}$ for every b corresponding to the superior channel state. To this end, the “BSC Decoder I” (for block b with the superior fading state) in this phase is described as follow:

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

in the order i from 1 to N , where

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

In this phase, one can reliably decode the information bits in blocks with respect to the superior channel states (with the knowledge of frozen symbols corresponding to \mathcal{B} indices).

For the blocks with the degraded channel states, information bits cannot be decoded reliably because frozen bits corresponding to set \mathcal{M} is not known at the decoder. At this point, we use the next phase to decode these bits using a BEC SC decoder. For that, a $B \times |\mathcal{M}|$ matrix, $\hat{\hat{\mathbf{U}}}$, is constructed by choosing each row as $\hat{u}^{(b)}$ with respect to bits in set \mathcal{M} for the superior channel states. The other elements in this matrix, i.e., the symbols corresponding to the degraded channel states, are set to erasures.

2) *Phase 2 (BEC Decoding)*: In this phase, we decode the frozen bits with respect to the degraded states using BEC SC decoders. More precisely, each column of matrix $\hat{\hat{\mathbf{U}}}$, denoted by $\hat{\hat{u}}^{(k)}$ for $k \in \{1, \dots, |\mathcal{M}|\}$, is considered as the input to the decoder, and the receiver aims to decode $\hat{v}^{(k)}$ from $\hat{\hat{u}}^{(k)}$ using the SC decoder with respect to channel $\tilde{W} = \text{BEC}(q_1)$. More formally, the “BEC Decoder” in this phase is expressed by the following:

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

for j from 1 to B , where

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

After Phase 2, the decoder outputs an $|\mathcal{M}| \times B$ matrix $\hat{\mathbf{V}}$ with rows $\{\hat{v}^{(k)}\}_{1:|\mathcal{M}|}$. Moreover, the decoder can reconstruct all bits erased in matrix $\hat{\hat{\mathbf{U}}}$, which is denoted as $\tilde{u}^{(b)}$ for each b corresponding to the degraded channel state. Using this information, we are able to decode the information bits in blocks with the degraded channel state in the next phase.

3) *Phase 3 (BSC Decoding II)*: In this phase, we decode the remaining blocks from Phase 1, using BSC SC decoders with respect to the degraded channel states. In particular, for each block b in the degraded channel state, the receiver could decode $\hat{u}^{(b)}$ from $y^{(b)}$ using $\text{BSC}(p_1)$ SC decoder by setting frozen bit as $\tilde{u}_i^{(b)}$ for each $i \in \mathcal{M}$ and 0 for each $i \in \mathcal{B}$. More formally, we have the “BSC Decoder II” (for block b with a degraded fading state) described as:

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

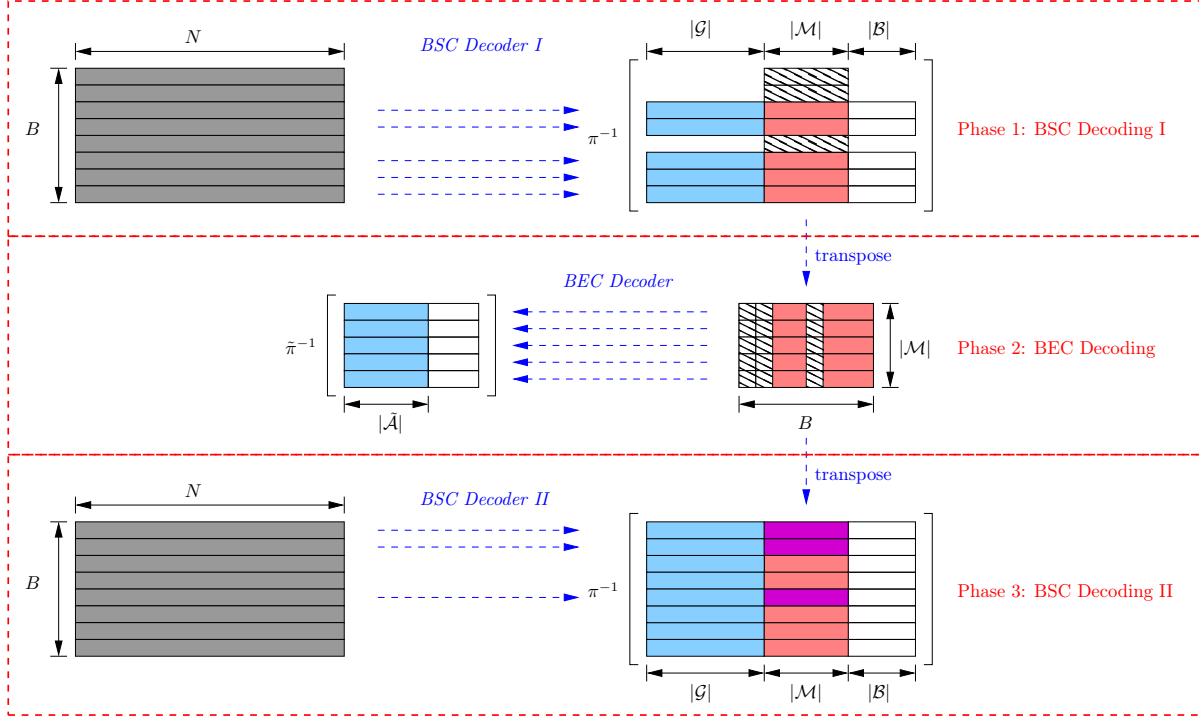


Fig. 3: **Illustration of proposed polar decoder for a fading binary symmetric channel with two states.** In Phase 1, decoder outputs all estimates using BSC SC decoders corresponding to the superior channel state. Selected columns are transposed and delivered as inputs to next phase, by adding all-erasures rows for blocks with the degraded channel state. In Phase 2, the decoder continues to use BEC SC decoders to decode all the blockwise information bits, and to recover all erased bits in shade. In Phase 3, the BSC SC decoders corresponding to the degraded channel state are utilized to decode the remaining information bits, by taking values of frozen bits in set \mathcal{M} as the decoded results from the previous phase. $\tilde{\pi}$ and π are column reordering permutations with respect to BEC and BSC, correspondingly.

in the order i from 1 to N , where

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

The whole decoding process for fading binary symmetric channel is illustrated in Fig. 3.

