

RESEARCH

Open Access



Polar codes for non-identically distributed channels

Jangseob Kim and Jungwoo Lee*

Abstract

We introduce new polar coding schemes for independent non-identically distributed parallel binary discrete memoryless channels. The first scheme is developed for the case where underlying channels are time invariant (the case of a deterministic channel parameters), while the other schemes deal with a scenario where underlying channels change based on a distribution (the case of random channel parameters). For the former case, we also discuss the importance of the usage of an interleaver Q to enhance system reliability, and for the latter case, we model the channel behavior of binary erasure channels. It is shown that the proposed polar coding schemes achieve a symmetric capacity.

Keywords: Polar codes, Non-identically distributed channels, Channel parameters, An interleaver scheme

1 Introduction

The polar codes previously introduced in [1] are a coding scheme that achieves the symmetric capacity of any discrete memoryless channel (DMC) by exploiting channel polarization phenomena. The block error rate converges to zero as the code length N goes to infinity.

One disadvantage of polar codes is their dependency on the underlying channel. If a channel is unknown to a transceiver, communication becomes unreliable and erroneous decoding can result, since the code design is based on a different channel. The performance gap due to imperfect channel knowledge was analyzed in [2], and a robust approximation technique has been proposed for this mismatch. In [3], it was demonstrated that polarization is possible even under a non-stationary channel.

In this paper, we propose an approach to the design of polar codes in the parallel channel model based on a martingale process and extend the model to the case in which the channel parameters are no longer fixed but exhibit some random behaviors. In addition, we briefly discuss the importance of the usage of the proper interleaver Q .

The communication scenario in which the transmitter and the receiver do not know the channel parameter but know the domain (set) to which it belongs is known as the compound channel scenario [4]. The authors of [4] defined the compound channel capacity as the rate

at which one can reliably transmit data without knowing the underlying channel parameter. In [5] and [6], polar codes that achieve the compound channel capacity were proposed. In [5], the unknown channel is deterministic during a codeword transmission, and in [6], the authors dealt with a deterministic compound parallel channels model.

Parallel channels are used to model statistical behaviors of communications in conveying multiple data simultaneously between nodes, connected with multiple links. The internal structure of storage systems or singular value decomposed multiple-input multiple-output (MIMO) channels are an example of parallel channel models. In these channels, the links do not need to be the same, and their statistical behaviors can differ. The question then arises about whether the parallel channel capacity is achievable via polar codes. In [7, 8], it is demonstrated, under a non-identically distributed channel assumption, that when all channel parameters (CP) are given to the transceiver, polar codes can achieve the capacity in both degraded and non-degraded channel settings.

In this paper, we show that the polarization phenomenon occurs in the parallel channel model and provide proof of achievability of polar codes for general binary discrete memoryless channels (B-DMC), under a non-identically distributed assumption, where deterministic CPs are given. Second, in contrast to previous literature, we also consider a case where CPs that describe binary erasure channels (BECs) are realizations of some

*Correspondence: junglee@snu.ac.kr
Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea

random variables, and only their probability distributions, and not their exact values, are known to the transmitter and the receiver.

This is similar to the compound channel scenario, since the transmitter does not know the CPs but differs since the CPs, which are a realization of random variables, might differ. The system model is depicted in Fig. 1. The role of the interleaver in this figure is discussed in the last part of the second section.

Developing polar codes for the randomly varying channel scenario is one of the noticeable issues. For the i th transmit BEC at time instant η , assume that an erasure probability $\epsilon_i(\eta)$ is a realization of a random variable ϵ that follows a distribution f_ϵ . In this case, conventional polar coding schemes fail to achieve the capacity since the encoder and the decoder do not know the exact value, $\epsilon_i(\eta)$, of the underlying channel. For example, in flash memories, the statistical responses of the voltage threshold may change asymmetrically with time and with the number of accesses to the cells. As the storage capacity increases, it is inefficient for a storage controller to trace and probe the exact states of every block or cell.

The remainder of this paper is organized as follows. In the second section, we analyze the polarization process on a time-invariant non-identical parallel channel. This corresponds to a parallel channel with deterministic CPs which could differ to each other. In the third and fourth sections, an independently non-identically distributed channel model with random CPs is discussed. In these sections, we assume that only the distributions of the CPs of each channel, $W_{(i)}$, ($\forall i \in [1, N]$), are known to an encoder and decoder. In the third section, the CPs may change in a block-by-block manner; however, their states are maintained within a block. In the fourth section, in contrast to the third section, CPs may change even within a block. In the fifth section, we discuss a special case of partially dependent channels and show that polar codes

based on the q -ary input can achieve the capacity. In the final section, concluding remarks are provided.

2 Non-identical channels with deterministic CP

Let $W : \mathcal{X} \rightarrow \mathcal{Y}$ denote a general symmetric binary input memoryless channel (B-DMC) and $W_N : \mathcal{X}^N \rightarrow \mathcal{Y}^N$ denote a vector channel. If channels are independent but not identical, then $W_N(y_1^N | x_1^N) = \prod_{i=1}^N W_{(i)}(y_i | x_i)$ where $W_{(i)} : \mathcal{X}_i \rightarrow \mathcal{Y}_i$, such that their transition probabilities $p_{(i)}(y|x)$ and $p_{(j)}(y|x)$ may differ if $i \neq j$. In parallel channels, a non-identical channel may correspond to a transmission scenario through links with different qualities. In terms of the channel model, this can be interpreted as a fast-fading channel in a time or frequency domain. This corresponds to the case when the time duration or the frequency gap between adjacent channels is larger than the coherence time or coherence frequency, respectively.

As in [1], given any B-DMC $W_{(i)}$, the same definitions of the symmetric capacity and the Bhattacharyya parameter are adopted as performance measures

- Symmetric capacity

$$I(W_{(i)}) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W_{(i)}(y|x) \log \frac{W_{(i)}(y|x)}{\frac{1}{2} W_{(i)}(y|0) + \frac{1}{2} W_{(i)}(y|1)} \quad (1)$$

- Bhattacharyya parameter

$$Z(W_{(i)}) \triangleq \sum_{y_i \in \mathcal{Y}} \sqrt{W_{(i)}(y_i|0) W_{(i)}(y_i|1)}. \quad (2)$$

According to Proposition 1 in [1], the two parameters satisfy

$$\log \frac{2}{1 + Z(W_{(i)})} \leq I(W_{(i)}) \leq \sqrt{1 - Z(W_{(i)})^2}. \quad (3)$$

For later use, we denote the bit channel $W_N^{(i)}$ by

$$W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) \triangleq \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-i}} W_N(y_1^N | u_1^N),$$

where $N = 2^n$ is the code length.

In this section, channels are assumed to be independent and not identically distributed, and the CPs are known to the transmitter and the receiver. For BEC with ϵ_1^N , these erasure probabilities may differ and are known in advance to the encoder and the decoder. Let us denote the sum symmetric capacity by $I_s = \sum_{i=1}^N I_i$, where $I_i = I(X_i; Y_i)$, and the sample mean by $\mathbb{E}[I_i] = \frac{I_s}{N}$. Then, the following theorem holds.

Theorem 1 *For any set of B-DMCs $\{W_{(i)}\}$, $i \in [1, N]$, for arbitrary small $\delta \leq 0$, there exist polar codes that achieve*

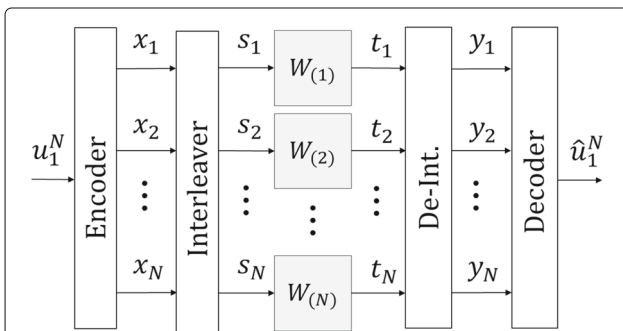


Fig. 1 System model. Non-identically distributed parallel channels. Interleaver Q is inserted to make a set of virtually ordered transmit channels from $\{W_{(i)}\}$ to $\{W'_{(i)}\}$, $i \in [1, N]$. By applying the interleaver, $\{W'_{(i)}\}$ is sorted in a way that maximize the reliability and the achievable rate

sum capacity I_s , in the sense that as N , which is the power of 2, goes to infinity, the fraction of indices $i \in [1, N]$ satisfies

$$\frac{|\{i|I(W_N^{(i)}) \in (1-\delta, 1]\}|}{N} \rightarrow \frac{I_s}{N}$$

$$\frac{|\{i|I(W_N^{(i)}) \in [0, \delta)\}|}{N} \rightarrow 1 - \frac{I_s}{N}.$$

For simple notations, denote $Z_N^{(i)} \triangleq Z(W_N^{(i)})$ and $I_N^{(i)} \triangleq I(W_N^{(i)})$. To prove the above theorem, we first need to clarify the recursive structures of $W_N^{(i)}$, $Z_N^{(i)}$, and $I_N^{(i)}$. Second, we need to prove that the values of $I_N^{(i)}$ and $Z_N^{(i)}$ converge to $\{0, 1\}$ as the code length N increases. For the second proof, the martingale convergence theorem and (3) will be used. These statements are proved via Lemma 1, Lemma 2, and Proposition 1. After proving Lemma 2, we summarize the proof of the theorem.

For a kernel $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ of the generator matrix $G_N = B_N F^{\otimes n}$, the recursively evolving structure of $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ is similar to that of the recursive equation in [1], except for the last recursions for the length 2.

Lemma 1 Suppose $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ for some set of binary input channels. Then,

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-1} | u_{2i}) = \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} | u_{2i})$$

$$\cdot W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} | u_{2i-1} \oplus u_{2i}) \quad (4)$$

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} | u_{2i})$$

$$= \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} | u_{2i-1} \oplus u_{2i}) \quad (5)$$

$$\cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} | u_{2i}),$$

where the last recursive channel relations is a mapping such that $(W_{(j)}, W_{(j+1)}) \mapsto (W_2^{(1)}, W_2^{(2)})$

$$W_2^{(1)}(y_j^{j+1} | u_j) = \sum_{u_{j+1}} \frac{1}{2} W_{(j)}(y_j | u_j \oplus u_{j+1})$$

$$\cdot W_{(j+1)}(y_{j+1} | u_{j+1}) \quad (6)$$

$$W_2^{(2)}(y_j^{j+1}, u_j | u_{j+1}) = \frac{1}{2} W_{(j)}(y_j | u_j \oplus u_{j+1})$$

$$\cdot W_{(j+1)}(y_{j+1} | u_{j+1}). \quad (7)$$

The evolution of the parallel channel transition probabilities is depicted in Fig. 2. Observe that the structure is

still recursive, but the number of inputs and the outputs of each evolution is identical.

