Scaling Exponent and Moderate Deviations Asymptotics of Polar Codes for the AWGN Channel

Silas L. Fong * and Vincent Y. F. Tan (ID)

Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, Singapore; vtan@nus.edu.sg
* Correspondence: silas.l.fong@gmail.com; Tel.: +65-6516-2133

Abstract: This paper investigates polar codes for the additive white Gaussian noise (AWGN) channel. The scaling exponent μ of polar codes for a memoryless channel $q_{Y|X}$ with capacity $I(q_{Y|X})$ characterizes the closest gap between the capacity and non-asymptotic achievable rates as follows: For a fixed $\varepsilon \in (0,1)$, the gap between the capacity $I(q_{Y|X})$ and the maximum non-asymptotic rate R_n^* achieved by a length-n polar code with average error probability ε scales as $n^{-1/\mu}$, i.e., $I(q_{Y|X}) - R_n^* = \Theta(n^{-1/\mu})$. It is well known that the scaling exponent μ for any binary-input memoryless channel (BMC) with $I(q_{Y|X}) \in (0,1)$ is bounded above by 4.714. Our main result shows that 4.714 remains a valid upper bound on the scaling exponent for the AWGN channel. Our proof technique involves the following two ideas: (i) The capacity of the AWGN channel can be achieved within a gap of $O(n^{-1/\mu}\sqrt{\log n})$ by using an input alphabet consisting of n constellations and restricting the input distribution to be uniform; (ii) The capacity of a multiple access channel (MAC) with an input alphabet consisting of n constellations can be achieved within a gap of $O(n^{-1/\mu}\log n)$ by using a superposition of $\log n$ binary-input polar codes. In addition, we investigate the performance of polar codes in the moderate deviations regime where both the gap to capacity and the error probability vanish as n grows. An explicit construction of polar codes is proposed to obey a certain tradeoff between the gap to capacity and the decay rate of the error probability for the AWGN channel.

Keywords: AWGN channel; moderate deviations; multiple access channel; polar codes; scaling exponent

1. Introduction

1.1. The Additive White Gaussian Noise Channel

This paper investigates low-complexity codes over the classical additive white Gaussian noise (AWGN) channel ([1], Chapter 9), where a source wants to transmit information to a destination and each received symbol is the sum of the transmitted symbol and an independent Gaussian random variable. More specifically, if X_k denotes the symbol transmitted by the source in the kth time slot, then the corresponding symbol received by the destination is

$$Y_k = X_k + Z_k \tag{1}$$

where Z_k is the standard normal random variable. When the transmission lasts for n time slots, i.e., each transmitted codeword consists of n symbols, it is assumed that $Z_1, Z_2, \ldots, Z_n$ are independent and each transmitted codeword $x^n \triangleq (x_1, x_2, \ldots, x_n)$ must satisfy the peak power constraint

$$\frac{1}{n}\sum_{k=1}^{n} x_k^2 \le P \tag{2}$$

where $P > 0$ is a constant which denotes the permissible power. For transmitting a uniformly distributed message $W \in \{1, 2, \ldots, \lceil 2^{nR} \rceil\}$ across this channel, Shannon [2] shows that the limit of the maximum coding rate R as n approaches infinity (i.e., capacity) is

$$C(P) \triangleq \frac{1}{2} \log(1 + P). \tag{3}$$

1.2. Polar Codes

In 1948, in his groundbreaking paper Shannon [2] proposed a systematic framework for studying the fundamental limits for transmitting information over noisy channels and provided a single-letter formula for the capacity of a memoryless channel. For decades, information and coding theorists sought to achieve these fundamental limits via low-complexity and capacity-achieving codes. In 2009, in his breakthrough paper Arıkan [3] proposed a class of codes—known as polar codes—whose encoding and decoding complexities are $O(n \log n)$ and provably achieve the capacity of any binary-input memoryless symmetric channel (BMSC).