E. Performance

Here, we summarize the performance of the proposed polar coding scheme. Intuitively, by using BSC SC decoders corresponding to the superior channel state, the output from Phase 1 successfully recovers all information bits, because the size of information set is equal to the size of \mathcal{G} . Then, for decoding at Phase 2, the input vector $\hat{u}^{(k)}$ can be considered as a q_1 -fraction erased polar codeword, hence, BEC SC decoder can decode all information bits in $v^{(k)}$ correctly for all $k \in \{1, \dots, |\mathcal{M}|\}$, and recover the erased entries correctly as well. Finally, in Phase 3 of decoding, the bits in \mathcal{M} have the correct frozen values, and by adopting BSC SC decoders corresponding to the degraded channel state, all the remaining information bits can be decoded correctly.

Therefore, as long as the designed rates of polar codes do not exceed the corresponding channel capacities, all information bits in our proposed polar coding scheme are reliably decodable. Hence, we have the following theorem.

Theorem 2. *The proposed polar coding scheme achieves any rate $R < C_{\text{CSI-D}}$, for sufficiently large N and B , and the decoding error probability scales as $\max\{O(B2^{-N^\beta}), O(N2^{-B^\beta})\}$ with $\beta < 1/2$. Moreover, the complexity of the encoding and decoding processes are both given by $O(NB \log(NB))$, where N is the block length and B is the number of blocks.*

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

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

where we have used (4), (5), (7), and

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

The proof for error exponent is obtained by utilizing error bound from polar coding. In Phase 1 and 3 of decoding, the error probability of recovering $u^{(b)}$ correctly for each $b \in \{1, \dots, B\}$ 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 $k \in \{1, \dots, |\mathcal{M}|\}$ 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 $N = o(2^{B^\beta})$.

Finally, since we have $|\mathcal{M}|$ number of B -length polar codes as well as B number of N -length polar codes utilized, the overall complexity of the coding scheme for both encoding and decoding is given by

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

■

This theorem shows that our proposed polar coding scheme achieves the capacity of fading BSC with low encoding and decoding complexity. In addition, the error scaling performance, which is inherited from polar codes, implies that long coherence intervals as well as large number of blocks are required for this coding scheme to make the error probability arbitrarily small.

F. Generalization

Here, we generalize the polar coding scheme to fading binary symmetric channel with arbitrary finite number of states. Consider S number of BSCs, each with a different transition probability. Without loss of generality, consider $W_1 \triangleq \text{BSC}(p_1), \dots, W_S \triangleq \text{BSC}(p_S)$, with $p_1 \geq p_2 \geq \dots \geq p_S$. Then, a fading BSC with S fading states is modeled as the channel being W_s with probability q_s for a given fading block, where $\sum_{s=1}^S q_s = 1$. The polarization of a fading BSC with S fading states is illustrated in Fig. 4, where the reconstructed channel indices are divided into $S+1$ sets after permutation π . In addition to \mathcal{G} and \mathcal{B} , there exist $S-1$ middle sets $\mathcal{M}_1, \dots, \mathcal{M}_{S-1}$ in this case. For each channel index in set \mathcal{M}_s , channels having statistics being one of W_1, \dots, W_s are polarized to be purely noisy and the remaining ones are noiseless. Therefore, for channel indices belonging to \mathcal{M}_s , we consider modeling them as BECs with erasure probability given by

$$e_s \triangleq \sum_{t=1}^s q_t, \quad 1 \leq s \leq S-1.$$

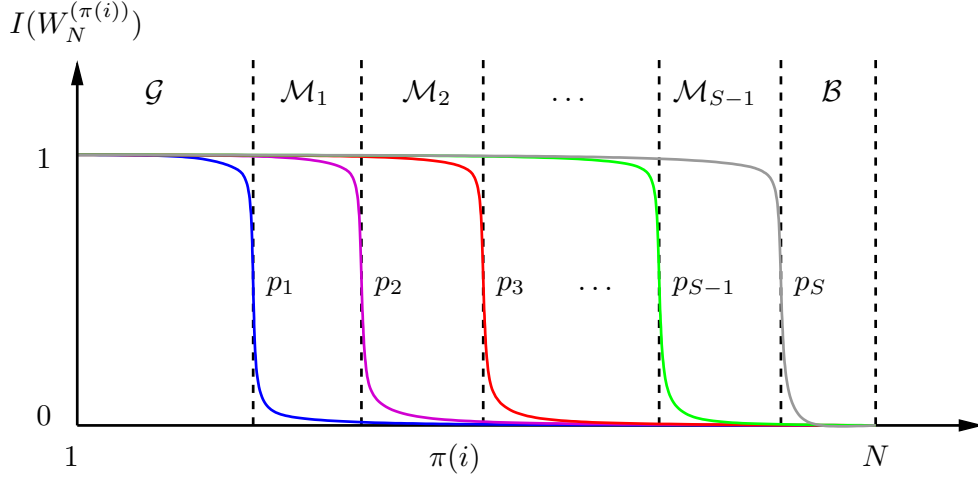


Fig. 4: **Illustration of polarization for a fading binary symmetric channel with S channel states.** Besides \mathcal{G} and \mathcal{B} , there are $S - 1$ middle sets, denoted as $\mathcal{M}_1, \dots, \mathcal{M}_{S-1}$.

Based on this, we have

$$\begin{aligned} |\mathcal{G}| &= |\mathcal{A}_1| = N[1 - H(p_1) - \epsilon], \\ |\mathcal{M}_s| &= N[H(p_s) - H(p_{s+1})], \quad 1 \leq s \leq S - 1, \\ |\mathcal{B}| &= N - |\mathcal{A}_S| = N[H(p_S) + \epsilon]. \end{aligned}$$

