Generalized Partial Orders for Polar Code Bit-Channels

Wei Wu and Paul H. Siegel, *Life Fellow, IEEE*

***Abstract*— We study partial orders (POs) for the synthesized bit-channels of polar codes. First, we give an alternative proof of an existing PO for bit-channels with the same Hamming weight and use the underlying idea to extend the bit-channel ordering to some additional cases. In particular, the bit-channel ordering for a given code block length is used to generate additional bit-channel ordering relationships for larger block lengths, generalizing previously known POs. Next, we consider POs especially for the binary erasure channel (BEC). We identify a symmetry property of the Bhattacharyya parameters of complementary bit-channel pairs on the BEC and provide a condition for the alignment of polarized sets of bit-channels for the BEC and general binary-input memoryless symmetric (BMS) channels. Numerical examples and further properties about the POs for the bit-channels with different Hamming weights are provided to illustrate the new POs. The bit-channels with universal ordering positions, which are independent of the channel erasure probability, are verified for all of the code block lengths. Finally, we show the threshold behavior of the Bhattacharyya parameters of some bit-channels by approximating the threshold values. The corresponding value for a bit-channel can be used to determine whether it is good or bad when the underlying channel is known.**

***Index Terms*— Polar codes, partial order, Bhattacharyya parameter, induction, erasure channel, function composition.**

I. Introduction

POLAR codes, as introduced by Arıkan [1], are the first family of codes proved to be capacity-achieving on binary-input memoryless symmetric (BMS) channels with low-complexity encoders and decoders. The code construction starts from a *channel transformation*, where N synthesized bit-channels $W_N^{(i)}$, $i = 0, 1, \ldots, N-1$, are obtained by applying a linear transformation to N independent copies of a BMS channel W. As the block length N goes to infinity, the synthesized bit-channels become either noiseless or completely noisy. Encoding has time complexity $O(N \log_2 N)$, as does decoding by means of successive cancellation (SC), whereby the bit-channels $W_N^{(i)}$ are sequentially decoded.

Manuscript received March 5, 2018; revised May 20, 2019; accepted June 23, 2019. Date of publication July 22, 2019; date of current version October 18, 2019. This work was supported in part by the Center for Memory and Recording Research at University of California, San Diego and in part by the NSF under Grant CCF-1405119 and Grant CCF-1619053. This paper was presented in part at the 55th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, October 2017.

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093 USA, and also with the Center for Memory and Recording Research, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: wew128@ucsd.edu; psiegel@ucsd.edu).

Communicated by M. Lentmaier, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2019.2930292

A polar code carries the information on the least noisy bit-channel positions and freezes the remaining ones to a predetermined value, usually chosen to be zero. However, except for the binary erasure channel (BEC), it is generally difficult to precisely measure the quality of the bit-channel $W_N^{(i)}$ because of the exponentially growing output alphabet size as a function of the bit-channel index. Several methods have been proposed to help select the information-bearing bit-channels: Monte Carlo simulation was discussed in [1], density evolution was used in [18], and a Gaussian approximation for density evolution was proposed in [30]. In [29], Tal and Vardy accurately approximated the error probabilities of bit-channels by using efficient degrading and upgrading quantization schemes.

Another important characteristic of polar codes is that the bit-channel orderings are channel-dependent. Although no general rule is known for completely ordering the bit-channels of a general BMS channel W, some partial orders (POs) that are independent of the underlying channel W have been found for selected bit-channels [3], [19], [28]. In [19], an ordering applicable to bit-channels with different Hamming weights was presented. (The Hamming weight of $W_N^{(i)}$ is defined as the number of ones in the binary expansion of i.) It stated that a bit-channel $W_N^{(j)}$ is stochastically degraded with respect to $W_N^{(i)}$ if the positions of 1 in the binary expansion of j are a subset of the positions of 1 in the binary expansion of i. The ordering in [28, Theorem 1] and [3] compared bit-channels with the same Hamming weight. It was based on the observation that a bit-channel $W_N^{(j)}$ is stochastically degraded with respect to $W_N^{(i)}$ if j is obtained by swapping a more significant 1 with a less significant 0 in the binary expansion of i. Both of these orderings are partial, in the sense that not all bit-channel pairs $(W_N^{(i)}, W_N^{(j)})$ are comparable.

Valuable insight into the structural properties of the information set of a polar code can be obtained with the help of POs. In particular, the POs mentioned above were used in [17] to show that it suffices to know the reliability of a sublinear number of bit-channels in order to construct a polar code. In a recent paper [33], a localization-based construction method with such sublinear complexity was proposed, making use of a group-based PO diagram to directly find the information set. In [11], a theoretical framework for fast code construction over additive white Gaussian noise (AWGN) channels was established using a mathematical concept called β-expansion that was based on the PO results. The POs were also used in [13], to study the algebraic structure and some invariant properties of the information set. The PO

in [28] and [3] traces back to an earlier paper [2], in which polar codes were applied in a public key cryptosystem.

There are several other works in the literature that relate to bit-channel orderings. The minimum distance of a polar code was shown to be 2 to the power of the minimum Hamming weight of bit-channels in the information set [9], [14]. In [23], a modified construction based on swapping weak information bits with strong frozen bits was proposed to improve the performance under belief propagation (BP) and successive cancellation list (SCL) decoding. In the asymptotic regime, it was proved that the distribution of frozen bits converges to a limit distribution [31]. Moreover, the information set was proved to have several properties that were common to fractals [7]. In the finite regime, the authors in [20] showed that each of the bit-channels obtained by applying polarization to a BEC has a sharp threshold for sufficiently large block length.

In this paper, we present further results related to bit-channel orderings. They are summarized as follows.

- *POs for BMS channels*: We provide an elementary proof based on mathematical induction for the PO proposed in [28] and [3]. Then, we use the proof idea to identify new general POs for bit-channels with different Hamming weights. Specifically, we show that when certain types of bit-channel orderings hold for a given code block length, some very general POs can be established for larger block lengths. The existence of the underlying bit-channel orderings is discussed for both BEC and other BMS channels.
- *POs for the BEC*: We consider the bit-channel pairs whose polarization processes are described by a pair of complementary binary sequences. Their Bhattacharyya parameters exhibit a symmetry property on the BEC, whose implications we explore. This motivates the investigation of conditions for the alignment of polarized sets of bit-channels for the BEC and other BMS channels. We present several examples to illustrate the new POs for the BEC and discuss several further ordering properties. In particular, we consider the ordering between the worst bit-channel with Hamming weight n and the best bit-channel with Hamming weight m at polarization step $t = n + m$, where $m < n$, as a function of the pair (m, n). Finally, we characterize bit-channels with universal ordering positions, which are independent of the channel erasure probability, for all of the code block lengths.
- *Threshold behavior*: We study the threshold value for the worst bit-channel with a given Hamming weight for the BEC by analyzing the polynomial expression for the Bhattacharyya parameter. This threshold value can be used to determine whether the bit-channel is good or bad when the erasure probability of the underlying channel is fixed. (A similar result about threshold behavior for the asymptotic case was presented in [7].) For general BMS channels, we determine the relationship between threshold values associated with lower and upper bounds on the Bhattacharyya parameter.

Portions of this paper were presented in [32]. In the present paper, we further analyze generalized POs for general BMS channels, more completely characterize the scenarios where bit-channel orderings at a given block length induce the most orderings at larger block lengths, provide a complete characterization of universal BEC bit-channels that preserve their ordering positions independent of the channel erasure probability, and study the threshold behavior of bit-channels for the BEC and general BMS channels.

The rest of this paper is organized as follows. In Section II, we present notation and definitions, as well as some basic results relating to key channel parameters. In Section III, we give an elementary proof of the PO proposed in [28] and [3]; then, we use the proof technique to derive some generalized POs. In Section IV, we concentrate on POs for the BEC and present numerical examples and further properties for the new generalized POs. The threshold behavior of bit-channels for both BEC and general BMS channels is discussed in Section V. Finally, Section VI summarizes the paper and discusses some directions for future research.

II. Preliminaries

A. Channel Parameters

Consider a BMS channel given by $W : \mathcal{X} \to \mathcal{Y}$, with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y}$ and transition probabilities $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$. Define the *Bhattacharyya parameter* of the channel W as

$$Z(W) = \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$$

and the error probability of W with uniform input under maximum-likelihood decoding as

$$P_e(W) = \frac{1}{2} \sum_{y \in \mathcal{Y}} \min\{W(y|0), W(y|1)\}.$$

Note that when $W = \mathrm{BEC}(\epsilon)$, i.e., the channel erasure probability is ϵ, we have $Z(W) = \epsilon$.

The mutual information between the input and output of W with uniform input distribution, which is also the symmetric capacity of W, can be written as

$$I(W) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log_2 \frac{2W(y|x)}{W(y|0) + W(y|1)}.$$

Each of the above three parameters can measure the reliability of a BMS channel W. Roughly speaking, a BMS channel W with higher reliability has a lower $Z(W)$, lower $P_e(W)$, and higher $I(W)$. But different parameters used for comparing two channels may lead to different channel orderings in some cases. In this paper, the Bhattacharyya parameter $Z(W)$ is used as the key parameter to measure the reliability of a BMS channel W.

The following relations between $Z(W)$ and $P_e(W)$ were given in [10]:

$$1 - \sqrt{1 - Z(W)^2} \leqslant 2P_e(W) \leqslant Z(W). \tag{1}$$

In addition, the following relations between $Z(W)$ and $I(W)$ were proved in [8]:

$$1 - Z(W) \leqslant I(W) \leqslant 1 - h_2\left(\frac{1 - \sqrt{1 - Z(W)^2}}{2}\right), \tag{2}$$

where $h_2(\cdot)$ denotes the binary entropy function, i.e., $h_2(x) = -x\log_2(x)-(1-x)\log_2(1-x)$. These lower and upper bounds are both tighter than those presented in [1].

Based on (1) and (2), we conclude with two lemmas that provide useful relationships among channel orderings with respect to the Bhattacharyya parameter, error probability, and symmetric capacity.

Lemma 1. For any two BMS channels W and Q, if $Z(W) \geqslant \beta_1 \geqslant \beta_2 \geqslant Z(Q)$, where β_1 and β_2 satisfy

$$h_2\left(\frac{1-\sqrt{1-\beta_1^2}}{2}\right) \geqslant \beta_2, \tag{3}$$

then we have $I(W) \leqslant I(Q)$. Since $h_2(\frac{x}{2}) \geqslant x$ holds for any $x \in [0,1]$, if β_1 and β_2 satisfy the stronger relation

$$1-\sqrt{1-\beta_1^2} \geqslant \beta_2,$$

then we have $I(W) \leqslant I(Q)$ and $P_e(W) \geqslant P_e(Q)$.

Lemma 2. For any two BMS channels W and Q, if $P_e(W) \geqslant \lambda_1 \geqslant \lambda_2 \geqslant P_e(Q)$, where λ_1 and λ_2 satisfy

$$2\lambda_1 \geqslant \sqrt{1-(1-2\lambda_2)^2},$$

then $Z(W) \geqslant Z(Q)$.

A stronger measure of channel ordering is provided by the channel degradation relation.

Definition 1. The channel $W : \mathcal{X} \to \mathcal{Y}$ is *stochastically degraded* with respect to the channel $Q : \mathcal{X} \to \mathcal{Z}$ if there exists a channel $P : \mathcal{Z} \to \mathcal{Y}$ such that

$$W(y|x) = \sum_{z\in\mathcal{Z}} Q(z|x)P(y|z)$$

for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$.

We write $W \preccurlyeq Q$ to denote that W is stochastically degraded with respect to Q. For clarity later on, we may also write it as $Q \succcurlyeq W$ to describe the same channel relation between W and Q. The following lemma states that the relation of channel parameters between different channels can be obtained from the channel degradation relation.

Lemma 3. [29, Lemma 3] For any two BMS channels W and Q, if $W \preccurlyeq Q$, then we have the following three inequalities:

$$\begin{aligned} Z(W) &\geqslant Z(Q), \\ P_e(W) &\geqslant P_e(Q), \\ I(W) &\leqslant I(Q). \end{aligned}$$

In general, however, the converse is not true. That is, the channel degradation relation does not hold even if all three of the inequalities are satisfied. However, when W and Q are two BECs, any one of the three inequalities is enough to imply the channel degradation relation.

B. Channel Transformation

Consider the channel transformation $W \to (W^0, W^1)$ defined in the following manner. Starting from a BMS channel $W : \{0,1\} \to \mathcal{Y}$, the channels $W^0 : \{0,1\} \to \mathcal{Y}^2$ and $W^1 : \{0,1\} \to \{0,1\} \times \mathcal{Y}^2$ are defined as

$$W^0(y_0, y_1|x_0) = \sum_{x_1\in\{0,1\}} \frac{1}{2} W(y_0|x_0 \oplus x_1)W(y_1|x_1), \tag{4}$$

$$W^1(y_0, y_1, x_0|x_1) = \frac{1}{2} W(y_0|x_0 \oplus x_1)W(y_1|x_1). \tag{5}$$

When there are $N = 2^t$ independent copies of BMS channel W, the channel transformation can be recursively repeated t times to produce N bit-channels $W_N^{(i)} : \{0,1\} \to \mathcal{Y}^N \times \{0,1\}^i, 0 \leqslant i \leqslant N-1$, defined as follows. For any $0 \leqslant i \leqslant N-1$, let $b^t = b_1 b_2 \cdots b_t$ be the binary expansion of i, where b_1 is the most significant digit. For example, if $t = 3$ and $i = 6$, the corresponding binary expansion of i is $b^3 = b_1\ b_2\ b_3 = 110$. The bit-channel $W_N^{(i)}$ is defined as

$$W_N^{(i)} = W^{b^t} \stackrel{\text{def}}{=} (((W^{b_1})^{b_2})^{\cdots})^{b_t}.$$

We denote the binary complement of b^t by $\bar{b^t} = \bar{b_1}\bar{b_2}\cdots\bar{b_t}$. Here, $b_j \oplus \bar{b_j} = 1$ for any $1 \leqslant j \leqslant t$. Then, $\bar{b^t}$ is the binary expansion of $N-1-i$, so $W_N^{(N-i-1)} = W^{\bar{b^t}}$. It should be noted that $W_N^{(i)}$ is also a BMS channel. In particular, when W is a BEC, each of the bit-channels $W_N^{(i)}$ is also a BEC.