The scaling exponent μ of polar codes for a memoryless channel $q_{Y|X}$ with capacity

$$I(q_{Y|X}) \triangleq \max_{p_X} I(X; Y) \tag{4}$$

characterizes the closest gap between the channel capacity and non-asymptotic achievable rates as follows: For a fixed $\varepsilon \in (0, 1)$, the gap between the capacity $I(q_{Y|X})$ and the maximum non-asymptotic rate R_n^* achieved by a length-n polar code with average error probability ε scales as $n^{-1/\mu}$, i.e., $I(q_{Y|X}) - R_n^* = \Theta(n^{-1/\mu})$. It has been shown in [4–6] that the scaling exponent μ for any BMSC with $I(q_{Y|X}) \in (0, 1)$ lies between 3.579 and 4.714. Indeed, the upper bound 4.714 remains valid for any general binary-input memoryless channel (BMC) ([7], Lemma 4). The scaling exponent of polar codes for a non-stationary channel has been recently studied in [8].

It is well known that polar codes are capacity-achieving for BMCs [9–12], and appropriately chosen ones are also capacity-achieving for the AWGN channel [13]. In particular, for any $R < C(P)$ and any $\beta < 1/2$, polar codes operated at rate R can be constructed for the AWGN channel such that the decay rate of the error probability is $O(2^{-n^\beta})$ [13] and the encoding and decoding complexities are $O(n \log n)$. However, the scaling exponent of polar codes for the AWGN channel has not been investigated yet.

In this paper, we construct polar codes for the AWGN channel and show that 4.714 remains a valid upper bound on the scaling exponent. Our construction of polar codes involves the following two ideas:(i) By using an input alphabet consisting of n constellations and restricting the input distribution to be uniform as suggested in [13], we can achieve the capacity of the AWGN channel within a gap of $O(n^{-1/\mu} \sqrt{\log n})$; (ii) By using a superposition of $\log n$ binary-input polar codes (in this paper, n is always a power of 2) as suggested in [14], we can achieve the capacity of the corresponding multiple access channel (MAC) within a gap of $O(n^{-1/\mu} \log n)$ where the input alphabet of the MAC has n constellations (i.e., the size of the Cartesian product of the input alphabets corresponding to the $\log n$ input terminals is n). The encoding and decoding complexities of our constructed polar codes are $O(n \log^2 n)$. On the other hand, the lower bound 3.579 holds trivially for the constructed polar codes because the polar codes are constructed by superposing $\log n$ binary-input polar codes whose scaling exponents are bounded below by 3.579 [5].

In addition, Mondelli et al. ([4], Section IV) provided an explicit construction of polar codes for any BMSC which obeys a certain tradeoff between the gap to capacity and the decay rate of the error probability. More specifically, if the gap to capacity is set to vanish at a rate of $\Theta\left(n^{-\frac{1-\gamma}{\mu}}\right)$ for some $\gamma \in \left(\frac{1}{1+\mu}, 1\right)$, then a length-$n$ polar code can be constructed such that the error probability

is $O\left(n \cdot 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\mu+\gamma-1}{\gamma\mu}\right)}}\right)$ where $h_2 : [0,1/2] \to [0,1]$ denotes the binary entropy function. This tradeoff was developed under the moderate deviations regime [15] where both the gap to capacity and the error probability vanish as n grows. For the AWGN channel, we develop a similar tradeoff under the moderate deviations regime by using our constructed polar codes described above.

1.3. Paper Outline

This paper is organized as follows. The notation used in this paper is described in the next subsection. Section 2 presents the background of this work, which includes existing polarization results for the BMC which are used in this work. Sections 3.1–3.3 state the formulation of the binary-input MAC and present new polarization results for the binary-input MAC. Sections 4.1–4.2 state the formulation of the AWGN channel and present new polarization results for the AWGN channel. Section 5 establishes the definition of the scaling exponent for the AWGN channel and establishes the main result—4.714 is an upper bound on the scaling exponent of polar codes for the AWGN channel. Section 6 presents an explicit construction of polar codes for the AWGN channel which obey a certain tradeoff between the gap to capacity and the decay rate of the error probability under the moderate deviations regime. Concluding remarks are provided in Section 7.