Here, the polarization result is similar to the case of two fading states, and we utilize a similar hierarchical coding scheme. In Phase 1 of encoding, transmitter generates $S - 1$ sets of polar codes, where each one is a G_B -coset code with parameter $(B, |\tilde{\mathcal{A}}_s|, \tilde{\mathcal{A}}_s, 0)$ with respect to $\text{BEC}(e_s)$, and all the encoded codewords are embedded into messages of Phase 2 in order. Then, in Phase 2 of encoding, we use BSC polar encoders with information set \mathcal{G} to generate the final codeword with length NB . At the receiver end, we need $2S - 1$ number of phases. Phase 1 utilizes the $\text{BSC}(p_S)$ SC decoders to decode blocks with respect to the best channel state (state S in this case). Consider all decoded bits in \mathcal{M}_{S-1} , as well as adding erasures to undecoded blocks, we could decode all erased bits by using $\text{BEC}(e_{S-1})$ SC decoders in Phase 2. Then, using the decoded information as frozen values for blocks with respect to state $S - 1$, $\text{BSC}(p_{S-1})$ SC decoders are adopted in Phase 3 to decode information bits in the blocks

corresponding to channel state $S - 1$. Recursively, all information bits for both BSC encoding and BEC encoding could be reliably decoded, as long as the designed rates of polar codes do not exceed corresponding channel capacities. Hence, by adopting this hierarchical 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 $\delta'(\epsilon) \triangleq \epsilon[1 + H(p_1) - H(p_S)]$ tends to 0 as $\epsilon \rightarrow 0$. Thus, to this end, 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 $\sum_{s=1}^{S-1} |\mathcal{M}_s| \leq N$. For the same reason, the decoding error bound also remains the same as the case of only two fading states. Thus, our proposed polar coding scheme achieves the capacity of fading binary symmetric channel with arbitrary finite number of fading states, and the encoding and decoding complexity are both guaranteed to be tractable in practice.

Noting again the relevancy of this scenario to the fading AWGN channels, we consider another fading channel model with analog noise statistics in the next section, where the polar coding scheme proposed above is utilized.

IV. POLAR CODING FOR FADING ADDITIVE EXPONENTIAL NOISE CHANNEL

A. System Model

We consider the fading additive exponential noise (AEN) channel given by

$$Y_{b,i} = X_{b,i} + Z_{b,i}, \quad b = 1, \dots, B, \quad i = 1, \dots, N, \quad (12)$$

where $X_{b,i}$ is channel input and restricted to be positive and with mean E_X ; N is block length; and B is the number of blocks. In this model, $Z_{b,i}$ are assumed to be identically distributed

within a block and follow an ergodic i.i.d. fading process over blocks. That is, if we consider a fading AEN channel with S states, then, with probability q_s channel noise $Z_{b,i}$ is distributed as an exponential random variable with parameter E_{Z_s} for a given b and all $i \in \{1, \dots, N\}$, i.e.,

$$f_{Z_{b,i}}(z) = \frac{1}{E_{Z_s}} e^{-\frac{z}{E_{Z_s}}}, \quad z \geq 0, \quad (13)$$

where $1 \leq s \leq S$ and $\sum_{s=1}^S q_s = 1$.

We first state the following upper bound on the ergodic channel capacity in the high SNR regime.

Lemma 3. *The ergodic capacity of a fading AEN channel, with channel state information known at the decoder, is upper bounded as follows.*

$$\lim_{E_X \rightarrow \infty} C_{\text{CSI-D}} \leq \sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_X}{E_{Z_s}} \right) \right] \quad (14)$$

Proof: Denote the channel state as a random variable G , which is discrete on set $\{1, 2, \dots, S\}$. If the channel state information is known at the decoder, then we have

$$\begin{aligned} & \lim_{E_X \rightarrow \infty} C_{\text{CSI-D}} \\ & \stackrel{(a)}{\leq} \lim_{E_X \rightarrow \infty} C_{\text{CSI-ED}} \\ & \stackrel{(b)}{=} \lim_{E_X \rightarrow \infty} \max_{\mathbb{E}[X] \leq E_X} I(X; Y|G) \\ & = \lim_{E_X \rightarrow \infty} \max_{\mathbb{E}[X] \leq E_X} h(Y|G) - h(Y|G, X) \\ & = \lim_{E_X \rightarrow \infty} \max_{X_s: \sum_s q_s \mathbb{E}[X_s] \leq E_X} \sum_{s=1}^S q_s [h(X_s + Z_s) - h(Z_s)] \\ & \stackrel{(c)}{=} \lim_{E_X \rightarrow \infty} \max_{E_{X_s}: \sum_s q_s E_{X_s} \leq E_X} \sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_{X_s}}{E_{Z_s}} \right) \right] \\ & \stackrel{(d)}{=} \sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_X}{E_{Z_s}} \right) \right], \end{aligned}$$

where (a) is due to upper bounding the channel capacity with the case where encoder also has CSI and adapts its coding according to the channel states; (b) is the ergodic capacity of the channel where both encoder and decoder has CSI (see, e.g., [17, pages 203-209]); and (c) holds as exponential distribution maximizes the differential entropy on positive support with a mean

constraint [14] [20, page 412]. Here, we choose X_s to be a weighted sum of an exponential distribution with mean $E_{X_s} + E_{Z_s}$ and a delta function in order to make the output $X_s + Z_s$ to be exponentially distributed random variable. That is, the pdf of X_s is given by

$$f_{X_s}(x) = \frac{E_{X_s}}{E_{X_s} + E_{Z_s}} \frac{e^{-x/(E_{X_s} + E_{Z_s})}}{E_{X_s} + E_{Z_s}} u(x) + \frac{E_{Z_s}}{E_{X_s} + E_{Z_s}} \delta(x), \quad (15)$$

where $\delta(x) = 1$ if $x = 0$, and $\delta(x) = 0$ if $x \neq 0$; $u(x) = 1$ if $x \geq 0$, and $u(x) = 0$ if $x < 0$. Finally, (d) follows by taking the limit. ■

In the following, we show that our proposed polar coding scheme achieves the upper bound above in the high SNR regime.

Remark 4. *Note that the capacity of the fading AEN channel with CSI-D approaches to the bound above in the high SNR regime. (For example, our coding scheme, as shown below, provides one such achievable rate.) This observation is similar to the Gaussian counterpart [17, pages 203-209], where in the high SNR regime, the performance obtained from a waterfilling strategy - the optimal solution for the case where encoder can adapt its power based on the channel state, i.e., CSI-ED - approaches to the performance of utilizing the same power allocation for each fading channel.*

Remark 5. *The model above assumes a mean constraint on the channel input where the average is over channel blocks and channel states. If the mean constraint is per block (abbreviated as MPB - Mean Per fading Block - in the following), i.e., $E[X_{b,i}] \leq E_X$ for each fading block b , then by following steps similar to the ones above, we have*

$$C_{\text{CSI-D, MPB}} \leq C_{\text{CSI-ED, MPB}} = \sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_X}{E_{Z_s}} \right) \right].$$

B. Binary Expansion of Exponential Distribution

In [9], expansion coding scheme is proposed for a static AEN channel, where the channel is expanded by the decomposition property of exponential random variable. Here, a similar scheme is adopted for the fading AEN channel. We first start with the following lemma, providing the theoretical basis for expansion coding.