The PO in [19] is based on the relation between two bit-channels obtained by one-step channel transformation, which is stated as follows.

Lemma 4. [26, Lemma 3] Let W be a BMS channel. Then, W^0 is stochastically degraded with respect to W^1,

$$W^0 \preccurlyeq W^1.$$

The channel transformation above yields the following results relating to the Bhattacharyya parameters of the channels W^0, W^1, and W [10]:

$$Z(W)\sqrt{2-Z(W)^2} \leqslant Z(W^0) \leqslant 2Z(W) - Z(W)^2, \tag{6}$$

$$Z(W^1) = Z(W)^2. \tag{7}$$

The lower bound and upper bound in (6) are achieved when W is a binary symmetric channel (BSC) and a BEC, respectively. In particular, when $W = \text{BEC}(\epsilon)$, we have $W^0 = \text{BEC}(2\epsilon - \epsilon^2)$ and $W^1 = \text{BEC}(\epsilon^2)$, with $Z(W^0) = 2\epsilon - \epsilon^2$ and $Z(W^1) = \epsilon^2$. As a result, for the BEC there is an explicit recursion for the bit-channels produced by the channel transformation.

Define $f_{-1}(x) = x\sqrt{2-x^2}$, $f_0(x) = 2x - x^2$, $f_1(x) = x^2$, $x \in [0,1]$, and

$$F_{b^t} = f_{b_t} \circ \cdots \circ f_{b_2} \circ f_{b_1}, \tag{8}$$

$$G_{b^t} = f_{2b_t-1} \circ \cdots \circ f_{2b_2-1} \circ f_{2b_1-1}. \tag{9}$$

When W is a BEC, we have $Z(W^{b^t}) = F_{b^t}(Z(W))$. For any BMS channel W and any $b^t \in \{0,1\}^t$, we have

$$G_{b^t}(Z(W)) \leqslant Z(W^{b^t}) \leqslant F_{b^t}(Z(W)), \tag{10}$$

where the upper bound is achieved when W is a BEC. We also define $f_j^{(k)} = \underbrace{f_j \circ f_j \circ \cdots \circ f_j}_{k}$ for $j = -1, 0, 1$ and $k \geqslant 1$.

The relationship between (8) and (9) is stated in the following lemma, whose proof is given in Appendix A.

Lemma 5. For any binary sequence $a^k \in \{0,1\}^k$, $k \geqslant 1$, and any $x \in [0,1]$, we have

$$G_{a^k}^2(x) = F_{a^k}(x^2). \tag{11}$$

In [5], similar relations involving the error probability were proved:

$$P_e(W^0) = 2P_e(W) - 2P_e(W)^2, \tag{12}$$

$$P_e(W^1) \geqslant 2P_e(W)^2. \tag{13}$$

For any $\varepsilon \in (0, \frac{1}{2})$, with ε usually taken to be very small, the following sets of *ε-good* and *ε-bad* bit-channels are often considered:

$$\mathcal{G}_N^{\varepsilon}(W) \triangleq \{i \in [N] : Z(W_N^{(i)}) \leqslant \varepsilon\}, \tag{14}$$

$$\mathcal{B}_N^{\varepsilon}(W) \triangleq \{i \in [N] : Z(W_N^{(i)}) \geqslant 1-\varepsilon\}, \tag{15}$$

where $[N]$ denotes the set $\{0, 1, \ldots, N-1\}$.

The channel transformation in (4) and (5) preserves the degradation relation when the underlying channels are BMS channels.

Lemma 6. [29, Lemma 5] Given two BMS channels W and Q, if $Q \preccurlyeq W$, then

$$Q^0 \preccurlyeq W^0 \quad \text{and} \quad Q^1 \preccurlyeq W^1.$$

Note that when both W and Q are BECs, either $Q^0 \preccurlyeq W^0$ or $Q^1 \preccurlyeq W^1$ implies $Q \preccurlyeq W$ because of the monotonicity of f_0 and f_1.

III. Partial Orders for BMS Channels

In this section, we give an elementary proof of the PO proposed in [28] and [3] based on mathematical induction. We then use the underlying idea to derive more generalized POs.

A. New Proof

The following basic lemma is deduced in [28] from a discussion of the case $N = 4$.

Lemma 7. For any BMS channel W, when $N = 4$, we have the bit-channel relation $W^{01} \preccurlyeq W^{10}$.

Remark 1. This implies that $Z(W^{10}) \leqslant Z(W^{01})$ for any BMS channel W, a fact that also follows immediately from (6) and (7), which give:

$$Z(W^{10}) \leqslant 2Z(W^1) - Z(W^1)^2 = 2Z(W)^2 - Z(W)^4,$$

$$Z(W^{01}) = Z(W^0)^2 \geqslant [Z(W)\sqrt{2 - Z(W)^2}]^2.$$

If W is replaced by W^{p^n}, for any $p^n \in \{0,1\}^n$ and $n \geqslant 1$, Lemma 7 implies

$$W^{p^n 01} \preccurlyeq W^{p^n 10}.$$

Applying Lemma 6 recursively yields

$$W^{p^n 01 q^m} \preccurlyeq W^{p^n 10 q^m} \tag{16}$$

for all $q^m \in \{0,1\}^m$ and $m \geqslant 1$. We now restate the main PO in [28] as follows.

Theorem 1. For any BMS channel W, the degradation relation

$$W^{p^n 0 r^l 1 q^m} \preccurlyeq W^{p^n 1 r^l 0 q^m} \tag{17}$$

holds for all $p^n \in \{0,1\}^n$, $r^l \in \{0,1\}^l$, $q^m \in \{0,1\}^m$, and $n, l, m \geqslant 0$.

Proof: The proof proceeds by mathematical induction. It is easy to see that when $l = 0$, (17) reduces to (16). Now assume (17) is true for $l = k$, $k \geqslant 0$, i.e.,

$$W^{p^n 0 r^k 1 q^m} \preccurlyeq W^{p^n 1 r^k 0 q^m}$$

holds for all $r^k \in \{0,1\}^k$, $p^n \in \{0,1\}^n$, $q^m \in \{0,1\}^m$, and $n, m \geqslant 0$.

For the induction step, let $l = k+1$, denoting the additional bit by r_{k+1}. There are two cases to consider.

1) If $r_{k+1} = 0$, the following relations hold:

$$\begin{aligned} W^{p^n 0 r^{k+1} 1 q^m} &= W^{p^n 0 r^k 01 q^m} \overset{(a)}{\preccurlyeq} W^{p^n 0 r^k 10 q^m} \\ &\overset{(b)}{\preccurlyeq} W^{p^n 1 r^k 00 q^m} = W^{p^n 1 r^{k+1} 0 q^m}. \end{aligned} \tag{18}$$

Here, (a) follows from (16) and (b) is based on the induction hypothesis for $l = k$.

2) If $r_{k+1} = 1$, the following relations hold:

$$\begin{aligned} W^{p^n 0 r^{k+1} 1 q^m} &= W^{p^n 0 r^k 11 q^m} \overset{(c)}{\preccurlyeq} W^{p^n 1 r^k 01 q^m} \\ &\overset{(d)}{\preccurlyeq} W^{p^n 1 r^k 10 q^m} = W^{p^n 1 r^{k+1} 0 q^m}. \end{aligned} \tag{19}$$

Here, (c) is based on the induction hypothesis for $l = k$ and (d) follows from (16).

Combining cases 1) and 2), we have

$$W^{p^n 0 r^{k+1} 1 q^m} \preccurlyeq W^{p^n 1 r^{k+1} 0 q^m}$$

for all $r^{k+1} \in \{0,1\}^{k+1}$, $p^n \in \{0,1\}^n$, $q^m \in \{0,1\}^m$, and $n, m \geqslant 0$. This completes the induction step. ∎

Remark 2. Theorem 1 presents the same PO as in [28, Theorem 1], and (16) leads to the Covering Relation (CR) of the POs proposed in [28].

B. Generalized PO

The PO in [28] and [3] applies to bit-channels with the same Hamming weight. Except for the PO in [19], little is known about relations among bit-channels with different Hamming weights. We now exploit the idea in the proof of Theorem 1 to derive some additional bit-channel orderings, thus generalizing the PO in [28] and [3] to bit-channels with possibly different Hamming weights.

Theorem 2. Let a^k be a binary sequence of length k and b^m, c^m be binary sequences of length m, for $k \geqslant 0$ and $m \geqslant 1$. Let W be a BMS channel, and assume $W^{b^m} \preccurlyeq W^{c^m}$. If the condition

$$W^{b^m a^k 1^n} \preccurlyeq W^{c^m a^k 0^n} \tag{$\star$}$$

holds for some $n \geqslant 1$, then

$$W^{b^m a^k d^h 1^n} \preccurlyeq W^{c^m a^k d^h 0^n} \tag{20}$$

holds for all $d^h \in \{0,1\}^h$ and $h \geqslant 1$.

Proof: We first let $h = 1$. There are two cases to consider.

1) If $d_1 = 0$, we have

$$W^{b^m a^k 0 1^n} \preccurlyeq W^{b^m a^k 1^n 0} \preccurlyeq W^{c^m a^k 0^n 0} = W^{c^m a^k 0 0^n}.$$

The first degradation relation follows from the PO in Theorem 1, and the second degradation relation is based on applying Lemma 6 to the condition ($\star$).

2) If $d_1 = 1$, we have

$$W^{b^m a^k 1 1^n} = W^{b^m a^k 1^n 1} \preccurlyeq W^{c^m a^k 0^n 1} \preccurlyeq W^{c^m a^k 1 0^n}.$$

The first degradation relation is based on applying Lemma 6 to the condition ($\star$), and the second degradation relation follows from the PO in Theorem 1.

Combining cases 1) and 2) gives

$$W^{b^m a^k d_1 1^n} \preccurlyeq W^{c^m a^k d_1 0^n}.$$

Repeating this argument for successive inserted bits $d_2, \ldots, d_h$ completes the proof. ∎

Remark 3. In fact, Theorem 2 is a generalization of Theorem 1 inspired by its proof. Taking $m = n = 1$ and $b^m = 0$, $c^m = 1$ the condition ($\star$) in Theorem 2 becomes the induction hypothesis in Theorem 1 (by omitting the same beginning and ending sequences, i.e., p^n and q^m used in Theorem 1). Note that the induction step in the proof of Theorem 1 does not depend on the sequences just ahead of r^k, i.e., 0 and 1; thus, if we substitute them with some other sequences with the degradation relation preserved, e.g., b^m and c^m ($W^{b^m} \preccurlyeq W^{c^m}$) in Theorem 2, the result (20) will follow provided that the condition ($\star$) holds.

Remark 4. In contrast to the assumption in Theorem 2, if $W^{b^m} \succcurlyeq W^{c^m}$, then

$$W^{b^m a^k 1^n} \succcurlyeq W^{c^m a^k 0^n}$$

holds for all $a^k \in \{0,1\}^k$ and $k \geqslant 0$. Hence, there is no a^k that satisfies the condition ($\star$).

Remark 5. If $W^{b^m} \preccurlyeq W^{c^m}$, we consider two scenarios, based upon the difference between the Hamming weights of the two binary sequences, $\mathrm{wt}(c^m) - \mathrm{wt}(b^m)$.

1) $\mathrm{wt}(c^m) - \mathrm{wt}(b^m) \geqslant n$: Applying the PO in [19] and the PO in Theorem 1, we get $W^{b^m 1^n} \preccurlyeq W^{c^m 0^n}$. According to Theorem 2, this implies that the condition ($\star$) is satisfied for all $a^k \in \{0,1\}^k$ and any $k \geqslant 0$.
2) $\mathrm{wt}(c^m) - \mathrm{wt}(b^m) < n$: This case is non-trivial since the a^k that satisfies the condition ($\star$) depends on the underlying channel W, the values of (m, n), and the sequences b^m, c^m. In Section III-C, we prove a strong result about the existence of such a^k when W is a BEC. A discussion for general BMS channels is also provided.

The following corollary is immediately obtained by setting $b^m = 0^m$ and $c^m = 1^m$ in Theorem 2, noting that $W^{0^m} \preccurlyeq W^{1^m}$ for any BMS channel W, and taking into consideration Remark 5.

Corollary 1. Let $b^m = 0^m$ and $c^m = 1^m$. If for some $k^* \geqslant 0$, $W^{0^m a^{k^*} 1^n} \preccurlyeq W^{1^m a^{k^*} 0^n}$ holds for all $a^{k^*} \in \{0,1\}^{k^*}$, then

$$W^{0^m a^k 1^n} \preccurlyeq W^{1^m a^k 0^n}$$

holds for all $a^k \in \{0,1\}^k$ and $k \geqslant k^*$. When $m \geqslant n$, the condition holds for $k^* = 0$.

In a similar manner, if W is a BEC, the PO for larger N can be used to deduce bit-channel orderings for smaller N.

Theorem 3. Let e^k be a binary sequence of length k, $k \geqslant 0$. Assume W is a BEC. If

$$W^{0^m e^k 1^n} \succcurlyeq W^{1^m e^k 0^n} \tag{†}$$

holds for some $m < n$, then

$$W^{0^m e^{k'} 1^n} \succcurlyeq W^{1^m e^{k'} 0^n} \tag{21}$$

holds for any $0 \leqslant k' \leqslant k$.

Proof: There are two cases to consider involving the last bit e_k of the middle binary sequence.

1) If $e_k = 0$, then

$$\begin{aligned} W^{0^m e^{k-1} 1^n 0} &\succcurlyeq W^{0^m e^{k-1} 0 1^n} = W^{0^m e^k 1^n} \\ &\succcurlyeq W^{1^m e^k 0^n} = W^{1^m e^{k-1} 0^n 0}. \end{aligned}$$

The first upgradation relation follows from Theorem 1, and the second upgradation relation is the condition (†).