1.4. Notation

The set of natural numbers, real numbers and non-negative real numbers are denoted by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}_+$ respectively. For any sets $\mathcal{A}$ and $\mathcal{B}$ and any mapping $f : \mathcal{A} \to \mathcal{B}$, we let $f^{-1}(\mathcal{D})$ denote the set $\{a \in \mathcal{A} \,|\, f(a) \in \mathcal{D}\}$ for any $\mathcal{D} \subseteq \mathcal{B}$. We let $\mathbf{1}\{\mathcal{E}\}$ be the indicator function of the set $\mathcal{E}$. An arbitrary (discrete or continuous) random variable is denoted by an upper-case letter (e.g., X), and the realization and the alphabet of the random variable are denoted by the corresponding lower-case letter (e.g., x) and calligraphic letter (e.g., $\mathcal{X}$) respectively. We use X^n to denote the random tuple $(X_1, X_2, \ldots, X_n)$ where each X_k has the same alphabet $\mathcal{X}$. We will take all logarithms to base 2 throughout this paper.

The following notations are used for any of the arbitrary X and Y random variables and any real-valued function g with domain $\mathcal{X}$. We let $p_{Y|X}$ and $p_{X,Y} = p_X p_{Y|X}$ denote the conditional probability distribution of Y given X and the probability distribution of (X, Y) respectively. We let $p_{X,Y}(x,y)$ and $p_{Y|X}(y|x)$ be the evaluations of $p_{X,Y}$ and $p_{Y|X}$ respectively at $(X, Y) = (x, y)$. To make the dependence on the distribution explicit, we let $\mathbb{P}_{p_X}\{g(X) \in \mathcal{A}\}$ denote $\int_{\mathcal{X}} p_X(x)\mathbf{1}\{g(x) \in \mathcal{A}\}\,\mathrm{d}x$ for any set $\mathcal{A} \subseteq \mathbb{R}$. The expectation of $g(X)$ is denoted as $\mathrm{E}_{p_X}[g(X)]$. For any (X, Y, Z) distributed according to some $p_{X,Y,Z}$, the entropy of X and the conditional mutual information between X and Y given Z are denoted by $H_{p_X}(X)$ and $I_{p_{X,Y,Z}}(X;Y|Z)$ respectively. For simplicity, we sometimes omit the subscript of a notation if it causes no confusion. The relative entropy between p_X and q_X is denoted by

$$D(p_X \| q_X) \triangleq \int_{\mathcal{X}} p_X(x) \log\left(\frac{p_X(x)}{q_X(x)}\right)\,\mathrm{d}x. \tag{5}$$

The 2-Wasserstein distance between p_X and p_Y is denoted by

$$W_2(p_X, p_Y) \triangleq \inf_{\substack{s_{X,Y}: \\ s_X=p_X, \\ s_Y=p_Y}} \sqrt{\int_{\mathcal{X}} \int_{\mathcal{Y}} s_{X,Y}(x,y)(x-y)^2 \mathrm{d}y\mathrm{d}x}. \tag{6}$$

We let $\mathcal{N}(\,\cdot\,; \mu, \sigma^2) : \mathbb{R} \to [0, \infty)$ denote the probability density function of a Gaussian random variable whose mean and variance are μ and σ^2 respectively, i.e.,

$$\mathcal{N}(z; \mu, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{7}$$

2. Background: Point-to-Point Channels and Existing Polarization Results

In this section, we will review important polarization results related to the scaling exponent of polar codes for binary-input memoryless channels (BMCs).