The proof of the Lemma 1 directly follows from [1]. The last recursion derives from the independently and non-identically distributed parallel channel environment.

We now extend Propositions 4–7 in [1], which were proved under the i.i.d. condition, into the non-identically distributed DMC case. This extension was performed in previous literatures, in the names of “parallel channels” [7, 8] and “non-stationary channels” [3] with the measure of the symmetric capacity.

2.1 Evolution of symmetric capacities

In this subsection, we consider the symmetric capacities of the recursively achieved bit channels. By inserting (6) and (7) into the definition of the symmetric capacities (1), the following proposition holds.

Proposition 1 Suppose $(W_{(1)}, W_{(2)}) \mapsto (W_2^{(1)}, W_2^{(2)})$ for any binary input discrete channels. Then,

$$I(W_2^{(1)}) + I(W_2^{(2)}) = I(W_{(1)}) + I(W_{(2)}) \quad (8)$$

$$I(W_2^{(2)}) \geq \max\{I(W_{(1)}), I(W_{(2)})\} \quad (9)$$

$$I(W_2^{(1)}) \leq \min\{I(W_{(1)}), I(W_{(2)})\}. \quad (10)$$

Proof of (8) Equation (8) can be proved as follows:

$$I(W_2^{(1)}) + I(W_2^{(2)}) = I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1)$$

$$= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2 | U_1)$$

$$= I(U_1, U_2; Y_1, Y_2)$$

$$= I(X_1, X_2; Y_1, Y_2)$$

$$= I(X_1; Y_1) + I(X_2; Y_2)$$

$$= I(W_{(1)}) + I(W_{(2)}).$$

Since the mapping from a message vector U_1^2 to an encoded codeword is deterministic through a generator matrix G_2 , no information loss occurs between the third equation and the forth equation. Also, the fifth equation derives from the independency among the transmit channels. \square

To prove (9), we focus on the second bit channel $W_2^{(2)}$

$$I(W_2^{(2)}) = I(U_2; Y_1, Y_2, U_1)$$

$$= I(U_2; Y_2) + I(U_2; Y_1, U_1 | Y_2)$$

$$= I(X_2; Y_2) + I(U_2; Y_1, U_1 | Y_2)$$

$$= I(W_{(2)}) + I(U_2; Y_1, U_1 | Y_2)$$

$$\geq I(W_{(2)}). \quad (11)$$

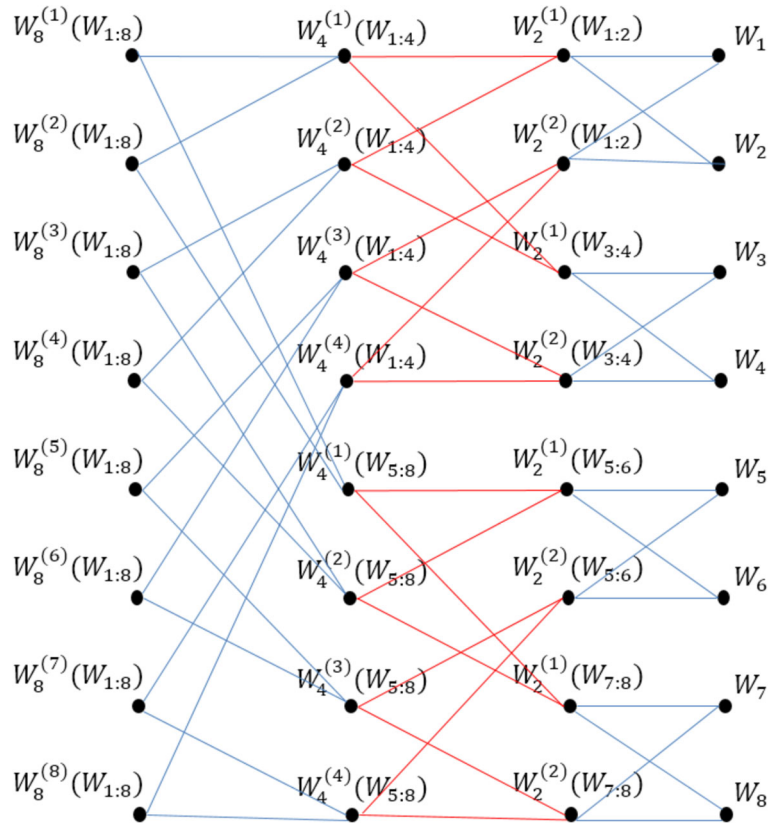


Fig. 2 Recursive parallel channel transitions $N = 8$. Note that the number of inputs and the outputs of each evolution is identical

Since the value of mutual information is always non-negative, $I(\cdot) \geq 0$. By inserting (11) into (8), we obtain $I(W_2^{(1)}) \leq I(W_{(1)})$.

However, ambiguity remains about whether (9) and (10) are true based on this relation and (11), since there are no additional conditions on the qualities of transmit channels $W_{(i)}$, $i \in [1, N]$ that make them order with measures of the symmetric capacity $I(\cdot)$ or the Bhattacharyya parameter $Z(\cdot)$.

Now, we can consider three necessary conditions for (9) and (10) to be true: (1) $I(W_{(1)}) \leq I(W_2^{(2)})$, (2) $I(W_{(2)}) \geq I(W_2^{(1)})$, and (3) $I(W_{(1)}) \leq I(W_{(2)})$.

Conditions 1 and 2 can be easily verified and hold for any set of BECs, binary symmetric channels (BSCs), and binary-input additive white Gaussian noise (BI-AWGN), and thus satisfy (9) and (10). However, the first and the second inequalities do not hold in general B-DMCs. Then, the third condition, $I(W_{(1)}) \leq I(W_{(2)})$, should be satisfied to achieve (9) and (10). For channels of $N = 2$, where $I(W_{(1)}) \geq I(W_{(2)})$, the third condition can be achieved simply by swapping $(W_{(1)}, W_{(2)}) \mapsto (W_{(2)}, W_{(1)})$ without degrading the achievable rate. This operation is available since the channel parameters are exposed to the

transceiver in advance. Therefore, we can conclude that (9), (10), and Proposition 1 are true for any B-DMCs.

In general, we can extend these relations to the case of $N = 2^n$ ($n \geq 1$). The following relations are thus true for all given channel parameters.

$$I_{2N}^{(2i)} \leq \min \left\{ I_N^{(i)}(W_{[1:N]}), I_N^{(i)}(W_{[N+1:2N]}) \right\} \quad (12)$$

$$I_{2N}^{(2i-1)} \geq \max \left\{ I_N^{(i)}(W_{[1:N]}), I_N^{(i)}(W_{[N+1:2N]}) \right\} \quad (13)$$

$$I_{2N}^{(2i-1)} + I_{2N}^{(2i)} = I_N^{(i)}(W_{[1:N]}) + I_N^{(i)}(W_{[N+1:2N]}). \quad (14)$$

The equality holds between $I_{2N}^{(2i)}$ and $I_{2N}^{(2i-1)}$ if and only if the underlying channels are either perfect or completely noisy.

According to (12)–(14), we can observe that the gap between these two evolved symmetric channel capacities $I_{2N}^{(2i-1)}$ and $I_{2N}^{(2i)}$ increases as recursions are repeated (or as the code length N is doubled). In addition, their values are lower and higher than those of the previous values, respectively. Recalling that $I(\cdot) \in [0, 1]$, a conjecture can

be formed suggesting that the evolved $I_{2N}^{(2i-1)}$ and $I_{2N}^{(2i)}$ converge to one of the extremal values, 0 or 1, as the process is repeated. We can now prove this conjecture with the aid of the concept of the martingale process and bounded martingale convergence theorem in the following subsection.

2.1.1 Martingale process I_n

Following the notation of the random tree process in [1], we first define the random sequence $\{I_n\}$ such that $I_N^{(i)} \mapsto I_n(b_1^n)$, where $I_n(b_1^n) \in \{I_n\}$ is represented as $I_{b_1^n}$. Let us introduce an average sequence E_n to replace $I_{b_1^n}$ such that $E_n(b_1^n) = \mathbb{E}[I_{b_1^n}]$. Based on the channel transitions in Fig. 2 and the definition of the average sequence, we can draw the alternative graphical transitions for each stage n as depicted in Fig. 3. The following lemma then shows that $\{I_n\}$ is a martingale process.

Lemma 2 $\{I_n\}$ is a martingale under average sequence $\{E_n\}$ in the sense that $\mathbb{E}[I_{n+1}|E_n] = E_n$, where $I_{b_1^n} \in \{I_n\}$ and $E_n(b_1^n) = \mathbb{E}[I_{b_1^n}]$

Proof Lemma 2 can be proved from the chain rule of mutual information that preserves rate.

$$I_{2N}^{(2i-1)} + I_{2N}^{(2i)} = I_N^{(i)}(W_{[1:N]}) + I_N^{(i)}(W_{[N+1:2N]}). \quad (15)$$

By taking an average on both sides, one can get

$$\mathbb{E}[I_n|b_1^n] = \frac{1}{2}((I_n(b_1^n), W_{[1:N]}) + I_n(b_1^n), W_{[N+1:2N]})) \quad (16)$$

which is equivalent to $E_n(b_1^n)$.

We can now exploit the ergodicity in B-DMCs. Let us denote the sum symmetric capacity as $I_s = \sum_1^N I_i$, where $I_i = I(X_i; Y_i)$, and the sample mean as $\mathbb{E}[I_i]$. Then, from

the ergodicity, $\mathbb{E}[I_i] = \frac{I_s}{N}$. Next, we borrow the tree process representation and the corresponding notations from Section 4 in [1].