2) If $e_k = 1$, then

$$\begin{aligned} W^{0^m e^{k-1} 1^n 1} &= W^{0^m e^k 1^n} \succcurlyeq W^{1^m e^k 0^n} \\ &= W^{1^m e^{k-1} 1 0^n} \succcurlyeq W^{1^m e^{k-1} 0^n 1}. \end{aligned}$$

The first upgradation relation is the condition (†), and the second upgradation relation follows from Theorem 1.

Combining cases 1) and 2) gives

$$W^{0^m e^{k-1} 1^n e_k} \succcurlyeq W^{1^m e^{k-1} 0^n e_k}. \tag{22}$$

If W and V are BECs, $W^0 \succcurlyeq V^0$ or $W^1 \succcurlyeq V^1$ implies $W \succcurlyeq V$. Therefore, the relation (22) implies

$$W^{0^m e^{k-1} 1^n} \succcurlyeq W^{1^m e^{k-1} 0^n}.$$

The above argument reduces the length of the middle binary sequence e^k by 1 in (†). We can repeat the same process to get the result for any shorter middle sequence in (21). ∎

Remark 6. Theorem 3 only considers the case $m < n$ because, according to part 1) of Remark 5, the condition (†) never holds when $m \geqslant n$.

C. The Condition ($\star$)

Theorem 2 provides a generalized PO that does not require the bit-channels to have the same Hamming weight, provided the ordering in the condition ($\star$) holds. The next proposition shows that, if the channel W is a BEC, then for any $m, n \geqslant 1$, the condition ($\star$) holds for some sequence a^k, for sufficiently large k.

Proposition 1. Let $W = \mathrm{BEC}(\epsilon)$ for some $\epsilon \in (0, 1)$, and $m, n \geqslant 1$. Let b^m, c^m be distinct binary sequences of length m such that $W^{b^m} \preccurlyeq W^{c^m}$. Then, for sufficiently large k, there exists a finite-length sequence $a^k \in \{0, 1\}^k$, such that

$$Z(W^{b^m a^k 1^n}) \geqslant Z(W^{c^m a^k 0^n}). \tag{23}$$

Therefore, $W^{b^m a^k 1^n} \preccurlyeq W^{c^m a^k 0^n}$.

Proof: According to (8), we have

$$Z(W^{b^m a^k 1^n}) = f_1^{(n)}[F_{a^k}(F_{b^m}(\epsilon))] = [F_{a^k}(F_{b^m}(\epsilon))]^{2^n},$$

$$Z(W^{c^m a^k 0^n}) = f_0^{(n)}[F_{a^k}(F_{c^m}(\epsilon))] = 1 - [1 - F_{a^k}(F_{c^m}(\epsilon))]^{2^n}.$$

We want to find a^k such that (23) is satisfied, that is

$$[F_{a^k}(F_{b^m}(\epsilon))]^{2^n} + [1 - F_{a^k}(F_{c^m}(\epsilon))]^{2^n} \geqslant 1. \tag{24}$$

Let $(a_\ell)_{\ell \geqslant 1}$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli($\frac{1}{2}$) random variables. Let a^k be a realization of the random vector $a_1\ a_2 \cdots a_k$. Then almost surely, the function F_{a^k} exhibits a threshold behavior as k grows large. To be more precise, according to [10, Lemma 11], there exists a point $\epsilon^* \in [0, 1]$ such that

$$\lim_{k\to\infty} F_{a^k}(z) = \begin{cases} 0, & z \in [0, \epsilon^*), \\ 1, & z \in (\epsilon^*, 1]. \end{cases} \tag{25}$$

Here, ϵ^* depends on the realization a^k and has a uniform distribution on [0, 1]. We use this result to complete the proof.

Since $W^{b^m} \preccurlyeq W^{c^m}$ and b^m, c^m are distinct, then we have $0 < F_{c^m}(\epsilon) < F_{b^m}(\epsilon) < 1$. We can therefore find a threshold point ϵ^* such that $F_{c^m}(\epsilon) < \epsilon^* < F_{b^m}(\epsilon)$. Therefore, there exists a realization a^k with a sufficiently large k such that

$$F_{a^k}(F_{b^m}(\epsilon)) \geqslant 1 - \theta(n),$$
$$F_{a^k}(F_{c^m}(\epsilon)) \leqslant \theta(n),$$

where $\theta(n)$ is a function of n. Here, to satisfy the inequality in (24), $\theta(n) \triangleq 1 - 2^{-\frac{1}{2^n}}$. These inequalities imply that

$$[F_{a^k}(F_{b^m}(\epsilon))]^{2^n} + [1 - F_{a^k}(F_{c^m}(\epsilon))]^{2^n} \geqslant 2[1 - \theta(n)]^{2^n} = 1$$

and thus (23) holds. The degradation relation follows from the fact that the bit-channels are BECs. ∎

Generalized POs for bit-channels with different Hamming weights are proposed in Theorem 2 for any BMS channel; however, the degradation relation in the condition ($\star$), which states that $W^{b^m a^k 1^n} \preccurlyeq W^{c^m a^k 0^n}$, may be difficult to verify for any BMS channel W other than BEC. It would therefore be desirable to have a bit-channel ordering result based upon a more easily verified condition involving the Bhattacharyya parameter or error probability. The following two propositions represent a step in that direction.

Proposition 2. For any BMS channel W and an $a^k \in \{0, 1\}^k$, if $Z(W^{b^m a^k 1^n}) \geqslant Z(W^{c^m a^k 0^n})$, then

$$Z(W^{b^m a^k 1 1^n}) \geqslant Z(W^{c^m a^k 1 0^n}).$$

Proposition 3. For any BMS channel W and an $a^k \in \{0, 1\}^k$, if $P_e(W^{b^m a^k 1^n}) \geqslant P_e(W^{c^m a^k 0^n})$, then

$$P_e(W^{b^m a^k 0 1^n}) \geqslant P_e(W^{c^m a^k 0 0^n}).$$

Remark 7. The proofs of Proposition 2 and Proposition 3 are derived from equations (7) and (12), respectively, using Lemma 3 and the PO in Theorem 1.

In connection with Proposition 2, we can prove a result for a general BMS channel analogous to Proposition 1 showing the existence of a^k satisfying the required Bhattacharyya parameter ordering condition for sufficiently large k. In Proposition 1, the assumption is $W^{b^m} \preccurlyeq W^{c^m}$. This is equivalent to $Z(W^{b^m}) \geqslant Z(W^{c^m})$ when W is a BEC. For a general BMS channel, the difference compared to Proposition 1 is that a stronger assumption about the Bhattacharyya parameter ordering is required.

Proposition 4. Let W be a general BMS channel and $m, n \geqslant 1$. Let b^m, c^m be binary sequences of length m such that $Z(W^{b^m})^2 > Z(W^{c^m})$. Then, for sufficiently large k, there exists a finite-length sequence $a^k \in \{0, 1\}^k$, such that

$$Z(W^{b^m a^k 1^n}) \geqslant Z(W^{c^m a^k 0^n}). \tag{26}$$

Proof: According to (10), we have

$$Z(W^{b^m a^k 1^n}) \geqslant [G_{a^k}(Z(W^{b^m}))]^{2^n},$$

$$Z(W^{c^m a^k 0^n}) \leqslant 1 - [1 - F_{a^k}(Z(W^{c^m}))]^{2^n}.$$

We want to find a^k such that (26) is satisfied. Combining the above two inequalities, we get a stronger result than (26), that is

$$[G_{a^k}(Z(W^{b^m}))]^{2^n} + [1 - F_{a^k}(Z(W^{c^m}))]^{2^n} \geqslant 1. \tag{27}$$

Hence, by Lemma 5, the inequality (27) is equivalent to

$$[F_{a^k}(Z(W^{b^m})^2)]^{2^{n-1}} + [1 - F_{a^k}(Z(W^{c^m}))]^{2^n} \geqslant 1.$$

Since $Z(W^{b^m})^2 > Z(W^{c^m})$, arguing as in the proof of Proposition 1, we can find a threshold point ϵ^* such that $Z(W^{b^m})^2 > \epsilon^* > Z(W^{c^m})$. Therefore, there exists a realization a^k with a sufficiently large k such that

$$F_{a^k}(Z(W^{b^m})^2) \geqslant 1 - \theta(n),$$
$$F_{a^k}(Z(W^{c^m})) \leqslant \theta(n).$$

These inequalities imply that

$$[F_{a^k}(Z(W^{b^m})^2)]^{2^{n-1}} + [1 - F_{a^k}(Z(W^{c^m}))]^{2^n}$$
$$\geqslant [1 - \theta(n)]^{2^{n-1}} + [1 - \theta(n)]^{2^n} > 1$$

and thus (27) holds, implying that (26) holds. ∎

IV. Partial Orders for The BEC

In this section, we concentrate on POs for the BEC. First, we consider complementary bit-channel pairs whose Bhattacharyya parameters on the BEC exhibit a symmetry property. This symmetry is used to gain insight into the alignment of polarized sets of bit-channels for the BEC and other BMS channels. Next, we give several examples of Theorem 2 and Corollary 1 and discuss further properties of BEC bit-channel orderings, with particular attention on the orderings of bit-channels with different Hamming weights. Finally, we show that some bit-channels have universal ordering positions which are independent of the channel erasure probability. For all of the code block lengths, we determine all the indices of bit-channels with universal ordering positions.

A. Complementary Bit-Channels

We study the properties of bit-channel pairs $(W^{b^t}, V^{\bar{b^t}})$, whose binary expansions represent complementary polarization sequences, by examining the effect of the basic one-step channel transformation on sums of Bhattacharyya parameters.

Proposition 5. For any two BMS channels W and V, if $Z(W)+Z(V) \leqslant a, a \in [0,1]$, then $Z(W^1)+Z(V^0) \leqslant f_0(a)$ and $Z(W^0)+Z(V^1) \leqslant f_0(a)$.

Proof: According to (6) and (7), we have $Z(W^1) = Z(W)^2$ and $Z(V^0) \leqslant 2Z(V) - Z(V)^2$. Since the Bhattacharyya parameter is non-negative, the condition $Z(W)+Z(V) \leqslant a$ implies $0 \leqslant Z(W) \leqslant a - Z(V)$ and $0 \leqslant Z(V) \leqslant a$. Therefore,

$$\begin{aligned} Z(W^1)+Z(V^0) &\leqslant Z(W)^2 + 2Z(V) - Z(V)^2 \\ &\leqslant [a - Z(V)]^2 + 2Z(V) - Z(V)^2 \\ &= a^2 + 2(1-a)Z(V) \\ &\leqslant a^2 + 2a(1-a) \\ &= 2a - a^2 \\ &= f_0(a). \end{aligned}$$

Interchanging the roles of W and V, we obtain $Z(W^0)+Z(V^1) \leqslant f_0(a)$. ∎

Proposition 6. For any two BMS channels W and V, if $Z(W)+Z(V) \leqslant a$, $a \in [0,1]$, then $Z(W^{b^t})+Z(V^{\bar{b^t}}) \leqslant f_0^{(t)}(a)$, for all $b^t \in \{0,1\}^t$ and $t \geqslant 1$.

Proof: Notice that $f_0(a) \in [0,1]$ for any $a \in [0,1]$. The result can be proved by applying Proposition 5 recursively t times. ∎

We have the following three corollaries based on Proposition 6.

Corollary 2. If $Z(W)+Z(V) \leqslant 1$, then $Z(W^{b^t})+Z(V^{\bar{b^t}}) \leqslant 1$, for all $b^t \in \{0,1\}^t$ and $t \geqslant 1$.

Remark 8. In Corollary 2, the equality, i.e., $Z(W^{b^t}) + Z(V^{\bar{b^t}}) = 1$, can be achieved when W and V are BECs and $Z(W)+Z(V)=1$.

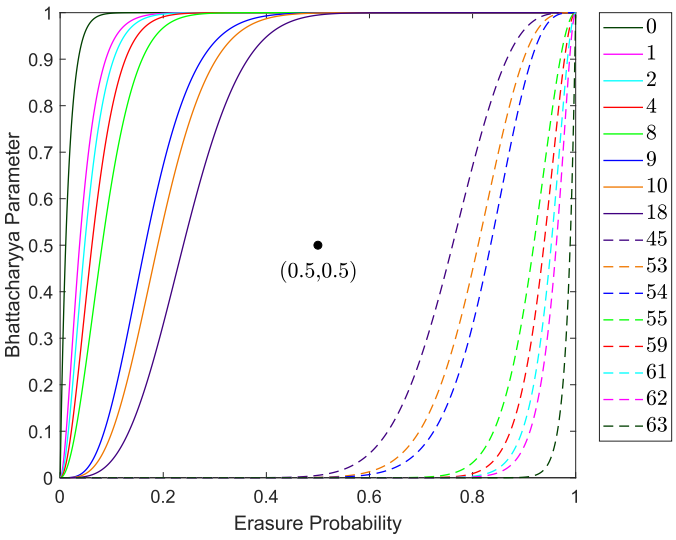


Fig. 1. BEC bit-channels with universal positions when $N = 64$.

Corollary 3. If $Z(W) \leqslant 1/2$, then $Z(W^{b^t})+Z(W^{\bar{b^t}}) \leqslant 1$, for all $b^t \in \{0,1\}^t$ and $t \geqslant 1$.

Remark 9. By Corollary 3, if W satisfies $Z(W) \leqslant 1/2$, then the two bit-channels W^{b^t} and $W^{\bar{b^t}}$ cannot both belong to $\mathcal{B}_N^{\varepsilon}(W)$ for any $b^t \in \{0,1\}^t$.

Corollary 4. Let $W = \mathrm{BEC}(\epsilon)$ and $V = W^c \stackrel{\text{def}}{=} \mathrm{BEC}(1-\epsilon)$, $\epsilon \in (0,1)$. So $Z(W) + Z(W^c) = 1$. Then, $Z(W^{b^t}) + Z((W^c)^{\bar{b^t}}) = 1$, for all $b^t \in \{0,1\}^t$ and $t \geqslant 1$.