2.1. Point-to-Point Memoryless Channels

Consider a point-to-point channel which consists of one source and one destination, denoted by s and d respectively. Suppose node s transmits information to node d in *n* time slots. Before any transmission begins, node s chooses message *W* destined for node d, where *W* is uniformly distributed over the alphabet

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\} \tag{8}$$

which consists of *M* elements. For each $k \in \{1, 2, \ldots, n\}$, node s transmits $X_k \in \mathcal{X}$ based on *W* and node d receives $Y_k \in \mathcal{Y}$ in time slot *k* where $\mathcal{X}$ and $\mathcal{Y}$ denote respectively the input and output alphabets of the channel. After *n* time slots, node d declares $\hat{W}$ to be the transmitted *W* based on Y^n. Formally, we define a length-*n* code as follows.

Definition 1. *An* (n, M) code *consists of the following:*

1. *A message set* $\mathcal{W}$ *as defined in* (8)*. Message W is uniform on* $\mathcal{W}$*.*
2. *An encoding function* $f_k : \mathcal{W} \to \mathcal{X}$ *for each* $k \in \{1, 2, \ldots, n\}$*, where* f_k *is used by node* s *for encoding* X_k *such that* $X_k = f_k(W)$*.*
3. *A decoding function* $\varphi : \mathcal{Y}^n \to \mathcal{W}$ *used by node* d *for producing the message estimate* $\hat{W} = \varphi(Y^n)$*.*

Definition 2. *The* point-to-point memoryless channel *is characterized by an input alphabet* $\mathcal{X}$*, an output alphabet* $\mathcal{Y}$ *and a conditional distribution* $q_{Y|X}$ *such that the following holds for any* (n, M) code*: For each* $k \in \{1, 2, \ldots, n\}$*,* $p_{W, X^n, Y^n} = p_{W, X^n} \prod_{k=1}^{n} p_{Y_k|X_k}$ *where* $p_{Y_k|X_k}(y_k|x_k) = q_{Y|X}(y_k|x_k)$ *for all* $x_k \in \mathcal{X}$ *and* $y_k \in \mathcal{Y}$*.*

For any (n, M) code defined on the point-to-point memoryless channel, let $p_{W, X^n, Y^n, \hat{W}}$ be the joint distribution induced by the code. By Definitions 1 and 2, we can factorize $p_{W, X^n, Y^n, \hat{W}}$ as,

$$p_{W, X^n, Y^n, \hat{W}} = p_W p_{X^n|W} \left(\prod_{k=1}^{n} p_{Y_k|X_k} \right) p_{\hat{W}|Y^n}. \tag{9}$$

2.2. Polarization for Binary-Input Memoryless Channels

Definition 3. *A point-to-point memoryless channel characterized by* $q_{Y|X}$ *is called a binary-input memoryless channel (BMC) if* $\mathcal{X} = \{0, 1\}$*.*

We follow the formulation of polar coding in [11]. Consider any BMC characterized by $q_{Y|X}$. Let p_X be the probability distribution of a Bernoulli random variable *X*, and let p_{X^n} be the distribution of *n* independent copies of $X \sim p_X$, i.e., $p_{X^n}(x^n) = \prod_{k=1}^{n} p_X(x_k)$ for all $x^n \in \mathcal{X}^n$. For each $n = 2^m$ where $m \in \mathbb{N}$, the polarization mapping of a length-*n* polar code is given by

$$G_n \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m} = G_n^{-1} \tag{10}$$

where $\otimes$ denotes the Kronecker power. Define $p_{U^n|X^n}$ such that

$$[U_1 \, U_2 \, \ldots \, U_n] = [X_1 \, X_2 \, \ldots \, X_n] G_n \tag{11}$$

where the addition and product operations are performed over GF(2), define

$$p_{Y_k|X_k}(y_k|x_k) \triangleq q_{Y|X}(y_k|x_k) \tag{12}$$

for each $k \in \{1, 2, \dots, n\}$ and each $(x_k, y_k) \in \mathcal{X} \times \mathcal{Y}$ where $q_{Y|X}$ characterizes the BMC (cf. (2)), and define

$$p_{U^n, X^n, Y^n} \triangleq p_{X^n} p_{U^n|X^n} \prod_{k=1}^n p_{Y_k|X_k}. \tag{13}$$