Lemma 6. Let A_l 's be independent Bernoulli random variables with parameters given by a_l , i.e., $\Pr\{A_l = 1\} \triangleq a_l$, and consider the random variable defined by

$$A \triangleq \sum_{l=-\infty}^{\infty} 2^l A_l. \quad (16)$$

Then, the random variable A is exponentially distributed with mean λ^{-1} , i.e., its pdf is given by

$$f_A(a) = \lambda e^{-\lambda a}, \quad a \geq 0, \quad (17)$$

if and only if the choice of a_l is given by

$$a_l = \frac{1}{1 + e^{\lambda 2^l}}. \quad (18)$$

This lemma reveals that one can reconstruct an exponential random variable from a set of independent Bernoulli random variables perfectly. The proof is given in [9], and a set of typical numerical values of a_l s for $\lambda = 1$ is shown in Fig. 5. It is evident that a_l approaches 0 for what we refer to as the “higher” levels and approaches 0.5 for the “lower” levels. Hence, the primary non-trivial levels meaningful for coding are the “middle” ones. This observation provides the basis for truncating the number of levels to a finite value without a significant loss in performance.

C. Expansion Coding

We consider the binary expansion of channel noise

$$\hat{Z}_{b,i} \triangleq \sum_{l=-L_1}^{L_2} 2^l Z_{b,i,l}, \quad (19)$$

where $Z_{b,i,l}$ is a discrete random variable taking value in $\{0, 1\}$. However, the distribution of $Z_{b,i,l}$ depends on the fading state. More precisely, if the noise for a fading block b is exponential with parameter E_{Z_s} , then $Z_{b,i,l}$ is a Bernoulli random variable with parameter

$$\tilde{p}_{l,s} \triangleq \Pr\{Z_{b,i,l} = 1\} = \frac{1}{1 + e^{2^l/E_{Z_s}}}. \quad (20)$$

Then, by Lemma 6, $\hat{Z}_{b,i} \xrightarrow{d} Z_{b,i}$ as L_1 and L_2 tend to infinity. In this sense, we approximate the original exponential noise perfectly by a set of discrete noises.

Similarly, we also expand channel input and output as follows,

$$\hat{Y}_{b,i} \triangleq \sum_{l=-L_1}^{L_2} 2^l Y_{b,i,l} = \sum_{l=-L_1}^{L_2} 2^l (X_{b,i,l} + Z_{b,i,l}), \quad (21)$$

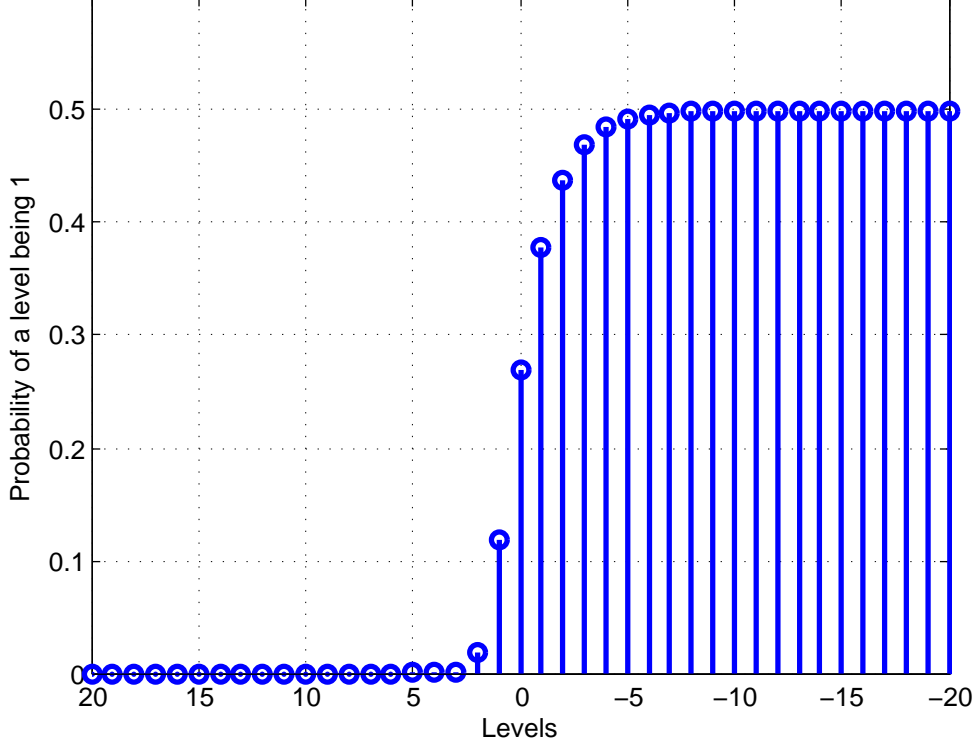


Fig. 5: **Numerical result.** A set of a_l s, the parameters from expanded levels, are shown, where the target random variable expanded from is exponentially distributed with $\lambda = 1$.

where $X_{b,i,l}$ is also a Bernoulli random variable with parameter $p_l \triangleq \Pr\{X_{b,i,l} = 1\}$. At this point, we model the expanded channels as

$$Y_{b,i,l} = X_{b,i,l} + Z_{b,i,l}, \quad l = -L_1, \dots, L_2. \quad (22)$$

Note that the summation is a real sum here, and hence, the channel is not a fading BSC for a given block. If we replace the real sum by modulo-2 sum, then, at level l , any capacity achieving code for fading BSC, for example the one constructed in Section III, can be utilized over this channel with optimal input probability distribution. In addition, instead of directly using the capacity achieving code, one can consider its combination with the method of Gallager [21, pages 74-76] [19] to achieve a rate corresponding to the one obtained by the mutual information $I(X_{b,l}; Y_{b,l})$ evaluated with a desired input distribution on $X_{b,l}$. Hence, a desired rate (evaluation of $I(X_{b,l}; Y_{b,l})$ with some distribution on $X_{b,l}$) at level l and fading block b can be achieved.