In Figure 1, we plot the Bhattacharyya parameters of selected bit-channels obtained by polarizing BECs, for $N = 64$. Each curve corresponds to a bit-channel index i. The corresponding index in the legend increases as the curves progress from left to right. The solid and dashed curves represent complementary bit-channel pairs $(W_N^{(i)}, W_N^{(N-1-i)})$. According to Corollary 4, the curve of bit-channel $W_N^{(i)}$ can be obtained by rotating the curve of bit-channel $W_N^{(N-1-i)}$ by an angle of 180° around the point (0.5, 0.5). (This symmetry was also noted in [20].) The bit-channels shown in the figure have universal positions with respect to bit-channel ordering in the sense that their positions in the complete ordering of the 64 bit-channels are independent of the channel erasure probability. The universality can be verified by using the POs in [3], [19], [28], and some computations. These universal ordering positions will be discussed further in Section IV-D.

We conclude with some results on the alignment of polarized sets of BEC and other BMS channels.

Theorem 4. Let $V = \mathrm{BEC}(\epsilon)$, $\epsilon \in (0,1)$. For any BMS channel W, if $Z(W) + Z(V^c) \leqslant 1$, i.e., $Z(W) \leqslant \epsilon$, then the ε-*good* and ε-*bad* sets defined in (14) and (15) satisfy

$$\mathcal{G}_N^{\varepsilon}(W) \supseteq \mathcal{G}_N^{\varepsilon}(V) \quad \text{and} \quad \mathcal{B}_N^{\varepsilon}(W) \subseteq \mathcal{B}_N^{\varepsilon}(V)$$

for all $t \geqslant 1$ and $\varepsilon \in (0, \frac{1}{2})$.

Proof: By Corollary 2, we have $Z(W^{b^t})+Z((V^c)^{\bar{b^t}}) \leqslant 1$, for all $b^t \in \{0,1\}^t$ and $t \geqslant 1$. Since V is a BEC, Corollary 4

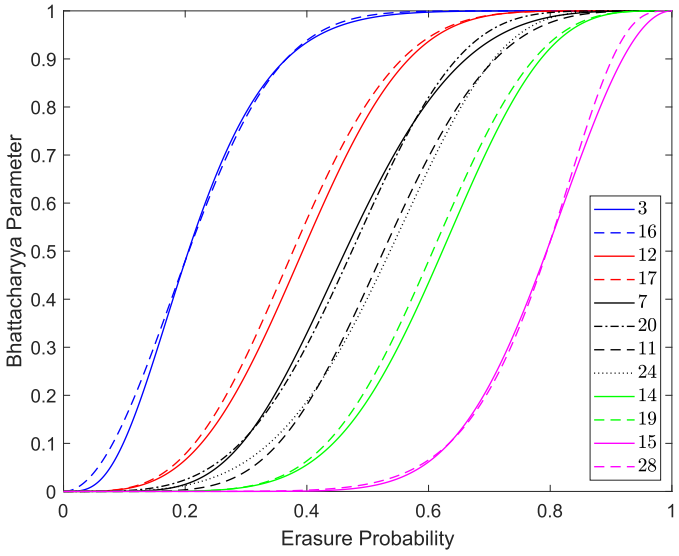


Fig. 2. BEC bit-channels with intersections when $N = 32$.

TABLE I

THE INTERSECTIONS OF ALL INTERSECTING PAIRS WHEN $N = 32$

$(3, 16)$	$(12, 17)$	$(7, 20)$	$(7, 24)$	$(11, 24)$	$(14, 19)$	$(15, 28)$
0.1997	0.7493	0.3077	0.1831	0.4228	0.2507	0.6358
0.3642		0.5772	0.8169	0.6923		0.8003

can be applied to get $Z(V^{b^t}) + Z((V^c)^{\bar{b^t}}) = 1$. Combining the above two relations, we get

$$Z(W^{b^t}) \leqslant Z(V^{b^t})$$

for all $b^t \in \{0, 1\}^t$ and $t \geqslant 1$. Therefore, $Z(V^{b^t}) \leqslant \varepsilon$ implies that $Z(W^{b^t}) \leqslant \varepsilon$, or $\mathcal{G}_N^{\varepsilon}(W) \supseteq \mathcal{G}_N^{\varepsilon}(V)$. Similarly, we find $\mathcal{B}_N^{\varepsilon}(W) \subseteq \mathcal{B}_N^{\varepsilon}(V)$. ∎

Remark 10. Theorem 4 can also be derived by using the results in (6) and (7). Further results about the alignment of polarized sets were given in [24].

B. Numerical Examples

We consider some applications of Theorem 2 and Corollary 1 to BECs in the case where $b^m = 0^m$ and $c^m = 1^m$.

1) Application of Theorem 2: The Bhattacharyya parameters of several synthesized bit-channels for BECs when $N = 32$ are shown in Figure 2. All bit-channels whose curves intersect are represented in the figure. The two pairs of intersecting curves at the left correspond to bit-channel index pairs $(3, 16)$ and $(12, 17)$, while the two pairs of intersecting curves at the right correspond to bit-channel index pairs $(14, 19)$ and $(15, 28)$. The four intersecting curves in the middle correspond to bit channel indices $(7, 20, 11, 24)$. Since the curves can be described explicitly as polynomial functions of the erasure probability ϵ, we can numerically determine the intersection points for any pair of intersecting curves. Table I lists all values of $\epsilon \in (0, 1)$ for which a pair of bit-channels $(W_N^{(i)}, W_N^{(j)})$, denoted by (i, j), has an intersection point. For example, the curves corresponding to $W_{32}^{(7)}$ and $W_{32}^{(20)}$ intersect at two locations, $\epsilon = 0.3077$ and $\epsilon = 0.5772$.

TABLE II

VALUES OF k_{min}^* FOR VARIOUS (m, n) WHEN $W = \text{BEC}(0.5)$

m \ n	1	2	3	4	5	6	7	8
1	0	3	8	10	12	13	14	15
2	0	0	0	1	3	4	5	5
3	0	0	0	0	0	0	0	1

Since the curve for $W_{32}^{(7)}$ is above the curve for $W_{32}^{(20)}$ when $0.3077 < \epsilon < 0.5772$, this implies that $W_{32}^{(7)} \preccurlyeq W_{32}^{(20)}$ when $W = \text{BEC}(\epsilon)$ for ϵ in this range. Applying Theorem 2, we obtain the following results.

Example 1. Suppose $W = \text{BEC}(\epsilon)$, where $0.3077 < \epsilon < 0.5772$. We have seen that $W_{32}^{(7)} \preccurlyeq W_{32}^{(20)}$, i.e., $W^{00111} \preccurlyeq W^{10100}$. Then, for $N = 64$, Theorem 2 implies that

$$W^{001011} \preccurlyeq W^{101000}, \text{ i.e., } W_{64}^{(11)} \preccurlyeq W_{64}^{(40)},$$
$$W^{001111} \preccurlyeq W^{101100}, \text{ i.e., } W_{64}^{(15)} \preccurlyeq W_{64}^{(44)}.$$

More generally, for $N = 2^{5+h}$, we have

$$W^{001d^h11} \preccurlyeq W^{101d^h00},$$

which holds for all $d^h \in \{0, 1\}^h$ and $h \geqslant 1$.

2) Application of Corollary 1: We define the minimum possible k^* in Corollary 1 as follows.

Definition 2. When $W = \text{BEC}(\epsilon)$, for a fixed pair (m, n), consider the set

$$S = \{k^* | W^{0^m a^{k^*} 1^n} \preccurlyeq W^{1^m a^{k^*} 0^n}, \forall a^{k^*} \in \{0, 1\}^{k^*}\}.$$

The minimum element in the set S is denoted by $k_{min}^*(m, n, \epsilon)$, i.e.,

$$k_{min}^*(m, n, \epsilon) = \min_{k^* \in S} k^*.$$

According to Corollary 1, $S = \{k^* | k^* \geqslant k_{min}^*(m, n, \epsilon)\}$. Therefore, by determining $k_{min}^*(m, n, \epsilon)$, we can deduce the largest number of new POs by adding a middle sequence with the length starting from $k_{min}^*(m, n, \epsilon)$. The properties of $k_{min}^*(m, n, \epsilon)$ will be discussed in Section IV-C. In Table II, we consider the case where $\epsilon = 0.5$ and list values of k_{min}^* for various (m, n). The values of k_{min}^* are obtained by checking all the binary sequences of the same length starting from $k^* = 1$. The following example illustrates how to interpret the results in Table II.

Example 2. Consider two cases $(m, n) = (1, 2)$ and $(m, n) = (2, 4)$. From Table II, we find that the corresponding k_{min}^* values are 3 and 1, respectively. It follows that, for $W = \text{BEC}(0.5)$, we have $W^{0a^k11} \preccurlyeq W^{1a^k00}$ for all $a^k \in \{0, 1\}^k$, $k \geqslant 3$. Similarly, $W^{00a^k1111} \preccurlyeq W^{11a^k0000}$ for all $a^k \in \{0, 1\}^k$, $k \geqslant 1$.

Remark 11. When $k_{min}^*(m, n, \epsilon) = 0$ for some (m, n, ϵ), the assumption in Corollary 1 holds with $k^* = 0$, that is, $W^{0^m 1^n} \preccurlyeq W^{1^m 0^n}$. Thus, one can derive from Corollary 1 the largest number of new POs by adding a middle sequence with any length.

C. Properties of $k^*_{min}(m, n, \epsilon)$

We now prove some additional properties of the parameter $k^*_{min}(m, n, \epsilon)$ defined in Section IV-B.2.

1) Structural Property of Table II: The following proposition considers "L"-shaped regions in tables for $k^*_{min}(m, n, \epsilon)$ with a fixed ϵ, similar to Table II ($\epsilon = 0.5$). For convenience, in the "L"-shaped region associated with $m \geqslant 1$ and $n \geqslant 1$ we write $k_1 = k^*_{min}(m+1, n, \epsilon)$, $k_2 = k^*_{min}(m+1, n+1, \epsilon)$, and $k_3 = k^*_{min}(m, n, \epsilon)$, as shown below.

k_3	
k_1	k_2

Proposition 7. The values in the "L"-shaped region above satisfy

$$k_1 \leqslant k_2 \leqslant k_3.$$

Proof: Since W is a BEC, we have

$$W^0 \preccurlyeq W \preccurlyeq W^1. \tag{28}$$

The proof includes two parts.

1) To prove $k_1 \leqslant k_2$, we consider any two adjacent values in the same row of the table for $k^*_{min}(m, n, \epsilon)$ with a fixed ϵ. Thus, for any $m \geqslant 1$ and $n \geqslant 1$, we need to show

$$W^{0^m a^k 1^{n+1}} \preccurlyeq W^{1^m a^k 0^{n+1}} \Rightarrow W^{0^m a^k 1^n} \preccurlyeq W^{1^m a^k 0^n}.$$

This follows from (28), which implies

$$W^{0^m a^k 1^n} \preccurlyeq W^{0^m a^k 1^{n+1}} \preccurlyeq W^{1^m a^k 0^{n+1}} \preccurlyeq W^{1^m a^k 0^n}.$$

2) To prove $k_2 \leqslant k_3$, for any $m \geqslant 1$ and $n \geqslant 1$, we need to show

$$W^{0^m a^k 1^n} \preccurlyeq W^{1^m a^k 0^n} \Rightarrow W^{0^{m+1} a^k 1^{n+1}} \preccurlyeq W^{1^{m+1} a^k 0^{n+1}}.$$

Applying Theorem 2 to the assumption gives

$$W^{0^m a^k 0 1^n} \preccurlyeq W^{1^m a^k 0 0^n}. \tag{29}$$

Furthermore, by applying the PO of Theorem 1, we have

$$W^{0^{m+1} a^k 1^n} = W^{0^m 0 a^k 1^n} \preccurlyeq W^{0^m a^k 0 1^n}. \tag{30}$$

Combining (29) and (30) gives

$$W^{0^{m+1} a^k 1^n} \preccurlyeq W^{1^m a^k 0^{n+1}}. \tag{31}$$

By applying Lemma 6 to (31), we get

$$W^{0^{m+1} a^k 1^{n+1}} = W^{0^{m+1} a^k 1^n 1} \preccurlyeq W^{1^m a^k 0^{n+1} 1}. \tag{32}$$

Again, by using the PO, $W^{a^k 0^{n+1} 1} \preccurlyeq W^{1 a^k 0^{n+1}}$, so

$$W^{1^m a^k 0^{n+1} 1} \preccurlyeq W^{1^m 1 a^k 0^{n+1}} = W^{1^{m+1} a^k 0^{n+1}}. \tag{33}$$

Combining (32) and (33), we get

$$W^{0^{m+1} a^k 1^{n+1}} \preccurlyeq W^{1^{m+1} a^k 0^{n+1}}.$$

■

Remark 12. This structural property in Proposition 7 shows that for a fixed ϵ, $k^*_{min}(m, n, \epsilon)$ increases as m decreases or as n increases. This observation helps reduce the complexity of computing the values of k^*_{min}.

2) Pairs (m, n) for which $k^*_{min}(m, n, 0.5) = 0$: Notice that $k^*_{min}(m, n, 0.5) = 0$ means $W^{0^m 1^n} \preccurlyeq W^{1^m 0^n}$, i.e.,

$$Z(W^{0^m 1^n}) \geqslant Z(W^{1^m 0^n}) \tag{34}$$

when $W = \text{BEC}(0.5)$. In general, when $W = \text{BEC}(\epsilon)$, the Bhattacharyya parameters of these two bit-channels can be expressed as

$$Z(W^{0^m 1^n}) = f_1^{(n)}(f_0^{(m)}(\epsilon)) = [1 - (1-\epsilon)^{2^m}]^{2^n}, \tag{35}$$

$$Z(W^{1^m 0^n}) = f_0^{(n)}(f_1^{(m)}(\epsilon)) = 1 - (1 - \epsilon^{2^m})^{2^n}. \tag{36}$$

Combining (35) and (36), (34) translates to

$$T(\epsilon) \triangleq (1 - \epsilon^{2^m})^{2^n} + [1 - (1-\epsilon)^{2^m}]^{2^n} \geqslant 1. \tag{37}$$

Specializing to $\epsilon = 0.5$, the inequality yields an upper bound on n, namely

$$n \leqslant -\log_2[2^m - \log_2(2^{2^m} - 1)] \triangleq U(m). \tag{38}$$

The following proposition, proved in Appendix B, shows that $n = 2^m - 1$ is the largest value for which $k^*_{min}(m, n, 0.5) = 0$.