In addition, for each $k \in \{1, 2, \dots, n\}$, define the Bhattacharyya parameter associated with time k as

$$Z^{[p_X;q_{Y|X}]}(U_k|U^{k-1}, Y^n)$$

$$\triangleq 2 \sum_{u^{k-1} \in \mathcal{U}^{k-1}} \int_{\mathcal{Y}^n} p_{U^{k-1}, Y^n}(u^{k-1}, y^n) \sqrt{p_{U_k|U^{k-1}, Y^n}(0|u^{k-1}, y^n) p_{U_k|U^{k-1}, Y^n}(1|u^{k-1}, y^n)} \mathrm{d}y^n \tag{14}$$

$$= 2 \sum_{u^{k-1} \in \mathcal{U}^{k-1}} \int_{\mathcal{Y}^n} \sqrt{p_{U_k, U^{k-1}, Y^n}(0, u^{k-1}, y^n) p_{U_k, U^{k-1}, Y^n}(1, u^{k-1}, y^n)} \mathrm{d}y^n \tag{15}$$

where the distributions in (14) and (15) are marginal distributions of p_{U^n, X^n, Y^n} defined in (13). To simplify notation, let

$$\beta \triangleq 4.714 \tag{16}$$

in the rest of this paper. The following lemma is based on Section III of [4] and it is a restatement of Lemma 2 of [7], which has been used in [7] to show that 4.714 is an upper bound on the scaling exponent for any BMC (including non-symmetric ones).

Lemma 1. *([7], Lemma 2) There exists a universal constant $t > 0$ such that the following holds. Fix any BMC characterized by $q_{Y|X}$ and any p_X. Then for any $m \in \mathbb{N}$ and $n \triangleq 2^m$, we have*

$$\frac{1}{n} \left| \left\{ k \in \{1, 2, \dots, n\} \,\middle|\, Z^{[p_X;q_{Y|X}]}(U_k|U^{k-1}, Y^n) \le \tfrac{1}{n^4} \right\} \right| \ge I_{p_X q_{Y|X}}(X; Y) - \frac{t}{n^{1/\beta}}. \tag{17}$$

This lemma continues to hold if the quantity $\frac{1}{n^4}$ is replaced by $\frac{1}{n^\nu}$ for any $\nu > 0$. The main result of this paper continues to hold if the quantity $\frac{1}{n^4}$ in this lemma is replaced by $\frac{1}{n^\nu}$ for any $\nu > 2$.

3. Problem Formulation of Binary-Input MACs and New Polarization Results

Polar codes have been proposed and investigated for achieving any rate tuple inside the capacity region of a binary-input multiple access channel (MAC) [14,16]. The goal of this section is to use the polar codes proposed in [14] to achieve the symmetric sum-capacity of a binary-input MAC.

3.1. Binary-Input Multiple Access Channels

Consider a MAC ([1], Section 15.3) which consists of N sources and one destination. Let $\mathcal{I} \triangleq \{1, 2, \dots, N\}$ be the index set of the N sources and let d denote the destination. Suppose the sources transmit information to node d in n time slots. Before any transmission begins, node i chooses message W_i destined for node d for each $i \in \mathcal{I}$, where W_i is uniformly distributed over

$$\mathcal{W}_i \triangleq \{1, 2, \dots, M_i\} \tag{18}$$

which consists of M_i elements. For each $k \in \{1, 2, \dots, n\}$, node i transmits $X_{i,k} \in \mathcal{X}_i$ based on W_i for each $i \in \mathcal{I}$ and node d receives $Y_k \in \mathcal{Y}$ in time slot k where $\mathcal{X}_i$ denotes the input alphabet for node i and $\mathcal{Y}$ denotes the output alphabet. After n time slots, node d declares $\hat{W}_i$ to be the transmitted W_i based on Y^n for each $i \in \mathcal{I}$.