A bit-channel $W_{N=2^n}^{(i)}$ can be uniquely represented via a binary sequence $b_1^n: W_{b_1^n}$. For example, $W_8^{(3)} \mapsto W_{011}$. Using this, we abbreviate $I(W_{b_1^n}^{(i)}) = I(W_{2^n}^{(i)})$ into $I_{b_1^n}$, where $I_{b_1^n}$ is an element of a set I_n . We use the same notation in the definition of the average sequence E_n ; $E_n(b_1^n) = \mathbb{E}[I_{b_1^n}] = \mathbb{E}[I_n]$ which is the average over all the elements of the set I_n . By exploiting E_n , we can draw a new tree process as depicted in Fig. 3, and we have shown that this process satisfies the property of a martingale process via (16).

Recalling the following properties, we can conclude that E_n converges to 0 or 1: (i) the symmetric capacity is bounded by 0 and 1, such that $0 \leq I_{b_1^n} \leq 1$, and (ii) a bounded martingale process converges to extremal values, which is the same technique as that used in the proof of [1].

Once E_n converges to one of the extremal values, we can also conclude that I_n converges to the same extremal, since E_n is defined as the average of I_n . From (16), the two newly evolving terms are equidistant from $E_n(b_1^n)$, so we denote them using distances α : $E_n(b_1^n) + \alpha$ and $E_n(b_1^n) - \alpha$. When the martingale $E_n(b_1^n)$ goes to 1, we have $\alpha \rightarrow 0$ since it is bounded above by 1. Similarly, when the martingale $E_n(b_1^n)$ goes to 0, we have $\alpha \rightarrow 0$ since it is bounded below by 0. Therefore, we conclude that I_n also converges to 0 or 1.

Consequently, it is concluded that sequence $\{I_n\}$ is a martingale sequence. \square

Since $\{I_n\}$ is a bounded martingale sequence, it converges to a random variable I_∞ with probability 1. Furthermore, from the martingale property, $\mathbb{E}[I_\infty] = \mathbb{E}[I_0]$,

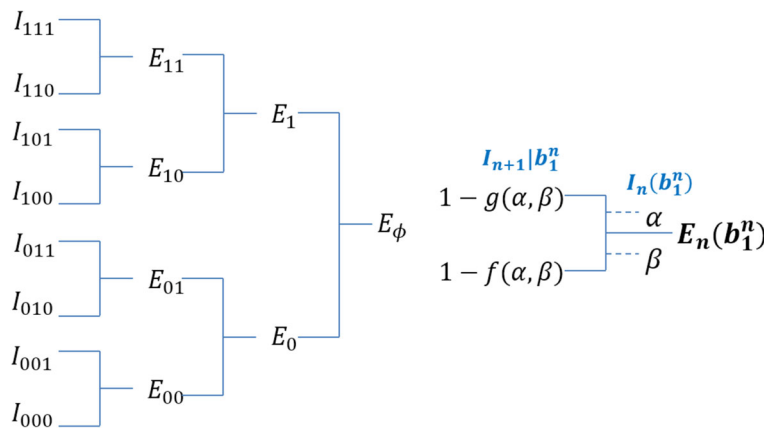


Fig. 3 Graphical representation of the evolution of $I_N^{(i)}$ for $N = 8$ parallel BECs

which is equivalent to $\frac{I_s}{N}$. By combining this relation with the martingale property of the random sequence $\{I_n\}$, we have

$$P(I_\infty = 0) = 1 - \frac{I_s}{N} \quad (17)$$

$$P(I_\infty = 1) = \frac{I_s}{N} \quad (18)$$

which concludes the proof of Theorem 1.

In summary, to prove Theorem 1, we showed that first, polarization takes place by using the martingale process, and second, the ratio of noiseless bit channels converges to the symmetric capacity as $N \rightarrow \infty$. This is because this equivalently indicates that the block error converges to zero as the code length goes to infinity as in [1].

2.1.2 Example on BECs

Consider BECs with ϵ_1^N , where ϵ_i can differ and $N = 2^{10}$, and we assume that they are chosen in the range of $[0.4, 0.5]$. As is known, the BEC has a convenient property such that I_N can be described in a recursive form as follows:

$$\begin{aligned} I_{2N}^{(2i-1)} &= 1 - Z_{2N}^{2i-1} \\ I_{2N}^{(2i)} &= 1 - Z_{2N}^{2i}, \end{aligned} \quad (19)$$

where

$$\begin{aligned} Z_{2N}^{(2i-1)} &= f(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})) \\ Z_{2N}^{(2i)} &= g(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})). \end{aligned} \quad (20)$$

Here, $f(\alpha, \beta) = \alpha + \beta - \alpha\beta$, $g(\alpha, \beta) = \alpha\beta$, and $W_{[1:N]} = \{W_{(j)}\}_{j=1}^N$.

Figure 4 depicts the polarization phenomenon in $I_N^{(i)}$. Observe that even under the non-identically distributed channel condition, symmetric capacities are polarized into $\{0, 1\}$. It should be verified that it approaches the channel capacity as it is in the i.i.d. scenario. In Table 1, the differences between the symmetric capacity and the rate according to the exponent of n are shown. Here, the difference is defined as $\frac{I_s - NR}{I_s}$. As can be seen, the size of the differences decreases as n increases, which means that the rate R approaches to the average symmetric capacity $\frac{I_s}{N}$. Equivalently, it can be said that I_s is achievable in terms of the sum rate.

2.2 Achievable scheme based on the symmetric capacity

2.2.1 Encoder

Given a set of B-DMCs $\{W_N^{(i)}\}$, the encoder calculates $\{I_N^{(i)}\}$ according to its definition. Only for BECs, $I_N^{(i)}$ calculation follows (19) with equality. In general B-DMCs, the error performance of bit channels should be tracked

with appropriate measures such as $Z_N^{(i)}$ by using the density evolution method [9]. Then, define an information index set $A = \{i | I_N^{(i)} \geq I_s^{(j)}, i \in A, j \notin A\}$ for all $i, j \in [1, N]$ such that $|A| = \lfloor I_s \rfloor$. We apply the interleaver Q to map the indices of the codeword to the ordered set of transmit channels: $Q : \mathcal{X}^N \mapsto \mathcal{W}^N$, where $x_1^N \in \mathcal{X}^N$, and $W_{[1:N]} \in \mathcal{W}^N$. This is deployed to help the polarization processes. The problem with finding the proper Q for given channels is discussed at the end of this section. G_N is a $N \times N$ generator matrix, and its kernel is $G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. The encoder outputs a codeword $x_1^N = u_1^N \cdot G_N \cdot Q$.

2.2.2 Decoder

First, the receiver de-maps an output sequence by a de-interleaver Q^{-1} , which produces y_1^N . Next, y_1^N is processed by a successive cancellation (SC) decoder in order to obtain an estimated message sequence \hat{u}_1^N . One can easily check that the same recursive relations exist under the non-identical condition as those under the i.i.d channel condition, by considering the evolutions of bit channels. The only change is the last recursion of a likelihood ratio (LR) equation: $L(y_i)$ changes to $L_i(y_i) = \frac{W_{(i)}(y_i|0)}{W_{(i)}(y_i|1)}$.

$$L_N^{(i)}(y_1^N, u_1^{i-1}) = \frac{W_N^{(i)}(y_1^N, u_1^{i-1}|0)}{W_N^{(i)}(y_1^N, u_1^{i-1}|1)}. \quad (21)$$

Thus, the decoding complexity is still maintained as $O(N \log N)$, and it has vanishing probability of error rate P_e as $N \rightarrow \infty$. The encoding and decoding process is summarized in Algorithm 1.

Algorithm 1 Encoding and decoding process

Encoding Process

- 1: Given $\{W_{(i)} | 1 \leq i \leq N\}$, calculate $\{I_N^{(i)} | 1 \leq i \leq N\}$
- 2: Define A of size $k = I_s$ s.t. $A = \{i | I_N^{(i)} \geq I_s^{(j)}, \forall i \in A, \forall j \notin A\}$
- 3: Fix u_F and insert data into $\{u_A | i \in A\}$
- 4: Encode $x_1^N = u_1^N \cdot G_N$
- 5: Interleaver $Q : x_1^N \mapsto s_1^N$
- 6: Transmit s_1^N

Decoding Process:

- 1: De-interleaver : $Q^{-1} : t_1^N \mapsto y_1^N$
 - 2: Given (A, u_F) , for $\forall i \in [1, N]$, calculate $L(y_1^N, u_1^{i-1})$ for u_i in the SC decoder based on modified LR in (21).
-

2.3 Evolution of Bhattacharyya parameters

We can now show the similar relationships of those in Proposition 1 on the Bhattacharyya functional $Z(\cdot)$.

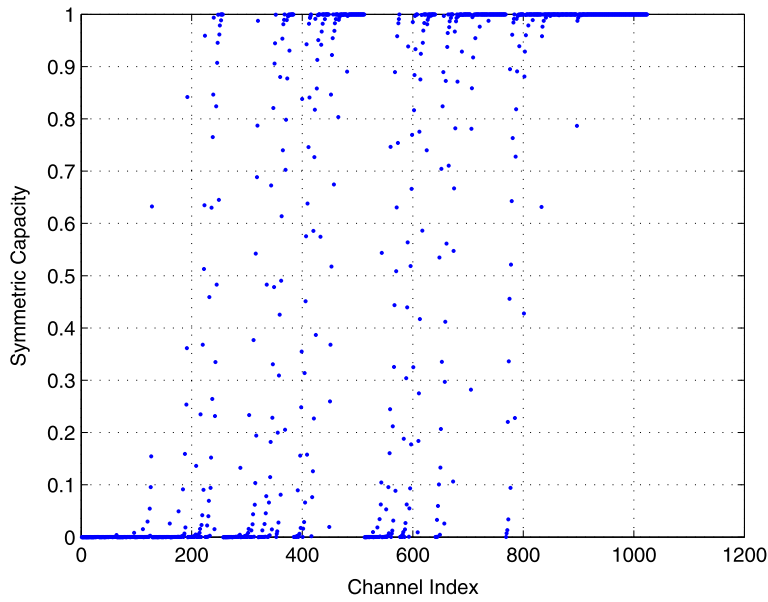


Fig. 4 Plot of $I(W_N^{(i)})$ for a non-identical BEC(ϵ_1^N), $N = 2^{10}$

Proposition 2 Suppose $(W_{(1)}, W_{(2)}) \mapsto (W_2^{(1)}, W_2^{(2)})$ for some binary input non-identically distributed discrete channels. Then, the following relations hold

$$Z(W_2^{(1)}) \leq Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)}) \quad (22)$$

$$Z(W_2^{(2)}) = Z(W_{(1)})Z(W_{(2)}) \quad (23)$$

$$Z(W_2^{(2)}) \leq Z(W_2^{(1)}) \quad (24)$$

The proofs of (22)-(24) is presented in the Appendix.