However, due to the real sum in the original channel above, carries exist between neighboring

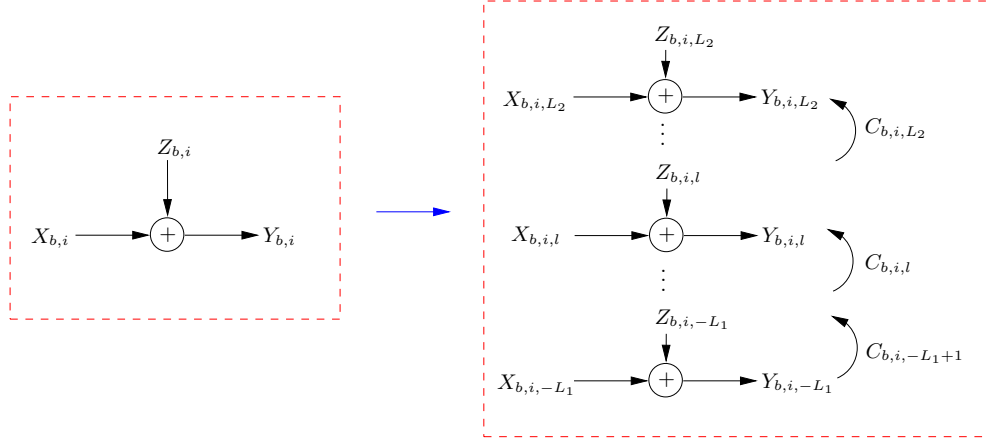


Fig. 6: **Illustration of expansion channel coding scheme.** An analog channel is expanded into a set of binary symmetric channels, where carries are considered between neighboring levels.

levels (see Fig. 6), which further implies that coding over levels are not independent. Hence, we do not have independent parallel channels to start with. Every level, except for the lowest one, is impacted by carries accumulated from lower levels. [9] proposed a scheme to get rid of this issue, where carries are decoded from the lowest level to the highest one. In this way, channels over levels are transformed to behave independently (assuming reliable decoding of each carry w.h.p.), and the total achievable rate is the summation of individual achievable rates over all levels.

Using this technique (to essentially remove carries) as suggested in [9], each level could be modeled as a fading BSC. Thus, expansion coding reduces the problem of coding over a fading exponential noise channel into a set of simpler subproblems, coding over fading BSCs. By adopting capacity achieving polar coding scheme proposed in Section III, we have the following achievable rate result for these channels.

Theorem 7. *By decoding carries in expansion coding, and adopting hierarchical polar coding scheme for fading BSC in each expanded level, the proposed scheme achieves the rate given by*

$$R = \sum_{l=-L_1}^{L_2} \sum_{s=1}^S q_s [H(p_l \otimes \tilde{p}_{l,s}) - H(\tilde{p}_{l,s})], \quad (23)$$

for any $L_1, L_2 > 0$, where $p_l \in [0, 0.5]$ is chosen to satisfy

$$\sum_{l=-L_1}^{L_2} 2^l p_l \leq E_X. \quad (24)$$

We note the followings. First, the achievable scheme we utilize satisfies the mean constraint on the channel input for each block, i.e., averaged over channel uses, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_{i,b} \leq E_X$ for each block b . (This implies satisfying power constraint averaged over the blocks as well.) Secondly, the maximum rate from our coding scheme could be considered as an optimization problem over finite number of parameters p_l , $-L_1 \leq l \leq L_2$. However, it is not clear how to solve this non-convex problem. Here, instead of searching for an optimal solution, we shift our focus to finding a sub-optimal choice of p_l such that the achievable rate is close the optimal one in the high SNR regime. From (15), we observe that the optimal input distribution for the case with the CSI at the transmitter could be approximated with an exponential with parameter E_{X_s} as $\text{SNR} = E_{X_s}/E_{Z_s}$ gets large. As we do not have CSI at the transmitter in our model, we consider choosing the same energy level, E_X , for each fading block. Noting again that the optimal input distribution is unknown for our fading model, the high SNR observation inspires us to choose p_l as

$$p_l = \frac{1}{1 + e^{2^l/E_X}}. \quad (25)$$

The following theorem gives the main result of our polar coding scheme over fading AEN channel. (We accompany the proof of this theorem with Fig. 7 in order to provide not only the details of the proof but also the intuition on how expansion approach is helpful.)

Theorem 8. *For any positive constant $\epsilon < 1$, if*

- $L_1 \geq -\log \epsilon - \min_s \log E_{Z_s}$;
- $L_2 \geq -\log \epsilon + \log E_X$;
- $\min_s \text{SNR}_s \geq 1/\epsilon$, where $\text{SNR}_s \triangleq E_X/E_{Z_s}$,

then by decoding carries and adopting hierarchical polar codes at each fading BSC after expansion, the achievable rate R given by (23), with a choice of p_l as (25), satisfies

$$R \geq \sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_X}{E_{Z_s}} \right) \right] - 5 \log e \cdot \epsilon.$$

Proof: We first state a bound for the entropy of channel noise with mean E_{Z_s} at level l

$$H(\tilde{p}_{l,s}) \leq 3 \log e \cdot 2^{-l+\eta_s} \quad \text{for } l > \eta_s, \quad (26)$$

$$H(\tilde{p}_{l,s}) \geq 1 - \log e \cdot 2^{l-\eta_s} \quad \text{for } l \leq \eta_s, \quad (27)$$

where $\eta_s \triangleq \log E_{Z_s}$. The proofs of these bounds for the case of $E_{Z_s} = 1$ are detailed in Lemma 4 of [9]. Here, we obtain these bounds by following the same steps given there. (Details are omitted for brevity.)

Now, if we denote $\xi \triangleq \log E_X$, then comparing the definitions of p_l and $\tilde{p}_{l,s}$, we get