To simplify notation, we use the following convention for any $T \subseteq \mathcal{I}$. For any random tuple $(X_1, X_2, \ldots, X_N)$, we let $X_T \triangleq (X_i : i \in T)$ be the corresponding subtuple, whose realization and alphabet are denoted by x_T and $\mathcal{X}_T$ respectively. Similarly, for each $k \in \{1, 2, \ldots, n\}$ and each random tuple $(X_{1,k}, X_{2,k}, \ldots, X_{N,k}) \in \mathcal{X}_{\mathcal{I}}$, we let $X_{T,k} \triangleq (X_{i,k} : i \in T)$ denote the corresponding random subtuple, and let $x_{T,k}$ and $\mathcal{X}_{T,k}$ denote respectively the realization and the alphabet of $X_{T,k}$. Formally, we define a length-n code for the binary-input MAC as follows.

Definition 4. *An $(n, M_{\mathcal{I}})$-code, where $M_{\mathcal{I}} \triangleq (M_1, M_2, \ldots, M_N)$, consists of the following:*

1. *A message set $\mathcal{W}_i$ for each $i \in \mathcal{I}$ as defined in (18), where message W_i is uniform on $\mathcal{W}_i$.*
2. *An encoding function $f_{i,k}^{MAC} : \mathcal{W}_i \to \mathcal{X}_i$ for each $i \in \mathcal{I}$ and each $k \in \{1, 2, \ldots, n\}$, where $f_{i,k}^{MAC}$ is used by node i for encoding $X_{i,k}$ such that $X_{i,k} = f_{i,k}^{MAC}(W_i)$.*
3. *A decoding function $\varphi^{MAC} : \mathcal{Y}^n \to \mathcal{W}_{\mathcal{I}}$ used by node d for producing the message estimate $\hat{W}_{\mathcal{I}} = \varphi^{MAC}(Y^n)$.*

Definition 5. *The* multiple access channel (MAC) *is characterized by N input alphabets specified by $\mathcal{X}_{\mathcal{I}}$, an output alphabet specified by $\mathcal{Y}$ and a conditional distribution $q_{Y|X_{\mathcal{I}}}$ such that the following holds for any $(n, M_{\mathcal{I}})$ code: For each $k \in \{1, 2, \ldots, n\}$,*

$$p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y^n} = \left(\prod_{i \in \mathcal{I}} p_{W_i, X_i^n} \right) \prod_{k=1}^n p_{Y_k | X_{\mathcal{I},k}} \tag{19}$$

where $p_{Y_k | X_{\mathcal{I},k}}(y_k | x_{\mathcal{I},k}) = q_{Y|X_{\mathcal{I}}}(y_k | x_{\mathcal{I},k})$ for all $x_{\mathcal{I},k} \in \mathcal{X}_{\mathcal{I}}$ and $y_k \in \mathcal{Y}$.

3.2. Polarization for Binary-Input MACs

Definition 6. *A MAC characterized by $q_{Y|X_{\mathcal{I}}}$ is called a* binary-input MAC *if $\mathcal{X}_{\mathcal{I}} = \{0, 1\}^N$.*

Consider any binary-input MAC characterized by $q_{Y|X_{\mathcal{I}}}$. For each $i \in \mathcal{I}$, let p_{X_i} be the probability distribution of a Bernoulli random variable X_i, and let $p_{X_i^n}$ be the distribution of n independent copies of $X_i \sim p_{X_i}$, i.e., $p_{X_i^n}(x_i^n) = \prod_{k=1}^n p_{X_i}(x_{i,k})$ for all $x_i^n \in \mathcal{X}_i^n$. Recall the polarization mapping G_n defined in (10). For each $i \in \mathcal{I}$, define $p_{U_i^n | X_i^n}$ such that

$$[U_{i,1} \ U_{i,2} \ \ldots \ U_{i,n}] = [X_{i,1} \ X_{i,2} \ \ldots \ X_{i,n}] G_n \tag{20}$$

where the addition and product operations are performed over GF(2), and define

$$p_{U_{\mathcal{I}}^n, X_{\mathcal{I}}^n, Y^n} \triangleq \left(\prod_{i \in \mathcal{I}} p_{X_i^n} p_{U_i^n | X_i^n} \right) \prod_{k=1}^n p_{Y_k | X_{\mathcal{I},k}}. \tag{21}$$