Proposition 8. For any $m \geqslant 1$,

$$2^m - 1 < U(m) < 2^m.$$

3) Range of ϵ where $k^*_{min}(m, n, \epsilon) = 0$: To explore the range of ϵ where $k^*_{min}(m, n, \epsilon) = 0$ for a given pair (m, n), we need to figure out when $T(\epsilon) \geqslant 1$. Referring to (37), we see that $T(\epsilon) = T(1-\epsilon)$ holds for any $\epsilon \in [0, 1]$ and any pair (m, n). Therefore, $T(\epsilon)$ is a symmetric function about the line $\epsilon = 0.5$ for any pair (m, n), so we only need to look at the behavior of $T(\epsilon)$ on the interval $(0, 0.5)$. The derivative of $T(\epsilon)$ is given in (39) and (40), as shown at the bottom of this page. Define $g(\epsilon, m) = \frac{1-(1-\epsilon)^{2^m}}{1-\epsilon^{2^m}}$. Here, we consider three different cases based on the value of the pair (m, n).

1) $n \leqslant m$: We make use of the following result.

Lemma 8. $1 > g(\epsilon, m+1) > g(\epsilon, m) > 0$, for any $\epsilon \in (0, 0.5)$ and any $m \geqslant 1$.

Proof: When $\epsilon \in (0, 0.5)$, we have

$$\begin{aligned} g(\epsilon, m+1) &= \frac{1 - (1-\epsilon)^{2^{m+1}}}{1 - \epsilon^{2^{m+1}}} \\ &= g(\epsilon, m) \cdot \frac{1 + (1-\epsilon)^{2^m}}{1 + \epsilon^{2^m}} \\ &> g(\epsilon, m). \end{aligned}$$

$$T'(\epsilon) = 2^{m+n}[(1 - (1-\epsilon)^{2^m})^{2^n - 1}(1-\epsilon)^{2^m - 1} - (1 - \epsilon^{2^m})^{2^n - 1}\epsilon^{2^m - 1}] \tag{39}$$

$$= 2^{m+n}[(\frac{1 - (1-\epsilon)^{2^m}}{1 - \epsilon^{2^m}})^{2^n - 1} - (\frac{\epsilon}{1-\epsilon})^{2^m - 1}](1-\epsilon)^{2^m - 1}(1 - \epsilon^{2^m})^{2^n - 1}. \tag{40}$$

It is also clear that $0 < g(\epsilon, m) < 1$ for any $\epsilon \in (0, 0.5)$ and any $m \geqslant 1$. ■

From Lemma 8, we have

$$[g(\epsilon, m)]^{2^n-1} \geqslant [g(\epsilon, m)]^{2^m-1} > [g(\epsilon, 0)]^{2^m-1}$$

for any $\epsilon \in (0, 0.5)$ and any $m \geqslant 1$. Referring to (40), we see that for $n \leqslant m$, $T'(\epsilon) > 0$ for any $\epsilon \in (0, 0.5)$. That is, $T(\epsilon)$ is increasing on $(0, 0.5)$. Noting the symmetry of $T(\epsilon)$, we conclude that

$$T(\epsilon) \geqslant T(0) = 1, \text{ for any } \epsilon \in [0, 1].$$

Note that this case also follows from Corollary 1.

2) $m < n \leqslant 2^m - 1$: We use the following lemma.

Lemma 9. For any $x \in (0, 1)$ and any $n \geqslant 0$, we have

$$1 - nx \leqslant (1 - x)^n \leqslant 1 - nx + \frac{n(n-1)}{2}x^2. \tag{41}$$

The lower bound is Bernoulli's inequality, and the upper bound follows from the Taylor series of $(1 - x)^n$. The proof is given in Appendix C.

Using the expression for $T(\epsilon)$ in (37) and Lemma 9, we get

$$T(\epsilon) \leqslant 1 - 2^n\epsilon^{2^m} + \frac{2^n(2^n-1)}{2}\epsilon^{2^{m+1}} + (2^m\epsilon)^{2^n} \tag{42}$$

$$\leqslant 1 - 2^n\epsilon^{2^m} + \frac{2^n(2^n-1)}{2}\epsilon^{2^{m+1}} + 2^{m2^n}\epsilon^{2^{m+1}} \tag{43}$$

$$= 1 + \epsilon^{2^m}[(\frac{2^n(2^n-1)}{2} + 2^{m2^n})\epsilon^{2^m} - 2^n], \tag{44}$$

where (43) follows from $n \geqslant m + 1$. By setting $(\frac{2^n(2^n-1)}{2} + 2^{m2^n})\epsilon^{2^m} - 2^n = 0$, we can find a value of ϵ, denoted by $\eta(m, n)$, such that $T(\eta(m, n)) \leqslant 1$. Then, for any $\epsilon \in (0, \eta(m, n))$, we have $(\frac{2^n(2^n-1)}{2} + 2^{m2^n})\epsilon^{2^m} - 2^n < 0$, which implies $T(\epsilon) < 1$. The upper bound $2^m - 1$ on n guarantees that $T(0.5) > 1$, so $T(\epsilon) = 1$ has at least one solution in $[\eta(m, n), 0.5)$. Denoting the largest one by $\epsilon^*_{max}(m, n)$, we conclude that

$$T(\epsilon) \geqslant 1, \text{ for any } \epsilon \in [\epsilon^*_{max}(m, n), 1 - \epsilon^*_{max}(m, n)].$$

3) $n \geqslant 2^m$: For any $\epsilon \in (0, 0.5)$, if $T(\epsilon) < 1$ for $n = 2^m$, then $T(\epsilon) < 1$ for $n > 2^m$ since $T(\epsilon)$ decreases as n increases for a fixed m. For $m = 1$, 2, 3, and $n = 2^m$, a computer search confirms that the only real roots of $T'(\epsilon) = 0$ occur at $\epsilon = 0$, 0.5, and 1. This implies that $T(\epsilon) < 1$, for any $\epsilon \in (0, 1)$, as can be seen in the corresponding curves in Figure 3, where the curves progress from bottom to top as m increases. Consequently, for $m = 1$, 2, 3, and any $\epsilon \in (0, 1)$, we have

$$W^{0^m1^{2^m}} \succcurlyeq W^{1^m0^{2^m}}. \tag{45}$$

We conjecture that this is also true for $m \geqslant 4$.

We can analyze the asymptotic behavior of bit-channels $W^{0^m1^{2^m}}$ and $W^{1^m0^{2^m}}$ based on the following lemma, whose proof is given in Appendix D.

Lemma 10.

$$\lim_{n\to\infty}(1 - \epsilon^n)^{2^n} = \begin{cases} 1, & \epsilon \in [0, 0.5), \\ e^{-1}, & \epsilon = 0.5, \\ 0, & \epsilon \in (0.5, 1]. \end{cases} \tag{46}$$

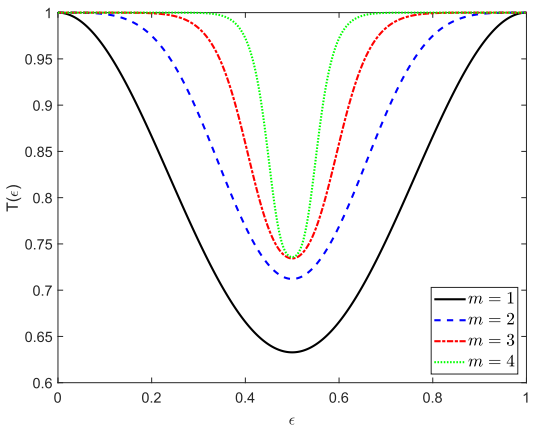


Fig. 3. The function $T(\epsilon)$ when $m = 1, 2, 3, 4$ and $n = 2^m$.

Combining (35), (36) and Lemma 10, we find the asymptotic values of the Bhattacharyya parameters for the two bit-channels:

$$\left(Z(W^{0^m1^{2^m}}), Z(W^{1^m0^{2^m}})\right) \xrightarrow{m\to\infty} \begin{cases} (0, 0), & \epsilon \in [0, 0.5), \\ (e^{-1}, 1 - e^{-1}), & \epsilon = 0.5, \\ (1, 1), & \epsilon \in (0.5, 1]. \end{cases}$$

Remark 13. For a fixed m, we note that $W^{0^m1^n}$ becomes more reliable as n increases, converging to a perfect channel as n goes to infinity. In contrast, $W^{1^m0^n}$ does the opposite, converging to a completely noisy channels as n goes to infinity. When $W = \text{BEC}(\epsilon)$, the channel upgradation relation in (45) provides the cut-off value 2^m for the smallest possible n such that $W^{0^m1^n} \succcurlyeq W^{1^m0^n}$ for any $\epsilon \in (0, 1)$. Furthermore, we note that for the bit-channels represented by binary sequences with the length $m + 2^m$, $W^{0^m1^{2^m}}$ is the worst one among the bit-channels with Hamming weight greater than or equal to 2^m, while $W^{1^m0^{2^m}}$ is the best one among the bit-channels with Hamming weight less than or equal to m. Setting $t = m + 2^m$, we consider the sets

$$S_1 = \{b^t \in \{0, 1\}^t | \text{wt}(b^t) \leqslant m\}$$

and

$$S_2 = \{b^t \in \{0, 1\}^t | \text{wt}(b^t) \geqslant 2^m\}.$$

When the channel upgradation relation in (45) holds, it provides a PO that applies to all pairs of bit-channels with one belonging to S_1 and the other belonging to S_2. That is, for any $b^t_1 \in S_1$ and any $b^t_2 \in S_2$, we have

$$W^{b^t_2} \succcurlyeq W^{b^t_1}.$$

We have seen that (45) holds for $m = 1, 2, 3$. If, as conjectured, (45) holds for $m \geqslant 4$, this PO would extend to all of the code block lengths with the form $t = m + 2^m$.

D. Universal Positions

Recall from Figure 1 that several bit-channels for the BEC have universal positions with respect to bit-channel ordering in the sense that their positions in the complete ordering of

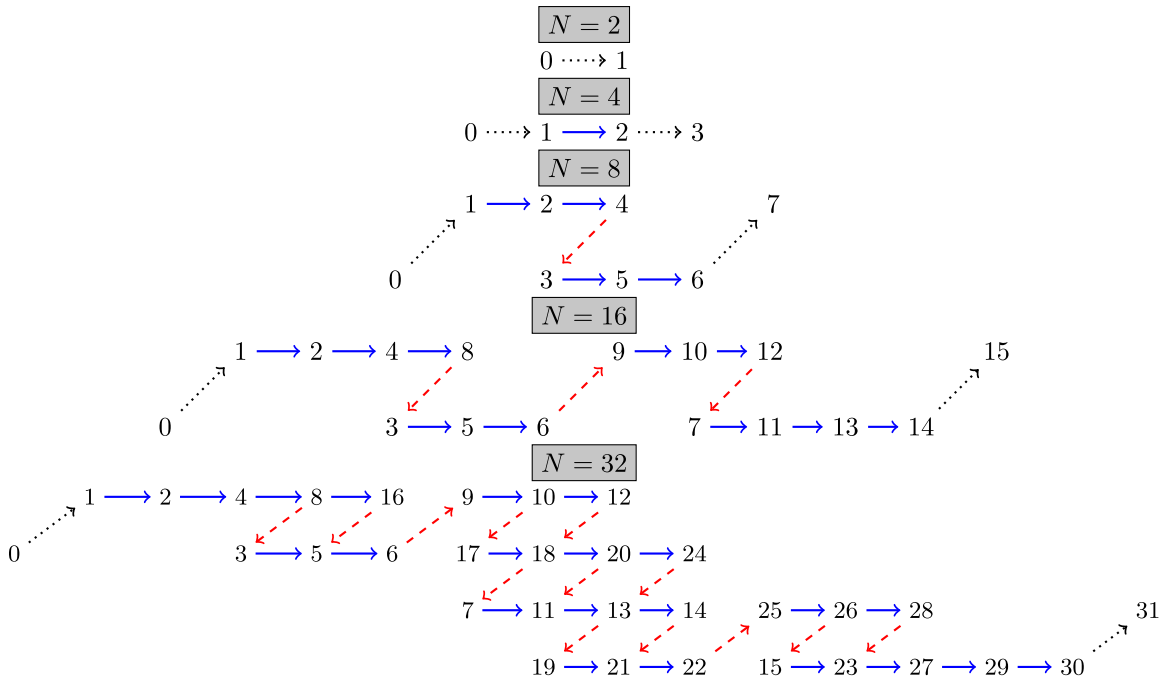


Fig. 4. Universal POs for the BEC when $N = 2, 4, 8, 16, 32$.

the 64 bit-channels are independent of the channel erasure probability. In other words, the bit-channels we show in Figure 1 are universally comparable with each of the other 63 bit-channels. If code rates are suitably chosen, we can have universal information sets which are independent of channel erasure probability.

Definition 3. Consider $W = \text{BEC}(\epsilon)$. Define the *universal set* $U_N \subseteq [N]$ to be the indices $i \in [N]$ such that for any $j \in [N] \setminus \{i\}$, either $W_N^{(i)} \preccurlyeq W_N^{(j)}$ for all $\epsilon \in [0, 1]$ or $W_N^{(i)} \succcurlyeq W_N^{(j)}$ for all $\epsilon \in [0, 1]$. (By Lemma 3, this property of bit-channels can be equivalently defined in terms of the Bhattacharyya parameter, error probability, or symmetric capacity.)