Corollary 1 Z_N can be described in a recursive form as follows:

$$\begin{aligned} Z_{2N}^{(2i-1)} &\leq f(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})) \\ Z_{2N}^{(2i)} &= g(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})), \end{aligned} \quad (25)$$

where $f(\alpha, \beta) = \alpha + \beta - \alpha\beta$, $g(\alpha, \beta) = \alpha\beta$, and $W_{[1:N]} = \{W_{(j)}\}_{j=1}^N$. The equality holds for BEC.

By applying the recursive channel structure of Lemma 1 to the definition of the Bhattacharyya parameter, the above corollary can be derived. With the aid of Lemma 1 and (22)-(24), we can derive Properties 1, 2, and 3 as follows:

1.

$$Z_{2N}^{(2i-1)} + Z_{2N}^{(2i)} \leq Z_N^{(i)}(W_{[1:N]}) + Z_N^{(i)}(W_{[N+1:2N]}). \quad (26)$$

2. (for some B-DMC W)

$$Z_{2N}^{(2i-1)} \geq \max \left\{ Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]}) \right\}. \quad (27)$$

To prove the above propositions, let us start from the first property. Property 1 can be derived from the definition of the Z -parameter and we can use the same proof process of (24) that exploits the inequality relation between the arithmetic mean and the geometric mean. By using the $f(\alpha, \beta)$ and $g(\alpha, \beta)$ notations as in Corollary 1, Property 2 holds if and only if the joint event of $\{\alpha\beta \leq \alpha\} \cap \{\alpha\beta \leq \beta\}$ holds, which is true because of the non-negativity of $\alpha, \beta \in [0, 1]$. Property 3 is true for BEC, BSC, and BI-AWGN which can be transformed to an equivalent BSC model.

2.3.1 BEC case

As previously mentioned, BEC has received attention for its special property of maximizing the Z -parameter value of the $f(\cdot)$ operation. For the BEC $W_{(i)}$ with erasure probability ϵ_i , $Z(W_{(i)}) = \epsilon_i$. For $N = 2$ with ϵ_1 and ϵ_2 , $Z_2^{(1)} = \epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2$, and it is larger than ϵ_1 and ϵ_2 . Therefore,

Table 1 Difference between capacities and achievable rates of polar codes under non-identical parallel channels

Number	8	10	12	14	16
Diff. [%]	0.6	0.044	0.038	0.003	0.002

without loss of generality, the following relation holds by repeating recursions:

$$Z_{2N}^{(2i-1)} = Z_N^{(i)}(W_{[1:N]}) + Z_N^{(i)}(W_{[N+1:2N]}) - Z_N^{(i)}(W_{[1:N]}) \cdot Z_N^{(i)}(W_{[N+1:2N]}).$$

Hence, $Z_{2N}^{(2i-1)} \geq Z_N^{(i)}(W_{[1:N]})$ and $Z_{2N}^{(2i-1)} \geq Z_N^{(i)}(W_{[N+1:2N]})$; thus, Property 3 holds.

2.3.2 BSC case

For $N = 2$ BSC with crossover probability ϵ_1 and ϵ_2 , $Z(W_{(i)}) = 2\sqrt{\epsilon_i(1-\epsilon_i)}$. Then, $Z_2^{(1)} = 2\sqrt{(1-\epsilon_s) \cdot \epsilon_s}$, where $\epsilon_s = \epsilon_1 + \epsilon_2 - 2\epsilon_1\epsilon_2$. Without loss of generality, assume that $\epsilon_1 \geq \epsilon_2$, if not, we swap these two channels. The numerical results in Fig. 5 show that differences of $Z_2^{(1)} - Z(W_{(1)})$ for all CPs are always non-negative.

Hence, we can conclude that Property 3 holds for the BSC case.

2.3.3 BI-AWGN case

Note that a binary-input AWGN channel can be transformed to the equivalent BSC by applying a line coding that maps the set of binary alphabets $\{0, 1\} \mapsto \{-1, +1\}$. Therefore, Property 3 holds.

2.4 Supermartingale Z_n

Let us define the random sequence $\{Z_n\}$ such that $Z_N^{(i)} \mapsto Z_n(b_1^n)$, where $Z_n(b_1^n) \in \{Z_n\}$ is represented as $Z_{b_1^n}$. Then, for some B-DMCs that satisfy four properties, we can draw a graphical evolving recursive structure of $Z_N^{(i)}$ (or Z_n) similar to Fig. 3. The graphical structure of Z_n has a common feature to that of the identically distributed channel case; however,

elements in the rightmost column are under a non-identically distributed scenario, where CPs may not be the same.

$\{Z_n\}$ can be considered as a supermartingale if this random tree process satisfies the relation $Z_n \geq \mathbb{E}[Z_{n+1}|b_1^n]$. Under the i.i.d. channel assumption, it is easily verified, since the one step transition from $Z_n(b_1^n)$ to $\{Z_{n+1}\}$ is a single variable to single variable mapping: $\mathcal{K} : \mathcal{Z} \rightarrow \mathcal{Z}$ such that $\mathcal{Z} = \{Z|Z \in \mathbb{R}, Z \in [0, 1]\}$. In contrast, with the non-identically distributed assumption, the transition becomes a two to one mapping: $\mathcal{K}' : \mathcal{Z}^2 \rightarrow \mathcal{Z}$. Each $Z_{b_1^n}$ is not a scalar but a 2×1 vector. In this case, the format of the condition is not appropriate due to dimension mismatch.

Using a similar process as that for I_n in proving the martingale property, we apply the average sequence $\{E_n\}$ to replace $Z_{b_1^n}$ such that $E_n(b_1^n) = \mathbb{E}[Z_{b_1^n}]$. Then, from (25), we can verify that $\mathbb{E}[Z_{n+1}|E_n(b_1^n)] \leq E_n$ which means Z_n is a supermartingale under E_n .

2.5 Convergence of $\{Z_n\}$

Since $\{Z_n\}$ is bounded within $[0, 1]$ and the supermartingale, from the martingale convergence theorem, there exists a random variable Z_∞ where the sequence $\{Z_n\}$ converges with probability 1 as $n \rightarrow \infty$. It is equivalent to the statement that in L^1 , $\lim_{n \rightarrow \infty} \mathbb{E}[|Z_{n+1} - E_n|] = 0$ which implies $\mathbb{E}[|Z_\infty - E_\infty|] = 0$. Given that $E_n(b_1^n) = \frac{\alpha + \beta}{2}$ and

$$Z_{n+1}(b_1^n) = \begin{cases} f(\alpha, \beta) & \text{w.p. } \frac{1}{2} \\ g(\alpha, \beta) & \text{w.p. } \frac{1}{2} \end{cases}, \quad (28)$$

where $\alpha \triangleq Z_N^{(i)}(W_{1:N})$ and $\beta \triangleq Z_N^{(i)}(W_{N+1:2N})$, it becomes an indeterminate equation with the condition of $\alpha, \beta \in$

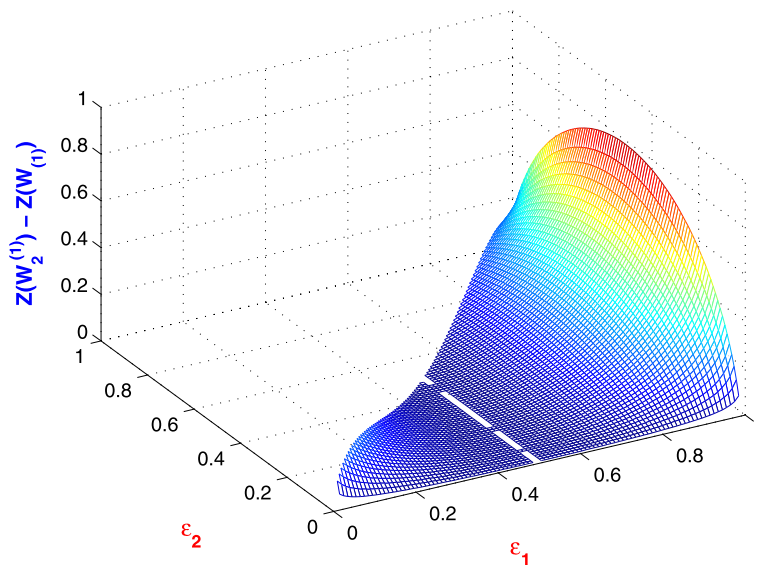


Fig. 5 General differences in Z-parameters for BSC ($N = 2$) for all $\epsilon_1 \in [0, 1]$ and $\epsilon_2 \in [0, 1 - \epsilon_1]$. Domains come from the constraints $\epsilon_1 \geq \epsilon_2$

$[0, 1]$. Solution pairs $(\alpha_\infty, \beta_\infty) = \{(0, 0), (1, 1)\}$ can be obtained. Therefore, since Z_∞ , which corresponds to α_∞ , converges to either 0 or 1 almost everywhere, we can conclude that all $Z_N^{(i)}$ are polarized in either a near perfect or totally random manner as $n \rightarrow \infty$.

2.6 Channel mapping via the interleaver Q

In this subsection, we discuss the role of an interleaver Q in polar coding systems under non-identically distributed B-DMCs and propose an algorithm that explains how to construct such an operation. The transceiver structure including an interleaver Q and a de-interleaver Q^{-1} is depicted in Fig. 1. To understand the importance of Q , let us consider the following example.

Example 1 For $N = 4$, let the set of erasure probabilities of parallel BECs be $\{\epsilon_1^N\} = \{0.1, 0.4, 0.6, 0.9\}$, of which the average is $\epsilon_m = 0.5$. Then, the evolved bit channel capacities are $\{I(W_N^{(i)})\} = \{0.02, 0.56, 0.44, 0.98\}$ and the encoder selects the information set $A = \{2, 4\}$. The resulting sum of the capacities of the selected bit channels is 1.54 [bits] for four channel uses.

We now apply the interleaver Q over the same set, which results in $\{\epsilon_1^N\} = \{0.6, 0.4, 0.9, 0.1\}$. The evolved set of symmetric capacities is $\{I(W_N^{(i)})\} = \{0.02, 0.31, 0.69, 0.97\}$.