In addition, for each $i \in \mathcal{I}$ and each $k \in \{1, 2, \ldots, n\}$, define $[i-1] \triangleq \{1, 2, \ldots, i-1\}$ and define the Bhattacharyya parameter associated with node i and time k as

$$Z^{[p_{X_{\mathcal{I}}}; q_{Y|X_{\mathcal{I}}}]}(U_{i,k} | U_i^{k-1}, X_{[i-1]}^n, Y^n) \triangleq$$

$$2 \sum_{u_i^{k-1} \in \mathcal{U}_i^{k-1}} \sum_{x_{[i-1]}^n \in \{0,1\}^{(i-1)n}} \int_{\mathcal{Y}^n} \sqrt{p_{U_{i,k}, U_i^{k-1}, X_{[i-1]}^n, Y^n}(0, u_i^{k-1}, x_{[i-1]}^n, y^n) p_{U_{i,k}, U_i^{k-1}, X_{[i-1]}^n, Y^n}(1, u_i^{k-1}, x_{[i-1]}^n, y^n)} \, dy^n \tag{22}$$

where the distributions in (22) are marginal distributions of $p_{U_{\mathcal{I}}^n, X_{\mathcal{I}}^n, Y^n}$ defined in (21). The following lemma is a direct consequence of Lemma 1.

Lemma 2. *There exists a universal constant $t > 0$ such that the following holds. Fix any binary-input MAC characterized by $q_{Y|X_{\mathcal{I}}}$ and any $p_{X_{\mathcal{I}}}$. Then for any $m \in \mathbb{N}$ and $n \triangleq 2^m$, we have*

$$\frac{1}{n} \left| \left\{ k \in \{1, 2, \ldots, n\} \left| Z^{[p_{X_{\mathcal{I}}}; q_{Y|X_{\mathcal{I}}}]}(U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n) \leq \frac{1}{n^4} \right. \right\} \right| \geq I_{p_{X_{\mathcal{I}}} q_{Y|X_{\mathcal{I}}}}(X_i; X_{[i-1]}, Y) - \frac{t}{n^{1/\beta}} \tag{23}$$

for each $i \in \mathcal{I}$. This lemma continues to hold if the quantity $\frac{1}{n^4}$ is replaced by $\frac{1}{n^\nu}$ for any $\nu > 0$. The main result of this paper continues to hold if the quantity $\frac{1}{n^4}$ in this lemma is replaced by $\frac{1}{n^\nu}$ for any $\nu > 2$.

Proof. Fix any $i \in \mathcal{I}$. Construct $p_{X_{[i-1]}, Y|X_i}$ by marginalizing $p_{X_{\mathcal{I}}} q_{Y|X_{\mathcal{I}}}$ and view $p_{X_{[i-1]}, Y|X_i}$ as the conditional distribution that characterizes a BMC. The lemma then follows directly from Lemma 1. $\square$

3.3. Polar Codes That Achieve the Symmetric Sum-Capacity of a Binary-Input MAC

Throughout this paper, let $p_{X_i}^*$ denote the uniform distribution on $\{0, 1\}$ for each $i \in \mathcal{I}$ and define $p_{X_{\mathcal{I}}}^* \triangleq \prod_{i \in \mathcal{I}} p_{X_i}^*$, i.e.,

$$p_{X_{\mathcal{I}}}^*(x_{\mathcal{I}}) = \frac{1}{2^N} \tag{24}$$

for any $x_{\mathcal{I}} \in \{0, 1\}^N$.