The following proposition presents two properties of U_N, which are useful in determining U_N.

Proposition 9.

1) *Symmetry*: If $i \in U_N$, then $N - 1 - i \in U_N$;
2) *Nesting*: $U_{2N} \cap [N] \subseteq U_N$.

Proof: Consider $W = \text{BEC}(\epsilon)$.

1) For $i \in U_N$, assume i has binary expansion $b^t = b_1\ b_2 \cdots b_t$, $t = \log_2 N$. Then, $N - 1 - i$ has binary expansion $\bar{b}^t$. Indeed, $(W_N^{(i)}, W_N^{(N-1-i)})$ is a complementary bit-channel pair. According to Corollary 4, we have

$$Z(W^{b^t}) + Z((W^c)^{\bar{b}^t}) = 1,$$

that is,

$$Z(W_N^{(i)}) + Z((W^c)_N^{(N-1-i)}) = 1, \tag{47}$$

where $W^c = \text{BEC}(1 - \epsilon)$. Universality of i implies that, for any $j \in [N] \setminus \{i\}$, $W_N^{(i)}$ is universally comparable with $W_N^{(j)}$. Therefore, according to (47), $(W^c)_N^{(N-1-i)}$ is also universally comparable with $(W^c)_N^{(N-1-j)}$. Since the erasure probability $1 - \epsilon$ of W^c ranges over the whole interval $[0, 1]$ when ϵ does, the index $N - 1 - i$ satisfies the condition presented in Definition 3 and thus $N - 1 - i \in U_N$.

2) Consider any $i \in U_{2N} \cap [N]$. The index i has binary expansion $0b^t$, where $t = \log_2 N$. Now, universality of i implies that $W_{2N}^{(i)}$ is universally comparable with $W_{2N}^{(j)}$ for any $j \in [2N] \setminus \{i\}$. In particular, if j is in the smaller half of $[2N]$, its binary expansion will have the form $0a^t$ for some $a^t \in \{0, 1\}^t$, implying that W^{0b^t} is universally comparable with W^{0a^t}. Setting $V = W^0$, we can restate this property as V^{b^t} is universally comparable with V^{a^t}, for any $a^t \in \{0, 1\}^t$. Noting that $V = W^0 = \text{BEC}(2\epsilon - \epsilon^2)$, we see that the erasure probability $2\epsilon - \epsilon^2$ of V ranges over the whole interval $[0, 1]$ when ϵ does, which implies that $W_N^{(i)} = W^{b^t}$ is universally comparable with $W_N^{(j)} = W^{a^t}$ for any $j \in [N] \setminus \{i\}$. We conclude that $i \in U_N$. ∎

In Figure 4, we show universal PO diagrams for the BEC when $N = 2, 4, 8, 16, 32$. The diagrams use the following conventions.

- Each number i represents the bit-channel $W_N^{(i)}$.
- The directed arrow from i to j between any two nodes indicates that $W_N^{(i)} \preccurlyeq W_N^{(j)}$.
- Every PO is universal, which means the ordering is independent of channel erasure probability.
- The dotted arrows are based on the PO in [19].
- The solid arrows are based on the PO in [28, Theorem 1] and [3].
- The dashed arrows are additional POs for the BEC beyond the existing POs.

Remark 14. Because of the explicit recursion for polarizing $W = \text{BEC}(\epsilon)$, we can express the Bhattacharyya parameter of each bit-channel as a polynomial function of ϵ. The orderings

TABLE III
UNIVERSAL SETS FOR BLOCK LENGTHS $N = 2^t$, $1 \leqslant t \leqslant 7$

N	U_N
2	$\{0, 1\}$
4	$\{0, 1, 2, 3\}$
8	$\{0, 1, 2, \ldots, 7\}$
16	$\{0, 1, 2, \ldots, 15\}$
32	$\{0, 1, 2, 4, 5, 6, 8, 9, 10, 13, 18, 21, 22, 23, 25, 26, 27, 29, 30, 31\}$
64	$\{0, 1, 2, 4, 8, 9, 10, 18, 45, 53, 54, 55, 59, 61, 62, 63\}$
128	$\{0, 1, 2, 4, 8, 119, 123, 125, 126, 127\}$

denoted by dashed arrows in Figure 4 are obtained by comparing the polynomial functions of ϵ associated with the corresponding bit-channels on $\epsilon \in [0, 1]$.

The universal sets for block lengths $N = 2^t$, $1 \leqslant t \leqslant 7$, are shown in Table III. The results for $N = 2, 4, 8, 16, 32$ are derived from Figure 4, the result for $N = 64$ is derived from Figure 1, and the result for $N = 128$ follows from the following proposition, which identifies the universal sets U_N for block lengths $N = 2^t$, $t \geqslant 7$.

Proposition 10. The universal set for $N = 2^t$, $t \geqslant 7$, is given by $U_N = \{0, 1, 2, 4, 8, N-9, N-5, N-3, N-2, N-1\}$.

Proof: Let $W = \text{BEC}(\epsilon)$. Consider the case where $N = 2^7 = 128$. After comparing the Bhattacharyya parameters of the corresponding bit-channels, we find that the orderings of the following pairs are not universal:
$(W_{128}^{(9)}, W_{128}^{(64)})$, $(W_{128}^{(10)}, W_{128}^{(64)})$, $(W_{128}^{(18)}, W_{128}^{(64)})$,
$(W_{128}^{(45)}, W_{128}^{(104)})$, $(W_{128}^{(53)}, W_{128}^{(104)})$, $(W_{128}^{(54)}, W_{128}^{(104)})$,
$(W_{128}^{(55)}, W_{128}^{(120)})$, $(W_{128}^{(59)}, W_{128}^{(120)})$, $(W_{128}^{(61)}, W_{128}^{(120)})$, and
$(W_{128}^{(62)}, W_{128}^{(120)})$. For example, to compare the first pair we can write

$$Z(W_{128}^{(9)}) = Z(W^{0001001}) = f_1 \circ f_0 \circ f_0 \circ f_1 \circ f_0 \circ f_0 \circ f_0(\epsilon), \tag{48}$$

$$Z(W_{128}^{(64)}) = Z(W^{1000000}) = f_0 \circ f_0 \circ f_0 \circ f_0 \circ f_0 \circ f_0 \circ f_1(\epsilon). \tag{49}$$

One of the intersection points of the curves corresponding to polynomial functions (48) and (49) is at $\epsilon = 0.0509$, which means the ordering of $(W_{128}^{(9)}, W_{128}^{(64)})$ depends on the value of ϵ. Similar arguments can be applied for the other pairs. Therefore, the indices 9, 10, 18, 45, 53, 54, 55, 59, 61, and 62 are not included in U_{128}. Note that the ordering of $(W_{128}^{(9)}, W_{128}^{(64)})$ is not universal, which also implies that 64 is not included in U_{128}. According to the symmetry property of U_N, 63 is excluded.

Next, we will show that the indices 0, 1, 2, 4, 8 are included in U_N for any $t \geqslant 4$. Referring to the universal PO diagram for $N = 16$ in Figure 4, we have the universal orderings $0 \to 1 \to 2 \to 4 \to 8 \to 3$, which translate to

$$W^{0000} \preccurlyeq W^{0001} \preccurlyeq W^{0010} \preccurlyeq W^{0100} \preccurlyeq W^{1000} \preccurlyeq W^{0011}. \tag{50}$$

Because of the universality, we can substitute $W^{0^{t-4}}$ for W in (50) and the orderings are still preserved. When $N = 2^t$, $t \geqslant 4$, by translating back to bit-channel indices, we get the universal orderings

$$W_N^{(0)} \preccurlyeq W_N^{(1)} \preccurlyeq W_N^{(2)} \preccurlyeq W_N^{(4)} \preccurlyeq W_N^{(8)} \preccurlyeq W_N^{(3)}.$$

In fact, for a fixed N, $W_N^{(3)}$ is the worst bit-channel among those with Hamming weight at least 2. In addition, among the bit-channels with Hamming weight 1, $W_N^{(1)}$, $W_N^{(2)}$, $W_N^{(4)}$, $W_N^{(8)}$ are the worst four bit-channels. Therefore, we conclude that, for any $j \in [N] \setminus \{0, 1, 2, 4, 8\}$,

$$W_N^{(0)} \preccurlyeq W_N^{(1)} \preccurlyeq W_N^{(2)} \preccurlyeq W_N^{(4)} \preccurlyeq W_N^{(8)} \preccurlyeq W_N^{(j)}$$

holds for any $t \geqslant 4$. Thus, 0, 1, 2, 4, 8 are included in U_N for any $t \geqslant 4$.

We also need to show that $N-9$, $N-5$, $N-3$, $N-2$, and $N-1$ are not included in U_{2N} for $t \geqslant 7$. When $N = 2^t$, the binary expansions of the above five indices are $1^{t-4}0111$, $1^{t-4}1011$, $1^{t-4}1101$, $1^{t-4}1110$, $1^{t-4}1111$. To consider whether these indices are included in U_{2N}, we extend the corresponding binary expansions by adding 0 in the most significant digit position. For example, the first one becomes $01^{t-4}0111$. According to Table I (or Fig. 2), we know that $(W_{32}^{(14)}, W_{32}^{(19)})$ and $(W_{32}^{(15)}, W_{32}^{(28)})$, i.e., (W^{01110}, W^{10011}) and (W^{01111}, W^{11100}), are not universally ordered pairs. Based on the orderings of these two pairs, we conclude that the following pairs are not universally ordered:

$$\begin{aligned}
&(W^{01^{t-4}0111} = W^{01110111}, W^{10011111}), && t = 7,\\
&(W^{01^{t-4}0111} = W^{011111^{t-8}0111}, W^{111001^{t-8}0111}), && t \geqslant 8,\\
&(W^{01^{t-4}1011} = W^{011111^{t-7}011}, W^{111001^{t-7}011}), && t \geqslant 7,\\
&(W^{01^{t-4}1101} = W^{011111^{t-7}101}, W^{111001^{t-7}101}), && t \geqslant 7,\\
&(W^{01^{t-4}1110} = W^{011111^{t-7}110}, W^{111001^{t-7}110}), && t \geqslant 7,\\
&(W^{01^{t-4}1111} = W^{011111^{t-7}111}, W^{111001^{t-7}111}), && t \geqslant 7.
\end{aligned}$$

Therefore, $N-9$, $N-5$, $N-3$, $N-2$, and $N-1$ are not included in U_{2N} for any $t \geqslant 7$. The proposition follows from the symmetry property of U_N presented in Proposition 9. ■

Remark 15. We note that the set inclusion in Proposition 9, part 2), is an equality for $N = 2, 4, 8$. It is a proper inclusion for $N = 2^t$, $t \geqslant 4$, with the intersection exactly equal to $\{0, 1, 2, 4, 8\}$ for $t \geqslant 6$.

Remark 16. For any index i in the set U_N, the ordering position of $W_N^{(i)}$ is called universal if it remains the same for any $\epsilon \in [0, 1]$. In other words, there is a nonzero code rate associated with each index $i \in U_N$ that produces a universal information set (with i included) when constructing a polar code with the block length N. The number of these code rates is equal to the size of U_N. When $N = 2, 4, 8, 16$, all the bit-channels are universally ordered for each block length, respectively. So each nonzero code rate for these block lengths produces a universal information set. When $N = 32$ and 64, there are 20 and 16 code rates with universal information sets, respectively. As an example, rate $\frac{22}{32}$ is associated with

index 10 when $N = 32$. In Figure 4, the diagram for $N = 32$ could be divided into two groups if a boundary is chosen between index 9 and 10; the frozen set consists of the indices in the left group and the information set consists of those in the right group. For block lengths $N = 2^t$, $t \geqslant 7$, Proposition 10 implies that there are 10 code rates with universal information sets; these rates are $\frac{k}{N}$ or $\frac{N-k}{N}$, where $k = 1, 2, 3, 4, 5$.

V. Threshold Behavior

In Section IV-D, we obtained several results relating to BEC bit-channel orderings by explicitly comparing the Bhattacharyya parameters of the corresponding bit-channels for short code block lengths. However, for larger block lengths, the computations become complex because of the exponentially growing degrees of the corresponding polynomials. It therefore becomes impractical to obtain additional POs by directly comparing the Bhattacharyya parameters. According to [20], each of the bit-channels has a sharp threshold for sufficiently large block length, in the sense that the function (8), whose asymptotic behavior is shown in (25), converges to a step function as the block length grows. This threshold value can be used to determine whether a bit-channel is good or bad when constructing a polar code for a BEC [7, Proposition 2].

In this section, we first explore the threshold behavior of bit-channels when the underlying channel is a BEC by considering the half-way point for $b^t \in \{0,1\}^t$, defined as follows.

Definition 4. For any $b^t \in \{0,1\}^t$, $t \geqslant 1$, there exists a *half-way* point, denoted by α_{b^t}, such that

$$F_{b^t}(\alpha_{b^t}) = \frac{1}{2}.$$

Since $f_0(x)$ and $f_1(x)$ are strictly increasing on $[0,1]$, it follows that so is $F_{b^t}(x)$ for any $b^t \in \{0,1\}^t$. Therefore, the half-way point α_{b^t} is unique for any $b^t \in \{0,1\}^t$.

Then, for general BMS channels, we combine the bounds of the bit-channel Bhattacharyya parameter shown in (10) with Lemma 5 and its auxiliary result to derive the relationship between the corresponding half-way points of the lower and upper bound (achieved by the BEC).

A. BEC

Recall $f_0(x) = 2x - x^2$, $f_1(x) = x^2$, and $F_{b^t}(x) = f_{b_t} \circ \cdots \circ f_{b_2} \circ f_{b_1}(x)$, for $b^t \in \{0,1\}^t$, $t \geqslant 1$. Assume $t = m + n$. Among all $b^t \in \{0,1\}^t$ with the same Hamming weight n, the value of α_{b^t} is minimized for

$$b^* = \underbrace{00\cdots0}_{m}\underbrace{11\cdots1}_{n} = 0^m 1^n.$$

According to (36), when $W = \mathrm{BEC}(\epsilon)$, we have

$$Z(W^{b^*}) = F_{b^*}(\epsilon) = F_{0^m 1^n}(\epsilon) = [1 - (1-\epsilon)^{2^m}]^{2^n}.$$

We know that the bit-channel W^{b^*} has a sharp threshold for sufficiently large t. In [20], the threshold behavior of the bit-channel W^{b^*} was analyzed at $\epsilon = \alpha_{b^*}$ as t goes to infinity. Here, we will give further results about the value of α_{b^*}.