The encoder chooses two indices with the highest $I(W_N^{(i)})$ values: $A = \{3, 4\}$, and the achievable sum is 1.66 [bits], which is a 7.8% enhancement compared to the previous index.

Note that as $\sum_{i \in A} I(W_N^{(i)})$ increases, the system becomes more reliable. The result of this example indicates that for a given set of channels $\{W_{(i)} | i \in [1, N]\}$, the optimal mapping $Q : x_1^N \mapsto s_1^N$ exists in the sense that the reliability (or the rate) is maximized by boosting polarizations among the symmetric capacities of bit channels. Finding the appropriate Q can be viewed as an optimization problem:

$$\begin{aligned} \text{Find } Q : x_1^N &\mapsto s_1^N \\ \text{Maximize } &\sum_{i \in A} I(W_{(i)}). \end{aligned}$$

Equivalently, the mapping can be interpreted as a channel permutation such that $Q : \{W_{(i)}\} \mapsto \{W'_{(i)}\}$. We discuss two methods that search for such a mapping Q .

2.6.1 Exhaustive search method with grouping

The simplest and most naive approach involves testing every possible combination over N channels and selecting the best combination. Obviously, $N!$ number of cases need to be checked. However, we can categorize every combination into equivalent groups since in each group,

all combinations output the same qualities of bit channels. The size of each group is 2^{N-1} . Hence, owing to the recursive channel evolving structure, the required number of tests is $\frac{N!}{2^{N-1}}$. The detailed proof is shown in the Appendix. This grouping technique considerably reduces the computational burdens: for $N = 8$, we need to test 315 representative combinations instead of $N! = 40,320$. However, the enhanced test set size would go beyond the computational capability for practical N lengths.

2.6.2 Heuristic method

The purpose of the channel combining and the splitting operation is to build virtual channels that are as close to the extremal channels as possible as the recursions are repeated. We already know that for any non-identical parallel B-DMCs, there exist polar codes that achieve the symmetric capacity as N goes to infinity. However, in a practical system that exploits finite code lengths, as we observed via Example 1, differences exist in the convergence speed for different mappings through the interleaver Q . A well-designed Q will polarize bit channel qualities fast. Therefore, the Q should sort the given set of transmit channels in order to create as many enhanced and degraded transmit channels as possible that are close to the extremals simultaneously. In this section, we propose an algorithm of Q that achieves such an object.

1. Sort transmit channels $W_{(i)}, i \in [1, N]$ in an ascending order of the capacity $I(W_{(i)})$.
2. Make $\frac{N}{2}$ pairs: the i^{th} pair includes the i^{th} smallest transmit channel and the i^{th} largest transmit channel.
3. Using the indices of $\frac{N}{2}$ pairs, $[1 : \frac{N}{2}]$, repeat the second procedure until the size of the index set becomes 4.

As can be seen, this algorithm has a recursive structure, and we can represent it in a matrix form. Let P_n represent the interleaver operation Q for length $N = 2^n$ and S_n indicate the second operation of the above algorithm. Then,

$$P_n = S_n \cdot P_{n-1}^{\otimes \mathcal{I}_2}, \quad (29)$$

where \mathcal{I}_2 is the 2×2 identity matrix and \otimes means the Kronecker product.

In Fig. 6, we depict achievable rates $\frac{1}{N} \sum_{i \in A} I(W_N^{(i)})$ of polar codes by exploiting different interleaving methods under parallel BECs of which the erasure probabilities are uniformly chosen within the range of $(0, 1)$ (hence, the average of 0.5). The black curve refers to the capacity, the red curve refers to the result of the proposed algorithm, and the blue curve refers the performance when using random shuffling. We can obtain the green curve when we apply the proposed algorithm only once without recursion: that is, $P_n = S_n$. Finally, the dark red curve refers to the performance when the transmit channels are sorted in

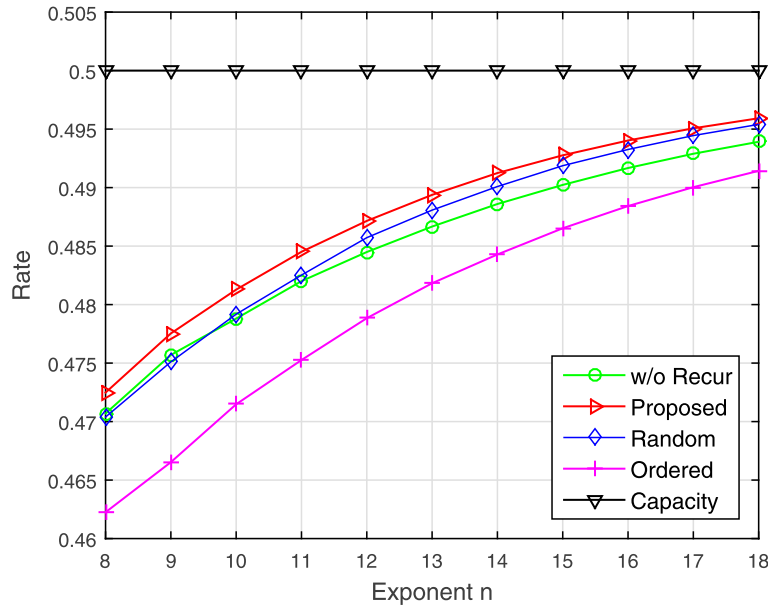


Fig. 6 Observation on the different usage of the interleaver Q . Shuffling the coded alphabets x_1^N as randomly as possible exhibits higher achievable rates compared to the ordered interleaver

capacities; whether the order is ascending or descending, the result is not affected.

From the figure, the proposed interleaving algorithm shows better performance in the rate than the other algorithms. It is also observed that when $W_{(i)}$ s forms an ordered set, it converges to the capacity slower than the other algorithms. Hence, if the symmetric capacities of the exposed channels are ordered in either way, they should be rearranged through Q .

3 Non-identical binary erasure channels with random erasure probabilities with single distribution

In previous studies, [8] and [3], it is assumed that the characteristics of underlying discrete memoryless channels are fully exposed to the transceiver; thus, the encoder and the decoder exploit this information. Under this condition, it is proved that polar codes can achieve the symmetric capacity.

In this section, we assume that channel parameters are not deterministic, but are realizations of a random variable. For BEC, the channel transition probability of a transmit channel $W_{(i)} : \mathcal{X}_i \mapsto \mathcal{Y}_i$ is fully described by the erasure probability ϵ_i . Hence, the channel features of the non-identically distributed parallel channels model are perfectly represented via the set of erasure probabilities $\{\epsilon_1^N\}$.

The existence of random erasure probabilities means that each ϵ_i is the realization of the random variable θ

such that $\epsilon_i \sim f_\theta(\epsilon_i), \forall i \in [1, N]$, where f_θ is a stationary probability distribution function.

We now assume that the realized set of erasure probability $\{\epsilon_1^N\}$ is exposed to neither the encoder nor the decoder. In this case, the only available information that can be extracted to the encoder and the decoder is the set of moments from the given distribution f_θ . We prove that under this scenario, polar codes can achieve the symmetric capacity.

Theorem 2 For a set of BECs $\{W_{(i)}\}$ with the unknown set of erasure probabilities $\{\epsilon_i\}, i \in [1, N]$, but the distribution f_θ is given to the transceiver, there exist polar codes, for arbitrary small $\delta \leq 0$, that achieve the symmetric capacity I_s , in the sense that as $N \rightarrow \infty$, the fraction of indices i satisfies

$$\lim_{N \rightarrow \infty} \frac{|\{i | I(W_N^{(i)}) \in (1 - \delta, 1]\}|}{N} = I_s$$

$$\lim_{N \rightarrow \infty} \frac{|\{i | I(W_N^{(i)}) \in [0, \delta)\}|}{N} = 1 - I_s,$$

where the symmetric capacity I_s is defined as an average of the individual transmit channel's capacities: $I_s(W_N) = \frac{1}{N} \sum_{i=1}^N I(W_{(i)})$, where $W_N : X_1^N \mapsto Y_1^N$.

3.1 Proof of Theorem 2

By the law of large numbers, the empirical channel behavior for a codeword can be described by the first moment

ϵ_m of the distribution f_θ . Also, as previously mentioned, since the transceiver is oblivious to the exact set of erasure probabilities of the underlying parallel channels, it has no choice but to exploit ϵ_m for consideration of constructing codewords.

Now, let us consider the meaning of I_s

$$I_s(W_N) = \frac{1}{N} \sum_{i=1}^N I(W_{(i)}) \quad (30)$$

$$= \frac{1}{N} \sum_{i=1}^N (1 - \epsilon_i) \quad (31)$$

$$= 1 - \frac{1}{N} \sum_{i=1}^N \epsilon_i \quad (32)$$

$$= (1 - \epsilon'), \quad (33)$$

where $\epsilon' = \frac{1}{N} \sum_{i=1}^N \epsilon_i$. Hence, we can conclude that the non-identically distributed N parallel BECs model is equivalent to the N identically distributed BECs model of which the erasure probability is ϵ' for a large N and we can achieve the symmetric capacity. Next, we need to verify whether (33) is reliably achievable based on information given by the transceiver. To that end, we need to demonstrate that the following two issues are true:

- The existence of polar codes over BEC with erasure probability ϵ_m
- $\epsilon' \geq \epsilon_m$ for all distributions f_θ

The first statement is proved to be true in Theorem 2 of [1] in that for any B-DMC W , there exists a sequence of information bit sets A_N such that $|A_N| \geq NR$ with arbitrary small error probability. In this scenario, $|A_N| = \lfloor N(1 - \epsilon') \rfloor$. The second statement is proved through the linearity of I_s to the erasure probability. In Fig. 7, I_s

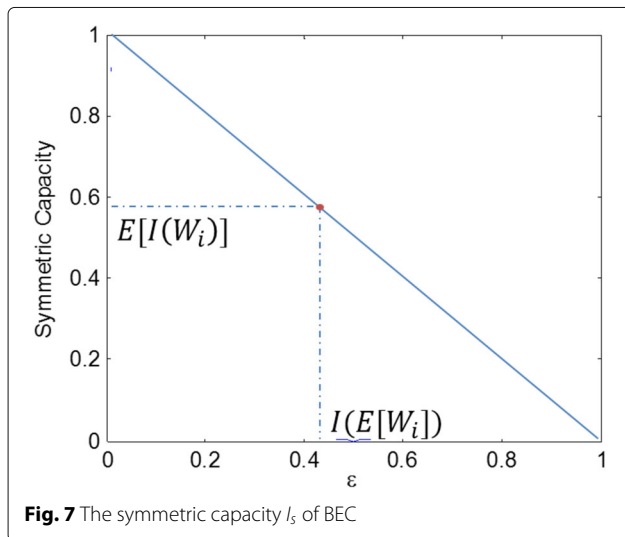


Fig. 7 The symmetric capacity I_s of BEC

is depicted. Note that the I_s is an affine function of the erasure probability ϵ . Therefore, $\epsilon' = \epsilon_m$ for any distribution f_θ since $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \epsilon_i = \epsilon_m$. Note that under erasure channels, the equality in the Jensen's inequality holds

$$I(E[W_{(i)}]) \leq E[I(W_{(i)})]. \quad (34)$$

This completes the proof of Theorem 2.