$$p_l = \frac{1}{1 + e^{2^l/E_X}} = \tilde{p}_{l+\eta_s-\xi,s}. \quad (28)$$

Based on these observations, we have

$$\begin{aligned} & \sum_{l=-L_1}^{L_2} [H(p_l \otimes \tilde{p}_{l,s}) - H(\tilde{p}_{l,s})] \\ & \stackrel{(a)}{\geq} \sum_{l=-L_1}^{L_2} [H(p_l) - H(\tilde{p}_{l,s})] \\ & \stackrel{(b)}{=} \sum_{l=-L_1}^{L_2} [H(\tilde{p}_{l+\eta_s-\xi,s}) - H(\tilde{p}_{l,s})] \\ & = \sum_{l=-L_1+\eta_s-\xi}^{L_2+\eta_s-\xi} H(\tilde{p}_{l,s}) - \sum_{l=-L_1}^{L_2} H(\tilde{p}_{l,s}) \\ & = \sum_{l=-L_1+\eta_s-\xi}^{-L_1-1} H(\tilde{p}_{l,s}) - \sum_{l=L_2+\eta_s-\xi+1}^{L_2} H(\tilde{p}_{l,s}) \\ & \stackrel{(c)}{\geq} \sum_{l=-L_1+\eta_s-\xi}^{-L_1-1} [1 - \log e \cdot 2^{l-\eta_s}] - \sum_{l=L_2+\eta_s-\xi+1}^{L_2} 3 \log e \cdot 2^{-l+\eta_s} \\ & \stackrel{(d)}{\geq} \xi - \eta_s - \log e \cdot 2^{-L_1-\eta_s} - 3 \log e \cdot 2^{-L_2+\xi} \\ & \stackrel{(e)}{\geq} \log \left(\frac{E_X}{E_{Z_s}} \right) - \log e \cdot \epsilon - 3 \log e \cdot \epsilon \\ & \stackrel{(f)}{\geq} \log \left(1 + \frac{E_X}{E_{Z_s}} \right) - \log e \cdot \frac{E_{Z_s}}{E_X} - \log e \cdot \epsilon - 3 \log e \cdot \epsilon \\ & \stackrel{(g)}{\geq} \log \left(1 + \frac{E_X}{E_{Z_s}} \right) - 5 \log e \cdot \epsilon, \end{aligned} \quad (29)$$

where

- (a) is due to $p_l \otimes \tilde{p}_{l,s} \triangleq p_l(1 - \tilde{p}_{l,s}) + \tilde{p}_{l,s}(1 - p_l) \geq p_l$, and then due to the fact that entropy function is increasing on $[0, 0.5]$ (and, we have $p_l \otimes \tilde{p}_{l,s} \leq 0.5$);
- (b) follows from equation (28);
- (c) follows from bounds (26) and (27);
- (d) follows as

$$\sum_{l=-L_1+\eta_s-\xi}^{-L_1-1} 2^{l-\eta_s} \leq 2^{-L_1-\eta_s},$$

and

$$\sum_{l=L_2+\eta_s-\xi+1}^{L_2} 2^{-l+\eta_s} = \sum_{l=-L_2+\eta_s}^{-L_2+\xi-1} 2^l \leq 2^{-L_2+\xi};$$

- (e) follows from theorem assumptions that $L_1 \geq -\log \epsilon - \min_s \eta_s$, and $L_2 \geq -\log \epsilon + \xi$;
- (f) is due to the fact that $\log(1 + E_X/E_{Z_s}) - \log(E_X/E_{Z_s}) = \log(1 + E_{Z_s}/E_X) \leq \log e \cdot E_{Z_s}/E_X$ (as $\ln(1+x) \leq x$ for any $x \geq 0$);
- (g) is due to the assumption in theorem that $\min_s \text{SNR}_s \geq 1/\epsilon$.

Then, using (29) in (23) of Theorem 7, we have

$$\begin{aligned} R &= \sum_{s=1}^S q_s \left\{ \sum_{l=-L_1}^{L_2} [H(p_l \otimes \tilde{p}_{l,s}) - H(\tilde{p}_{l,s})] \right\} \\ &\geq \sum_{s=1}^S q_s \left\{ \log \left(1 + \frac{E_X}{E_{Z_s}} \right) - 5 \log e \cdot \epsilon \right\} \\ &= \sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_X}{E_{Z_s}} \right) \right] - 5 \log e \cdot \epsilon. \end{aligned}$$

■

Remark 9. We note that the proposed scheme achieves a rate

$$\sum_{s=1}^S q_s \left[\log \left(1 + \frac{E_X}{E_{Z_s}} \right) \right] = C_{\text{CSI-ED, MPB}},$$

which is an upper bound on the capacity for the CSI-D case in the high SNR regime. (See Lemma 3.) Therefore, the proposed scheme achieves the capacity in the high SNR regime.

Note that, for an exponential distribution with mean $1/\lambda$, its binary expansion result can be considered as the expansion of an exponential distribution with mean 1 shifted by $\log(1/\lambda)$ number of levels. We show this phenomenon in Fig. 7. To this end, Theorem 8 shows that in order