For a constant $c \in [0,1]$, if there exists a positive and vanishing sequence $\{\delta(t)\}_{t=1}^{\infty}$ such that

$$F_{b^*}(c) \geqslant \frac{1}{2} \quad \text{and} \quad F_{b^*}(c - \delta(t)) \leqslant \frac{1}{2},$$

then $\alpha_{b^*} = c - o(1)$. Similarly, if the conditions become

$$F_{b^*}(c) \leqslant \frac{1}{2} \quad \text{and} \quad F_{b^*}(c + \delta(t)) \geqslant \frac{1}{2},$$

then $\alpha_{b^*} = c + o(1)$, where $o(1)$ stands for a vanishing positive number as t goes to infinity.

If we consider n to be a function of m, then $t = m + n$ can also be regarded as a function of m. Finding a sequence $\{\delta(t)\}_{t=1}^{\infty}$ translates to finding a sequence $\{\delta(m)\}_{m=1}^{\infty}$. The following proposition analyzes four cases, based on the functional dependence of n on m, denoted by $n(m)$.

Proposition 11.

1) When $\lim_{m\to\infty} \frac{n(m)}{2^m} = 0$, $\alpha_{b^*} = o(1)$.
2) When $\lim_{m\to\infty} \frac{n(m)}{2^m} \nearrow 1$ and $n(m) < 2^m$, $\alpha_{b^*} = \frac{1}{2} - o(1)$.
3) When $\lim_{m\to\infty} \frac{n(m)}{2^m} \searrow 1$ and $n(m) \geqslant 2^m$, $\alpha_{b^*} = \frac{1}{2} + o(1)$.
4) When $\lim_{m\to\infty} \frac{n(m)}{2^m} = \infty$, $\alpha_{b^*} = 1 - o(1)$.

The proof is given in Appendix E. We show several examples illustrating the half-way points α_{b^*} in Figure 5. Figure 5(a) considers the case where n varies as a function of m for $m = 5$. We see that for a fixed m, the value of the half-way point α_{b^*} increases as n increases, yielding a better bit-channel $W^{0^m 1^n}$. Figure 5(b) shows the results for $n = 2^m$ when $m = 1, 2, 3, 4, 5$. This is a special case of $n(m) = 2^m$ in the third scenario of Proposition 11. In Figure 5(b), the steepness of the curves increases as m increases.

Remark 17. The proposition generalizes the results about the half-way point in [20, Section II]. In particular, it was stated in [20] that $\alpha_{b^*} = \frac{1}{2} - o(1)$ when $m = \log_2 t$, which is a special case of $n(m) = 2^m - m$ in the second scenario of Proposition 11. It was also stated in [20] that $\alpha_{b^*} = 1 - o(1)$ when $m = \log_2 t - \log_2 \log_2 t$. If $t = 2^{2^k}$, $k \geqslant 0$, consider $m = \log_2 t - \log_2 \log_2 t = 2^k - k$, so that $n = t - m = 2^{2^k} - 2^k + k$. These values satisfy

$$2^m m - m < n(m) < 2^m \cdot 2m - m,$$

which is included in the fourth scenario of Proposition 11.

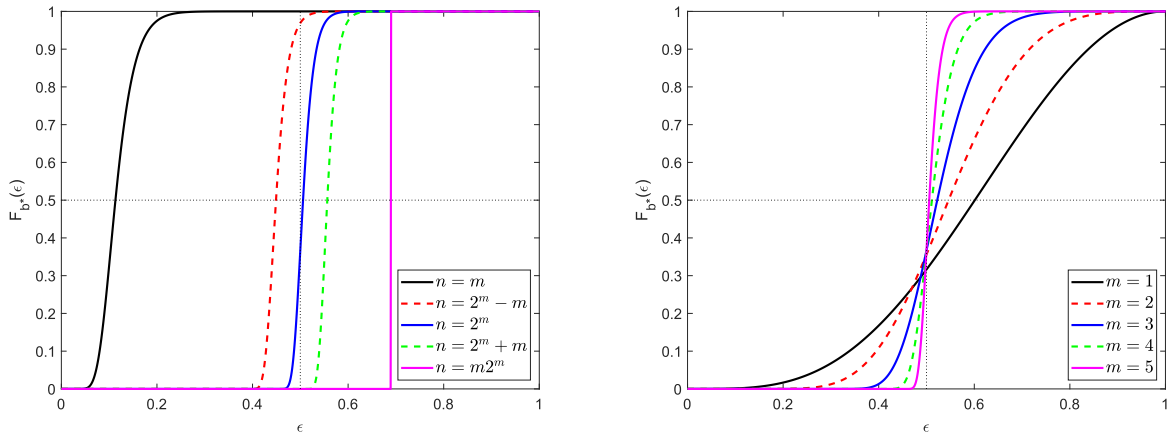
Remark 18. By the symmetry shown in Corollary 4, we get $\alpha_{1^m 0^n} = 1 - \alpha_{b^*}$. When $n = 2^m$,

$$\alpha_{1^m 0^{2^m}} = \frac{1}{2} - o(1) < \frac{1}{2} + o(1) = \alpha_{0^m 1^{2^m}},$$

which supports our conjecture about (45).

Assuming $N = 2^t$, we can deduce from Proposition 11 several results about the bit-channel $W^{0^m 1^n}$ when $W = \mathrm{BEC}(\epsilon)$.

1) When $\lim_{m\to\infty} \frac{n(m)}{2^m} = 0$, $\lim_{m\to\infty} Z(W^{0^m 1^{n(m)}}) = 1$ for any $\epsilon \in (0,1)$, i.e., $W^{0^m 1^{n(m)}}$ converges to a bit-channel in $\mathcal{B}_N^{\varepsilon}(W)$ for any $\epsilon \in (0,1)$.



(a) (b)

Fig. 5. Examples of Proposition 11: (a) The half-way points α_{b^*} when n varies as a function of m for $m = 5$. (b) The half-way points α_{b^*} when $n = 2^m$ for $m = 1, 2, 3, 4, 5$.

2) When $\lim_{m\to\infty} \frac{n(m)}{2^m} \nearrow 1$ and $n(m) < 2^m$, $\lim_{m\to\infty} Z(W^{0^m 1^{n(m)}}) = 1$ for any $\epsilon \in [\frac{1}{2}, 1)$, i.e., $W^{0^m 1^{n(m)}}$ converges to a bit-channel in $\mathcal{B}_N^{\varepsilon}(W)$ for any $\epsilon \in [\frac{1}{2}, 1)$.
3) When $\lim_{m\to\infty} \frac{n(m)}{2^m} \searrow 1$ and $n(m) \geqslant 2^m$, $\lim_{m\to\infty} Z(W^{0^m 1^{n(m)}}) = 0$ for any $\epsilon \in (0, \frac{1}{2}]$, i.e., $W^{0^m 1^{n(m)}}$ converges to a bit-channel in $\mathcal{G}_N^{\varepsilon}(W)$ for any $\epsilon \in (0, \frac{1}{2}]$.
4) When $\lim_{m\to\infty} \frac{n(m)}{2^m} = \infty$, $\lim_{m\to\infty} Z(W^{0^m 1^{n(m)}}) = 0$ for any $\epsilon \in (0, 1)$, i.e., $W^{0^m 1^{n(m)}}$ converges to a bit-channel in $\mathcal{G}_N^{\varepsilon}(W)$ for any $\epsilon \in (0, 1)$.

B. BMS Channels

Given a BMS channel with $Z(W) = z$, according to (10), for any $b^t \in \{0, 1\}^t$, we have

$$G_{b^t}(z) \leqslant Z(W^{b^t}) \leqslant F_{b^t}(z).$$

From Lemma 5, this is equivalent to

$$\sqrt{F_{b^t}(z^2)} \leqslant Z(W^{b^t}) \leqslant F_{b^t}(z).$$

There is an auxiliary result of Lemma 5. Let $(a_\ell)_{\ell \geqslant 1}$ be a sequence of i.i.d. Bernoulli($\frac{1}{2}$) random variables. For a realization a^k of the random vector $a_1 a_2 \cdots a_k$, the function F_{a^k} has the threshold point $\epsilon^* \in [0, 1]$ as shown in (25). Then, with the same realization a^k, G_{a^k} has a larger threshold point $\sqrt{\epsilon^*}$. Namely,

$$\lim_{k\to\infty} G_{a^k}(z) = \lim_{k\to\infty} \sqrt{F_{a^k}(z^2)} = \begin{cases} 0, & z \in [0, \sqrt{\epsilon^*}), \\ 1, & z \in (\sqrt{\epsilon^*}, 1]. \end{cases} \tag{51}$$

Definition 5. For any $b^t \in \{0, 1\}^t$, $t \geqslant 1$, there exists a *half-way* point, denoted by β_{b^t}, such that

$$G_{b^t}(\beta_{b^t}) = \frac{1}{2}.$$

Definition 6. For any $b^t \in \{0, 1\}^t$, $t \geqslant 1$, there exists a *quarter-way* point, denoted by γ_{b^t}, such that

$$F_{b^t}(\gamma_{b^t}) = \frac{1}{4}.$$

Applying an argument similar to that used for α_{b^t}, we see that both β_{b^t} and γ_{b^t} are unique for any $b^t \in \{0, 1\}^t$. In fact, $\beta_{b^t} = \sqrt{\gamma_{b^t}}$ holds for any $b^t \in \{0, 1\}^t$, $t \geqslant 1$. In the asymptotic regime, by the auxiliary result (51) of Lemma 5, we have

$$\lim_{t\to\infty} \beta_{b^t} = \lim_{t\to\infty} \sqrt{\alpha_{b^t}},$$

for the same realization b^t on both sides.

VI. Conclusion

We give an alternative proof based on mathematical induction for an existing PO proposed in [28] and [3]. Then, we use the proof idea to deduce further generalized POs. Since some of the POs rely on verifying degradation relations, which can be difficult for general BMS channels, we also give several examples of ordering relations based upon the Bhattacharyya parameter and error probability. The Bhattacharyya parameters of complementary bit-channel pairs on the BEC exhibit a symmetry property, whose implications we investigate, along with alignment properties of polarized sets for BEC and BMS channels. We illustrate the new POs for BEC bit-channels with different Hamming weights using numerical examples. Several further properties of ordering relations for particular classes of bit-channels are analyzed. The bit-channels with universal ordering positions, which are independent of the channel erasure probability, are characterized using symmetry and nesting properties. Finally, we approximate the half-way point of the Bhattacharyya parameter for the worst bit-channel among those with the same Hamming weight for the BEC and derive the relationship between the half-way points of the lower and upper bounds of the bit-channel Bhattacharyya parameter for general BMS channels.

There are several directions for future research on POs. In particular, since most of the work on POs concerns binary polar codes with Arıkan's 2×2 kernel for BMS channels, it would be of interest to explore possible POs in other scenarios, including non-binary inputs [21], [25], [27], larger kernels [4], [6], [15], [22], and asymmetric channels [12], [16].

APPENDIX A
PROOF OF LEMMA 5

Proof: When $k = 1$, there are two cases to consider for any $x \in [0, 1]$.

1) If $a_1 = 0$, then

$$G^2_{a_1}(x) = f^2_{-1}(x) = 2x^2 - x^4,$$
$$F_{a_1}(x^2) = f_0(x^2) = 2x^2 - x^4.$$

2) If $a_1 = 1$, then

$$G^2_{a_1}(x) = f^2_1(x) = x^4,$$
$$F_{a_1}(x^2) = f_1(x^2) = x^4.$$

Therefore, (11) is true for any $x \in [0, 1]$ when $k = 1$.

Assume it is also true when $k = s$, $s \geqslant 1$, i.e.,

$$G^2_{a^s}(x) = F_{a^s}(x^2)$$

holds for any $a^s \in \{0, 1\}^s$ and $x \in [0, 1]$.

Then, when $k = s + 1$, there are two cases to consider.

1) If $a_{s+1} = 0$, then

$$G^2_{a^{s+1}}(x) = [f_{-1}(G_{a^s}(x))]^2 = 2G^2_{a^s}(x) - G^4_{a^s}(x),$$
$$F_{a^{s+1}}(x^2) = f_0(F_{a^s}(x^2)) = 2F_{a^s}(x^2) - F^2_{a^s}(x^2).$$

2) If $a_{s+1} = 1$, then

$$G^2_{a^{s+1}}(x) = [f_1(G_{a^s}(x))]^2 = G^4_{a^s}(x),$$
$$F_{a^{s+1}}(x^2) = f_1(F_{a^s}(x^2)) = F^2_{a^s}(x^2).$$

Because of the assumption, we have

$$G^2_{a^{s+1}}(x) = F_{a^{s+1}}(x^2),$$

for any $a^{s+1} \in \{0, 1\}^{s+1}$ and $x \in [0, 1]$.