3.2 Achievable polar coding scheme

According to Theorem 2, the construction of the capacity achieving polar coding scheme is straightforward. First, given the distribution f_θ , the encoder calculates the first moment $\mathbb{E}[\epsilon] = \epsilon_m$. It then constructs the message vector u_1^N by determining the information index set A_N with the pre-defined frozen bits u_F . This message sequence is encoded through the generator matrix G_N and is transmitted through the non-identically distributed parallel BECs of $\{\epsilon_i^N\}$. The procedure is summarized in Algorithm 2.

Algorithm 2 Polar coding scheme

Encoding process

- 1: Given f_θ , calculate ϵ_m
- 2: Figure out the information sets A_N according to BEC(ϵ_m)
- 3: Encoding: $x_1^N = u_1^N \cdot G_N$

Decoding process

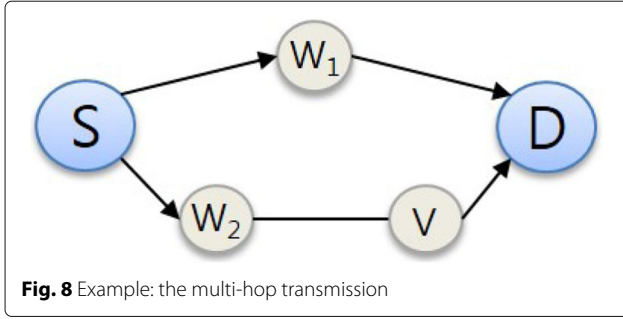
- 1: Given (A, u_F, ϵ_m) Perform SC decoding: $y_1^N \rightarrow \hat{u}_A$.
-

It should be noted that to achieve the symmetric capacity under non-identically distributed parallel BECs, with unknown channel parameters, the only constraint required is the code length $N \rightarrow \infty$.

4 Random erasure probabilities with non-identical distributions

In this subsection, we consider the case of N non-identically distributed BECs $W_{(i)}$: for $\forall i \in [1, N]$ that each distribution f_{θ_i} of the erasure probability ϵ_i for each transmit channel $W_{(i)}$ could differ for different bit channel indices. Thus, this scenario includes previous scenarios as a special case.

For example, consider the multi-hop communications between the two nodes in Fig. 8. In this figure, each data element in S is delivered to D through different paths in a multi-hop fashion. If the number of hops is increased, it becomes more difficult for S and D to track the exact parameters that model each path. Furthermore, these paths could have statistical variation due to unstable



media or interference from different noise sources such as the node V in the figure. In these cases, the proposed scenario is logical.

A simple coding scheme that is able to achieve the symmetric capacity is the conveyance of multiple codewords as a group for each decoding stage. In each stage, we exploit the set of parallel non-identical channels $L = 2^l$ times. The reason for this power of 2 format is to match with the length of polar codewords.

Let $W_{(i)}(j)$ mean the j th access to the channel $W_{(i)}$, $\epsilon_i(j)$ as the instantaneous erasure probability of the channel $W_{(i)}(j)$ that follows the distribution f_{θ_i} , and $I_{s,i}(W_{(i)})$ is the symmetric capacity of $W_{(i)}$ by accessing it L times.

The instantaneous capacity for L blocks are depicted in Table 2. The distributions of erasure probabilities can differ, and each transmit channel (row) shows ergodic behavior as shown in the last column of the table. Here, $\bar{I}(W_{(N)}) = \frac{1}{L} \sum_{j=1}^L I(W_{(i)}(j))$.

The symmetric capacity in this case becomes

$$I_s(W_{[N]}^L) = \frac{1}{L} \sum_{j=1}^L \frac{1}{N} \sum_{i=1}^N I(W_{(i)}(j)) \quad (35)$$

$$= \frac{1}{L} \sum_{j=1}^L \frac{1}{N} \sum_{i=1}^N (1 - \epsilon_i(j)) \quad (36)$$

$$= \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{1}{L} \sum_{j=1}^L \epsilon_i(j) \right) \quad (37)$$

$$\rightarrow \frac{1}{N} \sum_{i=1}^N (1 - \epsilon_{m_i}), \quad (38)$$

Table 2 Ergodic behaviors of instantaneous capacities

	1	2	...	L	Average
1	$I(W_{(1)}(1))$	$I(W_{(1)}(2))$...	$I(W_{(1)}(L))$	$\bar{I}(W_{(1)})$
2	$I(W_{(2)}(1))$	$I(W_{(2)}(2))$...	$I(W_{(2)}(L))$	$\bar{I}(W_{(2)})$
i	$I(W_{(i)}(1))$	$I(W_{(i)}(2))$...	$I(W_{(i)}(L))$	$\bar{I}(W_{(i)})$
N	$I(W_{(N)}(1))$	$I(W_{(N)}(2))$...	$I(W_{(N)}(L))$	$\bar{I}(W_{(N)})$

where ϵ_{m_i} is the first moment of $\epsilon_i \sim f_{\theta_i}$. Note that $\lim_{L \rightarrow \infty} I_s(W_{[N]}^L) \frac{1}{N} = \sum_{i=1}^N (1 - \epsilon_{m_i})$. The equality is due to the affinity of the symmetric capacity over the domain of erasure probability. We now consider two cases. In the first case, we assume that the encoder is able to be adapted to various code lengths. This means that the encoder can construct generator matrices G_N for any exponent n ($N = 2^n$). In the second case, the coding structure is fixed, thus parameter N (and the following G_N) cannot be changed. In the first case, let the encoder exploit any exponent l , and construct $L \times L$ generator matrix G_L where $L = 2^l$. The following proposition is then satisfied:

Proposition 3 For a set of non-identically distributed BECs $\{W_{(i)}\}$, with a set of random erasure parameter $\{\epsilon_i\}$, $i \in [1, N]$ that each ϵ_i follows non-identical f_{θ_i} , the symmetric capacity $I_s(W_{[N]}^L)$ is achievable by exploiting multiple streams of polar codewords.

The proof of Proposition 3 is the same as that of the existence of polar codes that achieve the set of individual capacities $\{\bar{I}(W_{(i)})\}$, since the symmetric capacity $I_s(W_{[N]}^L)$ is their sum. Also, in Theorem 2, we proved that there exist polar codes that achieve each capacity $\{\bar{I}(W_{(i)})\}$, in the sense that as $L \rightarrow \infty$ through the power of 2, the fraction of indices $j \in [1, L]$ of the i th message block is satisfied:

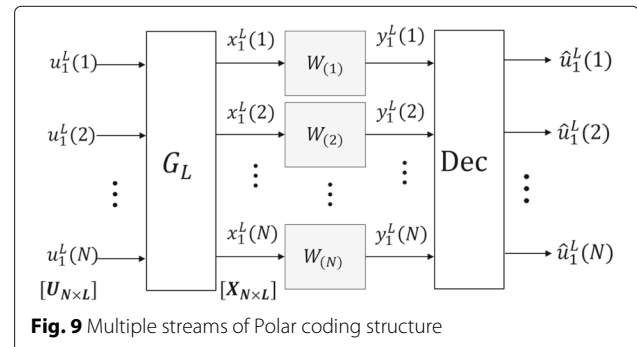
$$\lim_{L \rightarrow \infty} \frac{|\{j | I(W_L^{(j)}) \in (1 - \delta, 1]\}|}{L} = \bar{I}(W_{(i)})$$

$$\lim_{L \rightarrow \infty} \frac{|\{j | I(W_L^{(j)}) \in [0, \delta)\}|}{L} = 1 - \bar{I}(W_{(i)})$$

for an arbitrary small $\delta \geq 0$, and for all channel index $i \in [1, N]$. Therefore, Proposition 3 is true for any set of distributions $\{f_{\theta_i}\}$.

4.0.1 Achievable coding scheme

In Fig. 9, the encoding and the decoding procedures are depicted for all transmit channel indexes. Given



all distributions $\{f_{\theta_i} | \forall i \in [1, N]\}$, the encoder calculates the set of first moments $\{\epsilon_{m_i}\}$ and evolved bit channel capacities $\left\{I\left(W_N^{(i)}\right)\right\}$ with $\{\epsilon_{m_i}\}$ according to these recursive equations, (19) and (20). With these moments and symmetric capacities, the encoder then constructs N streams polar codewords $\{x_1^L(i)\}$ of length $L = 2^l$. To achieve this, the encoder first defines an information set $A_L(i) = \left\{j | I\left(W_N^{(j)}\right) \geq I\left(W_N^{(k)}\right), \forall j \in A_L, k \in A_L^c\right\}$, the size of which is $|A_L(i)| = \lfloor L(1 - \epsilon_{m_i}) \rfloor$.

With fixed frozen bits u_F , where $F = [1, L] \setminus A_L(i)$, message vectors $\{u_1^L(i)\}$ are stacked in the $N \times L$ message matrix U , which is then encoded with the $L \times L$ generator matrix G_L and output codewords matrix X . Each column of X propagates sequentially through the non-identically distributed parallel BECs.

The receiver saves L output vectors in matrix Y and produces estimates by applying the SC decoder row by row. This procedure is summarized in Algorithm 3. Note that there are no constraints on N . Actually, since the proposed coding scheme is not affected by N , it is robust to the deletion of some transmit channels. Assume that a set of channels W_J where J is a subset of $[1, N]$ is lost since the corresponding set of symmetric capacities are all zeros. The encoder and the decoder will then simply decrease N to $N - |J|$; the data is transmitted through $W_{[1, N] \setminus J}$. Note that the symmetric capacity $I_s\left(W_{[N] \setminus J}^L\right) = \sum_{\forall i \in [1, N] \setminus J} I(W_{(i)})$ is still achievable.