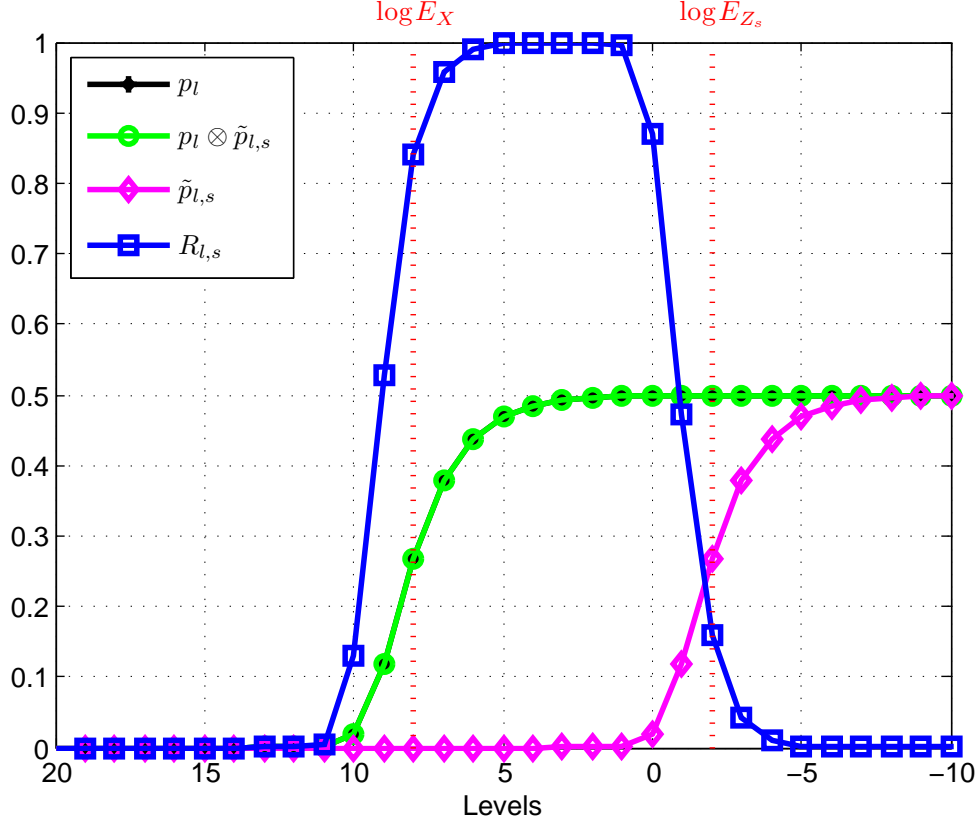


Fig. 7: **Illustration of signal, noise and rate at each level.** In this numerical result, we only concern about a single state s , and choose $E_X = 2^8$, $E_{Z_s} = 2^{-2}$. Note that, in this case, the curve of $p_l \otimes \tilde{p}_{l,s}$ (in black) almost coincides with the one of p_l (in green), because of high SNR. Moreover, the signal p_l (in green) is a left shifted version of noise $\tilde{p}_{l,s}$ (in purple) by $\log E_X - \log E_{Z_s}$ number of levels. Achievable rates at each level, $R_{l,s} \triangleq H(p_l \otimes \tilde{p}_{l,s}) - H(\tilde{p}_{l,s})$, are represented in blue.

to achieve the capacity of fading AEN channel, first, SNR should be large enough, and secondly, the number of expanded levels should also be large enough such that the highest level exceeds all the left shifted levels of expanded signal, and the lowest level exceeds the right shifted levels of expanded noises. Hence, in total, basically we need $\log \text{SNR}_{\max}$ ($\text{SNR}_{\max} \triangleq \max_s E_X / E_{Z_s}$) number of levels to cover all “non-trivial” levels for coding, as well as extra $-2 \log \epsilon$ number of levels to shoot for accuracy. At this point, the complexities of encoding and decoding are both given by $O((\log \text{SNR}_{\max} - 2 \log \epsilon) NB \log(NB))$, where $O(NB \log(NB))$ is the complexity

scale for fading BSC derived in the previous section.

D. Numerical Results

In this section, we analyze the rate obtained from Theorem 7 with parameter p_l chosen as (25). Numerical results are illustrated in Fig. 8, where we consider the case of two fading states. It is evident from the figure, and also from the theoretical analysis given in Theorem 8, that our proposed polar coding scheme together with expansion coding achieves the upper bound on the channel capacity (Lemma 3) in the high SNR regime. Therefore, the proposed coding scheme achieves the channel capacity for sufficiently large SNR.

We also note that the coding scheme does not perform well in the low SNR regime, which mainly results from two reasons. First, the upper bound we derived in Lemma 3, which is the target rate in our coding scheme, is not tight in the low SNR regime. Secondly, our choice of p_l only behaves as a good approximation for sufficiently high SNR, which limits the proposed scheme to be effective at the corresponding regime. However, as evident from the numerical results, for a fairly large set of SNR values the proposed scheme is quite effective. In addition, the upper bound curve is equal to $C_{\text{CSI-ED, MPB}}$, the capacity when the input mean constraint is imposed per block (instead of averaging over the blocks). Therefore, for the scenario of having input constraint per each fading block, the upper bound $C_{\text{CSI-D, MPB}} \leq C_{\text{CSI-ED, MPB}}$ holds at any SNR, and the only degradation in our coding scheme is due to the second point discussed above.

V. CONCLUSION

In this paper, polar coding schemes for fading binary symmetric channel (BSC) and fading additive exponential noise (AEN) channel are proposed. First, a hierarchical polar coding scheme is proposed for the fading BSC. This novel scheme, by exploiting an erasure decoding approach at the receiver, utilizes the polarization results of different BSCs. (These BSCs are defined over channel uses at a given fading block and over fading blocks at a given channel use index.) This novel polar coding technique is shown to be capacity achieving for fading BSC. Remarkably, the proposed scheme does not assume channel state information at the transmitter and fading BSC models the fading additive white Gaussian noise (AWGN) channel with a BPSK modulation. Therefore, our results are quite relevant to the practical channel models considered in wireless communications.

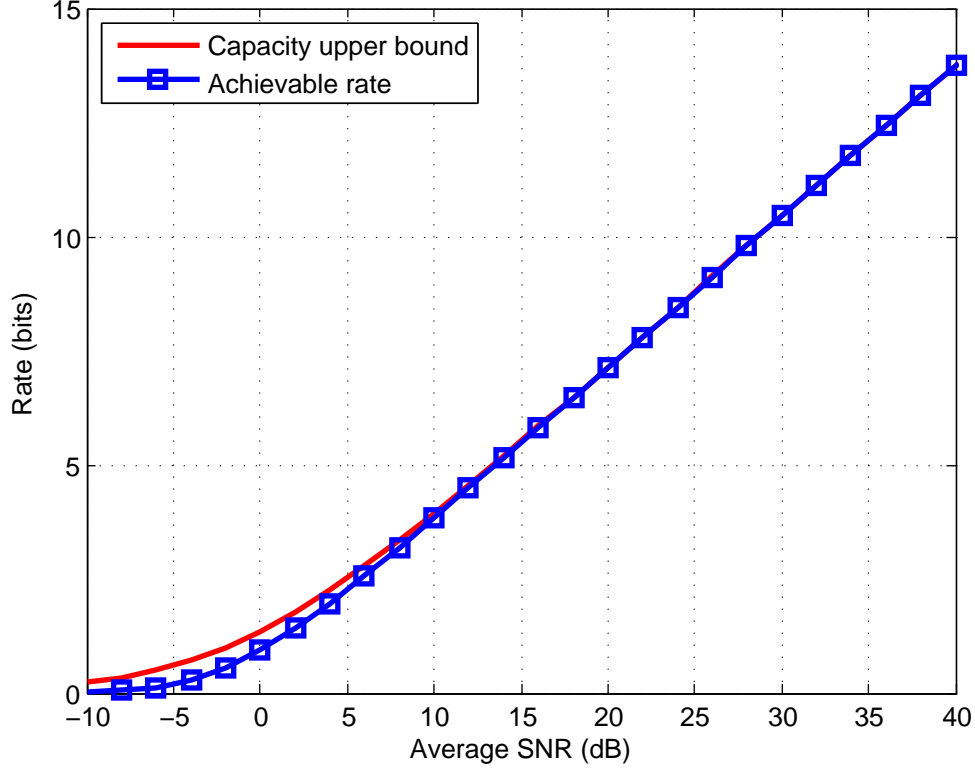


Fig. 8: **Numerical results.** The upper bound of ergodic capacity, $C_{\text{CSI-ED, MPB}}$, which is equal to $C_{\text{CSI-ED}}$ for sufficiently large SNR, is given by the red curve. The achievable rate is given by the blue curve. In this analysis, only two fading states are concerned, and the parameters are chosen as $E_{Z_1} = 0.5$, $E_{Z_2} = 3$, $q_1 = 0.8$, and $q_2 = 0.2$. Average SNR is defined as $E_X / (\sum_{s=1}^S q_s E_{Z_s})$.