The proof is complete. ∎

APPENDIX B
PROOF OF PROPOSITION 8

Proof: Referring to (38), we can rewrite $U(m)$ as

$$U(m) = -\log_2\left[-\log_2\left(1 - \frac{1}{2^{2^m}}\right)\right]. \tag{52}$$

Set $r(m) = \frac{1}{2^{2^m}}$. Then, from the Taylor series of $\ln(1 - x)$, we have

$$-\log_2(1 - r(m)) = \frac{1}{\ln 2}\sum_{i=1}^{\infty}\frac{r(m)^i}{i}. \tag{53}$$

The summation in (53) can be bounded from below and above, respectively, by

$$\sum_{i=1}^{\infty}\frac{r(m)^i}{i} > r(m), \tag{54}$$

and

$$\sum_{i=1}^{\infty}\frac{r(m)^i}{i} < \sum_{i=1}^{\infty} r(m)^i = \frac{r(m)}{1 - r(m)}. \tag{55}$$

Applying (53) and (54) in (52), we obtain the upper bound

$$\begin{aligned} U(m) &< -\log_2\left[\frac{1}{\ln 2}\cdot r(m)\right] \\ &= 2^m + \log_2 \ln 2 \\ &< 2^m. \end{aligned}$$

On the other hand, applying (53) and (55) in (52), we obtain the lower bound

$$U(m) > -\log_2\left[\frac{1}{\ln 2}\cdot\frac{r(m)}{1 - r(m)}\right] \tag{56}$$
$$= \log_2\left[(2^{2^m} - 1)\cdot \ln 2\right] \tag{57}$$
$$> \log_2(2^{2^m - 1}) \tag{58}$$
$$= 2^m - 1. \tag{59}$$

The inequality in (58) is based on the following lemma.

Lemma 11. For any $m \geqslant 1$,

$$(2^{2^m} - 1)\cdot \ln 2 > 2^{2^m - 1}. \tag{60}$$

Inequality (60) is equivalent to

$$1 - \frac{1}{2\ln 2} > \frac{1}{2^{2^m}}. \tag{61}$$

As the right-hand side of (61) decreases as m increases, it suffices to check the case $m = 1$, which can be verified numerically.

The proof is complete. ∎

APPENDIX C
PROOF OF LEMMA 9

Proof: The proof is by induction.

For $n = 0$, all three polynomials in (41) are equal to 1, so the statement holds.

Let $k \geqslant 0$ be given and suppose (41) is true for $n = k$. On the one hand, we have

$$\begin{aligned} (1 - x)^{k+1} &= (1 - x)^k(1 - x) \\ &\geqslant (1 - kx)(1 - x) \\ &= 1 - (k + 1)x + kx^2 \\ &\geqslant 1 - (k + 1)x. \end{aligned}$$

One the other hand, we have

$$\begin{aligned} (1 - x)^{k+1} &= (1 - x)^k(1 - x) \\ &\leqslant \left[1 - kx + \frac{k(k-1)}{2}x^2\right](1 - x) \\ &= 1 - (k + 1)x + \frac{k(k+1)}{2}x^2 - \frac{k(k-1)}{2}x^3 \\ &\leqslant 1 - (k + 1)x + \frac{k(k+1)}{2}x^2. \end{aligned}$$

Hence (41) holds for $n = k+1$, and the proof of the induction step is complete. ∎

APPENDIX D
PROOF OF LEMMA 10

Proof: We use the following well-known inequality:

$$(1+x)^r \leqslant e^{rx}, \text{ for any } x > -1 \text{ and } r > 0. \tag{62}$$

According to (62), we have

$$(1-\epsilon^n)^{2^n} \leqslant e^{-(2\epsilon)^n}. \tag{63}$$

When $\epsilon \in [0, 0.5)$, lower and upper bounds on the limit in (46) follow from Lemma 9 and (63), respectively; namely, we have

$$1-(2\epsilon)^n \leqslant (1-\epsilon^n)^{2^n} \leqslant e^{-(2\epsilon)^n}. \tag{64}$$

From (64), it follows that $\lim_{n\to\infty}(1-\epsilon^n)^{2^n} = 1$.

When $\epsilon = 0.5$, we use the expression $\lim_{x\to 0}(1+x)^{\frac{1}{x}} = e$ to obtain $\lim_{n\to\infty}(1-\frac{1}{2^n})^{2^n} = e^{-1}$.

When $\epsilon \in (0.5, 1]$, the upper bound in (63) implies $\lim_{n\to\infty}(1-\epsilon^n)^{2^n} = 0$. ∎

APPENDIX E
PROOF OF PROPOSITION 11

Proof: Combining (41) and (62) gives

$$1-nx \leqslant (1-x)^n \leqslant e^{-nx},$$

for any $n \geqslant 0$ and $x \in [0, 1]$. We use the above inequalities to complete the proof.

1) $\lim_{m\to\infty}\frac{n(m)}{2^m} = 0$. There exists a positive and vanishing sequence $\delta(m) = 1-2^{-\frac{n(m)+1}{2^m}}$ such that

$$\begin{aligned} F_{b^*}(\delta(m)) &= (1-(1-\delta(m))^{2^m})^{2^{n(m)}} \\ &\geqslant 1-2^{n(m)}(1-\delta(m))^{2^m} \\ &= \frac{1}{2}. \end{aligned}$$

It is straightforward that $F_{b^*}(0) = 0$. Therefore, the half-way point $\alpha_{b^*} = o(1)$.

2) $\lim_{m\to\infty}\frac{n(m)}{2^m} \nearrow 1$ and $n(m) < 2^m$. Set $p(m) = 2^m - n(m)$, where $p(m)$ is a positive integer and $\lim_{m\to\infty}\frac{p(m)}{2^m} = 0$. Then we have the following:

$$\begin{aligned} F_{b^*}\left(\frac{1}{2}\right) &= \left(1-\frac{1}{2^{2^m}}\right)^{2^{2^m-p(m)}} \\ &\geqslant 1-\frac{1}{2^{2^m}}\cdot 2^{2^m-p(m)} \\ &= 1-\frac{1}{2^{p(m)}} \\ &\geqslant \frac{1}{2}. \end{aligned}$$

There exists a positive and vanishing sequence $\delta(m) = \frac{1}{2}[(2^{p(m)}\cdot\ln 2)^{\frac{1}{2^m}} - 1]$ such that

$$\begin{aligned} F_{b^*}\left(\frac{1}{2}-\delta(m)\right) &= \left[1-\left(\frac{1}{2}+\delta(m)\right)^{2^m}\right]^{2^{2^m-p(m)}} \\ &\leqslant \exp\left\{-\frac{(1+2\delta(m))^{2^m}}{2^{p(m)}}\right\} \\ &= \frac{1}{2}. \end{aligned}$$

Therefore, the half-way point $\alpha_{b^*} = \frac{1}{2} - o(1)$.

3) $\lim_{m\to\infty}\frac{n(m)}{2^m} \searrow 1$ and $n(m) \geqslant 2^m$. Set $q(m) = n(m) - 2^m$, where $q(m)$ is a non-negative integer and $\lim_{m\to\infty}\frac{q(m)}{2^m} = 0$. Then we have the following:

$$\begin{aligned} F_{b^*}\left(\frac{1}{2}\right) &= \left(1-\frac{1}{2^{2^m}}\right)^{2^{2^m+q(m)}} \\ &\leqslant \exp\left\{-\frac{2^{2^m+q(m)}}{2^{2^m}}\right\} \\ &= \exp\left\{-2^{q(m)}\right\} \\ &\leqslant e^{-1}. \end{aligned}$$

There exists a positive and vanishing sequence $\delta(m) = \frac{1}{2}[1 - 2^{-\frac{q(m)+1}{2^m}}]$ such that

$$\begin{aligned} F_{b^*}\left(\frac{1}{2}+\delta(m)\right) &= \left[1-\left(\frac{1}{2}-\delta(m)\right)^{2^m}\right]^{2^{2^m+q(m)}} \\ &\geqslant 1-2^{2^m+q(m)}\cdot\left(\frac{1}{2}-\delta(m)\right)^{2^m} \\ &= \frac{1}{2}. \end{aligned}$$

Therefore, the half-way point $\alpha_{b^*} = \frac{1}{2} + o(1)$.

4) $\lim_{m\to\infty}\frac{n(m)}{2^m} = \infty$. Set $r(m) = \frac{n(m)}{2^m}$, where $\lim_{m\to\infty} r(m) = \infty$. Then there exists a positive and vanishing sequence $\delta(m) = \frac{(\ln 2)^{\frac{1}{2^m}}}{2^{r(m)}}$ such that

$$\begin{aligned} F_{b^*}(1-\delta(m)) &= (1-\delta(m)^{2^m})^{2^{2^m r(m)}} \\ &\leqslant \exp\left\{-(2^{r(m)}\delta(m))^{2^m}\right\} \\ &= \frac{1}{2}. \end{aligned}$$

It is straightforward that $F_{b^*}(1) = 1$. Therefore, the half-way point $\alpha_{b^*} = 1 - o(1)$. ∎

ACKNOWLEDGEMENT

The first author thanks Bing Fan for stimulating discussions at the preliminary stage of this work. Furthermore, we would like to thank the anonymous reviewers for their valuable comments that helped to improve this paper.

REFERENCES

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

[2] M. Bardet, J. Chaulet, V. Dragoi, A. Otmani, and J.-P. Tillich, "Cryptanalysis of the McEliece public key cryptosystem based on polar codes," in *Post-Quantum Cryptography* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2016.

[3] M. Bardet, V. Dragoi, A. Otmani, and J.-P. Tillich, "Algebraic properties of polar codes from a new polynomial formalism," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 230–234.

[4] S. Buzaglo, A. Fazeli, P. H. Siegel, V. Taranalli, and A. Vardy, "Permuted successive cancellation decoding for polar codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2618–2622.

[5] M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang, "Relaxed polar codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1986–2000, Apr. 2017.

[6] A. Fazeli and A. Vardy, "On the scaling exponent of binary polarization kernels," in *Proc. 52th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep./Oct. 2014, pp. 797–804.
[7] B. C. Geiger, "The fractality of polar and Reed–Müller codes," *Entropy*, vol. 20, no. 1, p. 70, Jan. 2018.
[8] A. G. I. Fàbregas, I. Land, and A. Martinez, "Extremes of error exponents," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2201–2207, Apr. 2013.
[9] J. Guo, A. G. I. Fàbregas, and J. Sayir, "Fixed-threshold polar codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 947–951.
[10] S. H. Hassani, K. Alishahi, and R. L. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014.
[11] G. He *et al.*, "β-expansion: A theoretical framework for fast and recursive construction of polar codes," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.
[12] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.
[13] D. Kim, K. Oh, D. Kim, and J. Ha, "Information set analysis of polar codes," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Jeju, South Korea, Oct. 2016, pp. 813–815.
[14] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.
[15] S. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6253–6264, Dec. 2010.
[16] M. Mondelli, R. Urbanke, and S. H. Hassani, "How to achieve the capacity of asymmetric channels," in *Proc. 52nd Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep./Oct. 2014, pp. 789–796.
[17] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "Construction of polar codes with sublinear complexity," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2782–2791, May 2019.
[18] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun./Jul. 2009, pp. 1496–1500.
[19] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, vol. 13, no. 7, pp. 519–521, Jul. 2009.
[20] E. Ordentlich and R. M. Roth, "On the pointwise threshold behavior of the binary erasure polarization subchannels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 859–863.
[21] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 955–969, Feb. 2013.
[22] N. Presman, O. Shapira, S. Litsyn, T. Etzion, and A. Vardy, "Binary polarization kernels from code decompositions," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2227–2239, May 2015.
[23] M. Qin, J. Guo, A. Bhatia, A. G. I. Faàbregas, and P. Siegel, "Polar code constructions based on LLR evolution," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1221–1224, Jun. 2017.
[24] J. M. Renes, D. Sutter, and S. H. Hassani, "Alignment of polarized sets," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 2, pp. 224–238, Feb. 2016.
[25] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7839–7857, Dec. 2013.
[26] G. Sarkis, I. Tal, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Flexible and low-complexity encoding and decoding of systematic polar codes," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2732–2745, Jul. 2016.
[27] E. Şaşoğlu, E. Telatar, and E. Arikan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Taormina, Italy, Oct. 2009, pp. 144–148.
[28] C. Schürch, "A partial order for the synthesized channels of a polar code," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 220–224.
[29] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.
[30] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3221–3227, Nov. 2012.
[31] Y. Wang, K. R. Narayanan, and A. A. Jiang, "Exploiting source redundancy to improve the rate of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 864–868.
[32] W. Wu, B. Fan, and P. H. Siegel, "Generalized partial orders for polar code bit-channels," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2017, pp. 541–548.
[33] R. Zhang, Y. Ge, H. Saber, W. Shi, and X. Shen, "Localization-based polar code construction with sublinear complexity," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

Wei Wu received the B.E. degree in Information Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013 and the M.S. degree in Electrical Engineering from NYU Tandon School of Engineering, New York, NY, in 2015. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of California, San Diego, where he is also affiliated with the Center for Memory and Recording Research. He is interested in information and coding theory. His current research focuses on polar coding and its applications.

Paul H. Siegel (M'82–SM'90–F'97–LF'19) received the S.B. and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1975 and 1979, respectively. He held a Chaim Weizmann Postdoctoral Fellowship with the Courant Institute, New York University, New York, NY, USA. He was with the IBM Research Division, San Jose, CA, USA, from 1980 to 1995. He joined the faculty at University of California, San Diego, CA, USA, in 1995, where he is currently a Distinguished Professor of Electrical and Computer Engineering with the Jacobs School of Engineering. He is affiliated with the Center for Memory and Recording Research where he holds an Endowed Chair and served as Director from 2000 to 2011. His research interests include information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission. He is a Member of the National Academy of Engineering. He was a Member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2009 to 2014. He was the 2015 Padovani Lecturer of the IEEE Information Theory Society. He was a recipient of the 2007 Best Paper Award in Signal Processing and Coding for Data Storage from the Data Storage Technical Committee of the IEEE Communications Society. He was the co-recipient of the 1992 IEEE Information Theory Society Paper Award and the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award. He served as a Co-Guest Editor of the 1991 Special Issue on Coding for Storage Devices of the IEEE TRANSACTIONS ON INFORMATION THEORY. He served as an Associate Editor of Coding Techniques of the IEEE TRANSACTIONS ON INFORMATION THEORY from 1992 to 1995, and as the Editor-in-Chief from 2001 to 2004. He was also a Co-Guest Editor of the 2001 two-part issue on The Turbo Principle: From Theory to Practice and the 2016 issue on Recent Advances in Capacity Approaching Codes of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.