Algorithm 3 Encoding and decoding process

Encoding process: Repeat $\forall i \in [1, N]$

- 1: Calculate $\bar{I}(W_{(i)})$
- 2: Find ϵ_{m_i} s.t. $I(\epsilon_{m_i}) = \bar{I}(W_{(i)})$
- 3: Define index set A_i based on (L, ϵ_{m_i})
- 4: $x_1^L(i) = u_1^L(i) \cdot G_L$
- 5: Store $x_1^L(i)$ in the i^{th} row of X
- 6: Transmit each column of X

Decoding Process:

- 1: Stack N blocks into matrix Y row by row.
 - 2: Given (A_i, L, u_F) , perform SC decoder: $Y(i) \mapsto \hat{u}_{A_i} \quad \forall i \in [1, N]$
-

4.0.2 Proof of achievability

From the set of BECs $\{W_{(i)}\}$ for $\forall i \in [1, N]$, the size of each information set $|A_L(i)| = \lfloor L(1 - \epsilon_{m_i}) \rfloor$, the unit of which is in bits per L channel uses. It is known that the SC decoder recovers each message u_{A_i} with vanishing probability of error as $L \rightarrow \infty$ for $\forall i$. Now, define an individual rate R_i as $\frac{|A_i|}{L}$; it then converges to $\bar{I}(W_{(i)})$ under the same condition. That is, for any

$\delta_i \in [0, 1)$

$$\begin{aligned} R_i &= \frac{1}{L} \lfloor L(1 - \epsilon_{m_i}) \rfloor \\ &= (1 - \epsilon_{m_i}) - \frac{\delta_i}{L} \\ &\xrightarrow{L \rightarrow \infty} \bar{I}(W_{(i)}). \end{aligned}$$

For the symmetric capacity, $I_s\left(W_{[N]}^L\right)$ is the same as the arithmetical mean of its parts $\{\bar{I}(W_{(i)}) | \forall i \in [1, N]\}$: $I_s\left(W_{[N]}^L\right) = \frac{1}{N} \sum_{i=1}^N \bar{I}(W_{(i)})$. We can conclude that it is achievable from the proposed scheme. The complexity of this polar coding scheme is $O(NL \log L)$, since it is a concatenation of N SC decoders of length L .

In addition, consider a transmitter in which the encoding structure cannot be changed. Then, the only choice is to utilize the fixed size of the $N \times N$ generator matrix G_N , where $N = 2^n$ and the produced codewords are of the same length, N . We can treat this problem by setting L to be identical to N . The encoder defines the collection of information index sets $\{A_i\}$ from $\{\epsilon_{m_i}\}$, where $|A_i| = \lfloor N \cdot I(\epsilon_{m_i}) \rfloor$. Using a common generator matrix G_N , the encoder sequentially produces N of polar codewords $x_1^N(i) = u_1^N(i) \cdot G_N$ for $i \in [1, N]$. These codewords $\{x_1^N(k)\}$ are stacked in rows of the matrix X . The complexity of this polar coding scheme is $O(N^2 \log N)$, since it is a concatenation of N SC decoders.

5 Polarizations on non-independent channels

In the previous sections, it was assumed that all transmit channels are independent from each other from

$$W_N(y_1^N | x_1^N) = \prod_{i=1}^N W_{(i)}(y_i | x_i), \quad (39)$$

where $W_{(i)} : \mathcal{X}_i \rightarrow \mathcal{Y}_i$. Now, consider the case where these channels are correlated; thus, the equality in (39) does not hold. Then, the previous polar coding scheme for independent B-DMCs might not achieve the capacity. For example, we can check the polarization phenomenon in a measure of the symmetric capacity for $N = 2$, where the transition probability of the channel $W_2 : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{Y}_1 \times \mathcal{Y}_2$, where \mathcal{X}_1 and \mathcal{X}_2 are $GF(2)$, and $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, e\}$ is defined as follows:

$$W_2(y_1^2 | x_1^2) = \begin{cases} 1 - \epsilon, & y_1^2 = x_1^2 \\ \epsilon & y_1^2 = e \end{cases}. \quad (40)$$

The element e denotes the erasure symbol. Without losing information, W_2 can be modeled as a single quaternary erasure channel (QEC) W' with CP ϵ , of which the capacity is $I(W') = 2(1 - \epsilon)$. We now apply the same generator matrix G_2 to the encoder and calculate the symmetric

capacities of split bit channels $I(W_2^{(1)})$ and $I(W_2^{(2)})$ via (1). We can then obtain $I(W_2^{(1)}) = I(W_2^{(2)}) = 1 - \epsilon$, which means that no polarization occurs with the coding strategy for independent channels. By relabeling, the set of binary input vectors $X_1^2 = \{00, 01, 10, 11\}$ is mapped to a quaternary symbol set $S = \{0, 1, 2, 3\}$, as depicted in Fig. 10, where $r = 2$ with the channel notation change from the correlated vector channel W_2 to the single independent q -ary DMC W' . The problem of non-polarization is proved to occur in some q -ary DMCs when the cardinality of the set S is a composite number [10]. However, in [11], the authors proved the existence of polar codes for the composite cardinality of $q = 2^r$.

Under the parallel channel model, when $N = 2^n$, the transmit channels are pairwise correlated from

$$W_N(y_1^N | x_1^N) = \prod_{i=1}^{N/2} W_{2(i)}(y_{2i-1}^{2i} | x_{2i-1}^{2i}). \quad (41)$$

The transmit channels can be represented as N' length of independent quaternary input DMCs

$$W_{N'}(t_1^{N'} | s_1^{N'}) = \prod_{i=1}^{N'} W'_{(i)}(t_i | s_i), \quad (42)$$

where $N' = \frac{N}{2}$.

For generalization, assume that r transmit channels are correlated in similarly to (41). If we assume $r = 2^\alpha$ ($1 \leq \alpha \ll n - 1$) for simplicity, then the new code length becomes $N' = 2^{n-\alpha}$, and we have

$$W_N(y_1^N | x_1^N) = \prod_{i=1}^{N'} W'_{r(i)}(y_{r(i-1)+1}^{ri} | x_{r(i-1)+1}^{ri}). \quad (43)$$

In a q -ary representation, (42) still holds, where the input alphabet cardinality is $q = 2^r$. If these relabeled independent q -ary DMCs are identically distributed, such that $W'_{(i)} = W'_{(j)}$ for $\forall i, j \in [1, N']$, it is proved in [11] that polar codes for q -ary input DMCs achieves the symmetric channel capacity when $q = 2^r$ by exploiting the same kernel as in [1]. Therefore, the following proposition holds for a general q -ary DMC W' .

Proposition 4 *There exist polar codes for non-independent DMCs W' , which achieve the symmetric $I(W')$, by relabeling r -bits binary sequence to a q -ary ($q = 2^r$) symbol as $N' \rightarrow \infty$ through the power of 2.*

Proof The proof follows the proof of the existence of polar codes in [11]. The authors proved the existence of polar codes that achieve the symmetric capacity for any q -ary input alphabet discrete memoryless channel when q is the power of 2. In our model, by grouping and relabeling alphabets in the same length, the model can be mapped to an equivalent q -ary alphabet model, without any information loss because these operations are deterministic. We can then exploit the same proof as that in [11]; hence, the achievability still holds in this partially non-independent channels model. \square

6 Conclusions

In Section 2, we proved that for deterministic CPs in non-identical channel models, polar codes can achieve the sample mean of bit channel capacities. In addition, we provided an example that demonstrates the importance of the use of the proper channel interleaver Q for achievable rate enhancement and proposed the heuristic mapping algorithm.

In Sections 3 and 4, the key contribution is a new system model where the transmitter and the receiver know only the channel parameter distribution and not the channel parameter itself. If the underlying channel type is BEC, the coding scheme can become simpler. Note that for a BEC with erasure probability ϵ , its symmetric capacity I is the affine function of ϵ . Then, we have the relation $E[I(\epsilon)] = I(\bar{\epsilon})$, where $\bar{\epsilon}$ is the expectation of the random variable $\epsilon \sim f_\epsilon(\epsilon)$.

By applying multiple streams of polar codewords, we prove that the average capacity of BECs under our scenarios is achievable. However, this is obtained by sacrificing the latency and complexity, since they stack multiple blocks during the encoding and decoding process. Hence, these schemes might not be suitable in systems where low latency or low complexity is required. Rather, it is more practical in storage systems such as flash memory devices where throughput is more important than latency. Especially, for flash memories, statistical responses such as a voltage threshold would change with time and with the number of accesses to a cell block. Hence, as the storage

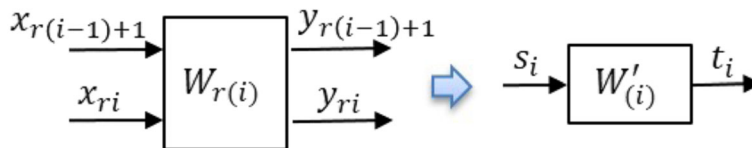


Fig. 10 Relabeling of a length $r = 2$ binary vector channel to a single quaternary channel

capacity increases, it is inefficient for a storage controller to determine the exact states of every block or cell. If statistics on their changes are given instead, we can manage cells more efficiently using the proposed polar coding scheme.

In addition, in the case of parallel channels, where statistically different random disturbances exist across channels, it is difficult to track all the channel parameters. However, if their statistics are known to the transmitter and the receiver, we can deliver data up to the average capacity through polar codes by sacrificing latency. In such cases, polar codes are a promising option that maximizes the throughput.

Under the non-independent channel scenario, we assume that the N transmit channels are grouped into channels with size r which is a power of 2, so that we can treat the scenario as a non-binary system. If N is not divisible by r ($N \bmod r \neq 0$), puncturing may be used to fit the system into a q -ary system [12]. The proposed polar codes appear to be promising for applications where only the knowledge of the channel parameter distribution is available and can be practical for storage applications such as flash memory devices.