Towards utilizing the proposed techniques for encoding over another fading channel model, we focused on fading AEN channel. For this model, expansion coding [9] is adopted to convert the problem of coding over an analog channel into coding over discrete channels. By performing this expansion approach and making the resulting channels independent (via decoding the underlying carries), a fading AEN channel is decomposed into multiple independent fading BSCs (with a reliable decoding of the carries). By utilizing the hierarchical polar coding scheme for fading BSC, both theoretical proof and numerical results showed that the proposed approach achieves the capacity of this fading channel in the high SNR regime.

We remark that the advantages of polar codes in rate and complexity are both inherited in the proposed coding schemes. More precisely, as polar codes achieve channel capacity of BSC and

BEC, our hierarchical utilization of polar codes also achieves the capacity of fading BSC, and this result is further utilized to guarantee that expansion coding scheme can achieve the capacity of fading AEN channel in the high SNR regime (with low complexity in all cases).

Although the discussion in this paper focuses only on fading BSC and fading AEN channel, the proposed coding scheme could be generalized to more general cases. For example, by utilizing non-binary polar codes, our polar coding scheme can be generalized to a fading non-binary discrete symmetric channel. This result can then be utilized for an AWGN fading channel with more constellation points, such as QPSK. Moreover, the expansion coding scheme can also be used for other analog channels that have noise statistics other than exponential, e.g., Gaussian. Here, even though these distributions may not be perfectly approximated by a set of independent discrete random variables, expansion coding scheme can still perform well, especially at high SNR.

Finally, we note that the proposed coding scheme requires long codeword lengths to make the error probability arbitrarily small. This requirement translates to requiring long coherence intervals and large number of fading blocks as our approach utilizes coding over both channel uses and fading blocks. (This is somewhat similar to the analyses in Shannon theory, where the guarantee of the coding is that the error probability vanishes as the block length gets large.) Therefore, our coding scheme fits to the fading channels with moderate/long coherence time and large number of fading blocks. Here, we comment on applicability of the proposed coding scheme in typical wireless systems. As reported in [22, page 219], LTE systems operating at 1.8GHz frequency with 20MHz bandwidth typically have fading durations of 2.8×10^5 to 1.0×10^7 channel uses. In addition, WiFi systems operating at 5GHz frequency with 20MHz bandwidth typically have fading durations of 7.7×10^5 to 1.8×10^7 channel uses [23, pages 98-99]. (Here, a mobile speed of 1m/s is assumed for both systems.) Polar codes, on the other hand, typically have error rates around 10^{-6} when the blocklength is around 2^{10} , and a smaller error probability is even possible, when the decoding is implemented with a better decoder [13]. For instance, instead of the classical SC decoder, a list decoder [13] can be utilized. Finally, besides long coherence intervals, another requirement for the proposed coding scheme is to have large number of fading blocks. This requirement can be satisfied in many practical scenarios at the expense of having large decoding delays. To summarize, for a given wireless system and a choice of one of the encoding/decoding strategies discussed above, the resulting error rate and its

propagation in the proposed decoding algorithm should be studied further. We leave the analysis of such applications of the proposed techniques to a future work.

REFERENCES

- [1] E. Arıkan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):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)*, pages 1493–1495, Jun. 2009.
- [3] E. Şaşıoğlu, E. Arıkan, and E. Telatar. Polarization for arbitrary discrete memoryless channels. In *Proc. 2009 IEEE Information Theory Workshop (ITW 2009)*, pages 144–148, Oct. 2009.
- [4] E. Arıkan. Source polarization. In *Proc. 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, pages 899–903, Jun. 2010.
- [5] S. B. Korada and R. L. Urbanke. Polar codes are optimal for lossy source coding. *IEEE Transactions on Information Theory*, 56(4):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)*, pages 909–912, Jun. 2010.
- [7] E. Abbe and A. Barron. Polar coding schemes for the AWGN channel. In *Proc. 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, pages 194–198, Jul. 2011.
- [8] E. Abbe and E. Telatar. Polar codes for the m -user multiple access channel. *IEEE Transactions on Information Theory*, 58(8):5437–5448, Aug. 2012.
- [9] 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)*, pages 1932–1936, Jul. 2012.
- [10] Joseph J. Boutros and Ezio Biglieri. Polarization of quasi-static fading channels. In *Proc. 2013 IEEE International Symposium on Information Theory Proceedings (ISIT 2013)*, pages 769–773, Jul. 2013.
- [11] Angel Bravo-Santos. Polar codes for the rayleigh fading channel. *IEEE Communications Letters*, 17(12):2352–2355, Dec. 2013.
- [12] E. Arıkan. Systematic polar coding. *IEEE Communications Letters*, 15(8):860–862, Aug. 2011.
- [13] B. Li, H. Shen, and D. Tse. An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check. *IEEE Communications Letters*, 16(12):2044–2047, Dec. 2012.
- [14] S. Verdú. The exponential distribution in information theory. *Problems of Information Transmission*, 32(1):100–111, Jan. 1996.
- [15] A. Martinez. Communication by energy modulation: The additive exponential noise channel. *IEEE Transactions on Information Theory*, 57(6):3333–3351, Jun. 2011.
- [16] S. Y. Le Goff. Capacity-approaching signal constellations for the additive exponential noise channel. *IEEE Wireless Communications Letters*, 1(4):320–323, Aug. 2012.
- [17] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge University Press, 2005.
- [18] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge University Press, 2011.
- [19] S. B. Korada. *Polar codes for channel and source coding*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2009.
- [20] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 1991.

- [21] R. G. Gallager. *Information theory and reliable communication*. John Wiley & Sons, 1968.
- [22] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.
- [23] Eldad Perahia and Robert Stacey. Next generation wireless LANs. *Throughput, robustness and reliability in 802.11 n*, 2008.