Appendix

Proof of (22)

Proof

$$\begin{aligned}
 & Z(W_2^{(1)}) \\
 &= \sum_{y_1^2} \sqrt{W_2^{(1)}(y_1^2|u_1=0) W_2^{(1)}(y_1^2|u_1=1)} \\
 &= \sum_{y_1^2} \sqrt{\sum_{u_2} \frac{1}{2} W_{(1)}(y_1|u_2) W_{(2)}(y_2|u_2)} \\
 &\quad \cdot \sqrt{\sum_{u_2'} \frac{1}{2} W_{(1)}(y_1|1+u_2') W_{(2)}(y_2|u_2')} \\
 &= \sum_{y_1^2} \frac{1}{2} \sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|0) + W_{(1)}(y_1|1) W_{(2)}(y_2|1)} \\
 &\quad \cdot \sqrt{W_{(1)}(y_1|1) W_{(2)}(y_2|0) + W_{(1)}(y_1|0) W_{(2)}(y_2|1)} \\
 &\leq \sum_{y_1^2} \frac{1}{2} \left[\sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|0)} + \sqrt{W_{(1)}(y_1|1) W_{(2)}(y_2|1)} \right] \\
 &\quad \cdot \left[\sqrt{W_{(1)}(y_1|1) W_{(2)}(y_2|0)} + \sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|1)} \right] \\
 &\quad - \sum_{y_1^2} \sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|0) W_{(1)}(y_1|1) W_{(2)}(y_2|1)};
 \end{aligned}$$

after calculations, it becomes

$$\begin{aligned}
 &= \frac{1}{2} \sum_{y_1^2} \left(W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \right. \\
 &\quad + W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \\
 &\quad + W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \\
 &\quad \left. + W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \right) \\
 &\quad - \sum_{y_1^2} \sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|0) W_{(1)}(y_1|1) W_{(2)}(y_2|1)} \\
 &= Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)}).
 \end{aligned}$$

Therefore, $Z(W_2^{(1)}) \leq Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)})$ is satisfied for any binary input channel parameters. \square

Proof of (23)

Proof

$$\begin{aligned}
 & Z(W_2^{(2)}) \\
 &= \sum_{y_1^2, u_1} \sqrt{W_2^{(2)}(y_1^2, u_1|u_2=0) W_2^{(2)}(y_1^2, u_1|u_2=1)} \\
 &= \sum_{y_1^2, u_1} \frac{1}{2} \sqrt{W_{(1)}(y_1|u_1) W_{(2)}(y_2|0)} \\
 &\quad \cdot \sqrt{W_{(1)}(y_1|u_1+1) W_{(2)}(y_2|1)} \\
 &= \sum_{y_2} \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \\
 &\quad \cdot \sum_{y_1, u_1} \frac{1}{2} \sqrt{W_{(1)}(y_1|u_1) W_{(1)}(y_1|u_1+1)} \\
 &= Z(W_{(2)})Z(W_{(1)}).
 \end{aligned}$$

\square

The fourth equation comes from the fact that the summation over u_1 is a sum of two similar terms. In addition, the following relation holds with the aid of (23):

$$Z(W_2^{(2)}) \leq \min(Z(W_{(1)}), Z(W_{(2)})).$$

The relation can be verified simply by subtracting either the right hand terms from the left hand term.

Proof of (24)

Proof We can prove (24) simply by applying the arithmetic-geometric mean inequality to $Z(W_2^{(1)})$. Let us review the development process.

$$\begin{aligned}
& Z(W_2^{(1)}) \\
&= \sum_{y_1^2} \sqrt{W_2^{(1)}(y_1^2|u_1=0) W_2^{(1)}(y_1^2|u_1=1)} \\
&= \sum_{y_1^2} \sqrt{\sum_{u_2} \frac{1}{2} W_{(1)}(y_1|u_2) W_{(2)}(y_2|u_2)} \\
&\quad \cdot \sqrt{\sum_{u_2'} \frac{1}{2} W_{(1)}(y_1|1+u_2') W_{(2)}(y_2|u_2')} \\
&= \sum_{y_1^2} \frac{1}{2} \sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|0) + W_{(1)}(y_1|1) W_{(2)}(y_2|1)} \\
&\quad \cdot \sqrt{W_{(1)}(y_1|1) W_{(2)}(y_2|0) + W_{(1)}(y_1|0) W_{(2)}(y_2|1)}.
\end{aligned}$$

Defining shorthand notations of $A = W_{(1)}(y_1|0)$, $B = W_{(1)}(y_1|1)$, $C = W_{(2)}(y_2|0)$, and $D = W_{(2)}(y_2|1)$, we can rewrite the above equation as follows:

$$\begin{aligned}
Z(W_2^{(1)}) &= \sum_{y_1^2} \frac{1}{2} \sqrt{ABC^2 + CDA^2 + CDB^2 + ABD^2} \\
&\geq \sum_{y_1^2} \frac{1}{2} \sqrt{4 \cdot ABCD} \\
&= \sum_{y_1} \sqrt{AB} \sum_{y_2} \sqrt{CD} \\
&= \sum_{y_1} \sqrt{W_{(1)}(y_1|0) W_{(1)}(y_1|1)} \\
&\quad \cdot \sum_{y_2} \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \\
&= Z(W_{(1)}) Z(W_{(2)}) \\
&= Z(W_2^{(2)})
\end{aligned}$$

which the inequality is from the arithmetic and geometric mean relation. \square

Proof of the number of equivalent channel combinations

According to the basic theory of polar codes, channels are recursively evolved such that $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ for some set of binary input discrete memoryless channels. We denote each mapping as follows:

$$\mathcal{F}_{n+1} : W_N^{(i)}, W_N^{(i)} \mapsto W_{2N}^{(2i-1)} \quad (44)$$

$$\mathcal{G}_{n+1} : W_N^{(i)}, W_N^{(i)} \mapsto W_{2N}^{(2i)}, \quad (45)$$

and it was proved that polarizations occur in non-identically distributed channels.

Recall that functional \mathcal{F}_n and \mathcal{G}_n share the same set of functions as an input. Furthermore, these input functions

are also results of functionals \mathcal{F}_{n-1} or \mathcal{G}_{n-1} due to the recursive structure of channel evolution. Let us define a swapping operation $S : (\alpha, \beta) \mapsto (\beta, \alpha)$. We can then easily check that \mathcal{F}_n and \mathcal{G}_n are invariant to S . We also define a functional \mathcal{H}_n which includes both \mathcal{F}_n and \mathcal{G}_n . \mathcal{H}_n can represent its members since \mathcal{F}_n and \mathcal{G}_n are equivalent in the number of cases; hence, it is also invariant to S . In a similar way, we define \mathcal{W}_{n+1} which includes $W_{2N}^{(2i-1)}$ and $W_{2N}^{(2i)}$. We now rewrite (44) and (45) using \mathcal{H} and \mathcal{W} , respectively.

$$\mathcal{W}_{n+1} = \mathcal{H}_{n+1}(\mathcal{W}_n, \mathcal{W}_n).$$

Due to the recursive structure in channel evolutions, each $W_N^{(i)}$ is calculated again through \mathcal{H}_n . It is then generally in the form of

$$\mathcal{W}_{n+1} = \mathcal{H}_{n+1}(\mathcal{H}_n(\mathcal{W}_{n-1}, \mathcal{W}_{n-1}), \mathcal{H}_n(\mathcal{W}_{n-1}, \mathcal{W}_{n-1})).$$

Defining the number of operations of \mathcal{H} that are required to recursively reach \mathcal{H}_n as χ_n then

$$\chi_{n+1} = 2\chi_n + 1.$$

By solving this recurrence formula, we can obtain $\chi_n = N - 1$. Recalling that \mathcal{H} is invariant to S which indicates there are two cases for each \mathcal{H} , the number of combinations of the B-DMCs that result in the same information set of polar codes is 2^{N-1} . Hence, the number of representative combinations that may have different information sets for length N parallel polar coding systems is $\frac{N!}{2^{N-1}}$.

Acknowledgements

This research was supported in part by the Basic Science Research Programs (NRF-2013R1A1A2008956 and NRF-2015R1A2A1A15052493) through NRF funded by MOE and MSIP, the ICT R&D Program of MSIP/IITP (B0717-16-0023), the Technology Innovation Program (10051928) funded by MOTIE, the Bio-Mimetic Robot Research Center funded by DAPA (UD130070ID), INMC, and BK21-plus.

Authors' contributions

JK and JL prove the achievability of polar codes in non-identically distributed channels and propose the channel interleaver algorithm to enhance the system reliability. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 27 July 2015 Accepted: 27 November 2016

Published online: 16 December 2016

References

1. E Arikan, Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inform. Theory*. **55**(7), 3051–3073 (2009)
2. M Alsan, Conditions for robustness of polar codes in the presence of channel mismatch (2013). e-print ArXiv:1303.2379v2
3. M Alsan, E Telatar, in *IEEE International Symposium on Information Theory (ISIT) 2014*. A simple proof of polarization and polarization for non-stationary channels, (Honolulu, 2014), pp. 301–305
4. D Blackwell, L Breiman, AJ Thomasian, The capacity of a class of channels. *Ann. Math. Stat.* **3**(4), 1229–1241 (1959)
5. SH Hassani, R Urbanke, in *IEEE International Symposium on Information Theory (ISIT) 2014*. Universal polar codes, (Honolulu, 2014), pp. 1451–1455

6. E Sasoglu, L Wang, in *IEEE International Symposium on Information Theory (ISIT) 2014*. Universal polarization, (Honolulu, 2014), pp. 1456–1460
7. E Hof, I Sason, S Shamai, in *IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*. Polar coding for degraded and non-degraded parallel channels, (Eilat, 2010), pp. 550–554
8. E Hof, I Sason, S Shamai, C Tian, Capacity-achieving codes for arbitrarily permuted parallel channels. *IEEE Trans. Inform. Theory*. **59**(3), 1505–1516 (2013)
9. R Mori, T Tanaka, Performance of Polar codes with the construction using density evolution. *IEEE Commun. Lett.* **13**(7), 519–521 (2009)
10. E Sasoglu, in *IEEE International Symposium on Information Theory Proceedings (ISIT) 2012*. Polar codes for discrete alphabets, (Cambridge, 2012), pp. 2137–2141
11. W Park, A Barg, Polar codes for q -ary channels, $q = 2^r$. *IEEE Trans. Inform. Theory*. **59**(2), 955–969 (2013)
12. A Eslami, H Pishro-Nik, in *IEEE International Symposium on Information Theory Proceedings (ISIT) 2011*. A practical approach to polar codes, (St. Petersburg, 2011), pp. 16–20

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com