Definition 7. *For a binary-input MAC characterized by $q_{Y|X_{\mathcal{I}}}$, the symmetric sum-capacity is defined to be $C_{sum} \triangleq I_{p_{X_{\mathcal{I}}}^* q_{Y|X_{\mathcal{I}}}}(X_{\mathcal{I}}; Y)$.*

The following definition summarizes the polar codes for the binary-input MAC proposed in Section IV of [14].

Definition 8. *([14], Section IV) Fix an $n = 2^m$ where $m \in \mathbb{N}$. For each $i \in \mathcal{I}$, let $\mathcal{J}_i$ be a subset of $\{1, 2, \ldots, n\}$, define $\mathcal{J}_i^c \triangleq \{1, 2, \ldots, n\} \setminus \mathcal{J}_i$, and let $b_{i, \mathcal{J}_i^c} \triangleq (b_{i,k} \in \{0, 1\} : k \in \mathcal{J}_i^c)$ be a binary tuple. An $(n, \mathcal{J}_{\mathcal{I}}, b_{\mathcal{I}, \mathcal{J}_{\mathcal{I}}^c})$ polar code, where $\mathcal{J}_{\mathcal{I}} \triangleq (\mathcal{J}_i : i \in \mathcal{I})$ and $b_{\mathcal{I}, \mathcal{J}_{\mathcal{I}}^c} \triangleq (b_{i, \mathcal{J}_i^c} : i \in \mathcal{I})$, consists of the following:*

1. *An index set for information bits transmitted by node i denoted by $\mathcal{J}_i$ for each $i \in \mathcal{I}$. The set $\mathcal{J}_i^c$ is referred to as the index set for frozen bits transmitted by node i.*
2. *A message set $\mathcal{W}_i \triangleq \{1, 2, \ldots, 2^{|\mathcal{J}_i|}\}$ for each $i \in \mathcal{I}$, where W_i is uniform on $\mathcal{W}_i$.*
3. *An encoding bijection $f_i^{MAC} : \mathcal{W}_i \to \mathcal{U}_{i, \mathcal{J}_i}$ for encoding W_i into $|\mathcal{J}_i|$ information bits denoted by $U_{i, \mathcal{J}_i}$ for each $i \in \mathcal{I}$ such that*

$$U_{i, \mathcal{J}_i} = f_i^{MAC}(W_i) \tag{25}$$

where $\mathcal{U}_{i, \mathcal{J}_i}$ and $U_{i, \mathcal{J}_i}$ are defined as $\mathcal{U}_{i, \mathcal{J}_i} \triangleq \prod_{k \in \mathcal{J}_i} \mathcal{U}_k$ and $U_{i, \mathcal{J}_i} \triangleq (U_{i,k} : k \in \mathcal{J}_i)$ respectively. Since message W_i is uniform on $\mathcal{W}_i$, $f_i^{MAC}(W_i)$ is a sequence of independent and identically distributed (i.i.d.) uniform bits such that

$$\mathbb{P}\{U_{i, \mathcal{J}_i} = u_{i, \mathcal{J}_i}\} = \frac{1}{2^{|\mathcal{J}_i|}} \tag{26}$$

for all $u_{i, \mathcal{J}_i} \in \{0, 1\}^{|\mathcal{J}_i|}$, where the bits are transmitted through the polarized channels indexed by $\mathcal{J}_i$. For each $i \in \mathcal{I}$ and each $k \in \mathcal{J}_i^c$, let

$$U_{i,k} = b_{i,k} \tag{27}$$

be the frozen bit to be transmitted by node i in time slot k. After U_i^n has been determined, node i transmits X_i^n where

$$[X_{i,1} \ X_{i,2} \ \ldots \ X_{i,n}] \triangleq [U_{i,1} \ U_{i,2} \ \ldots \ U_{i,n}] G_n^{-1}. \tag{28}$$

4. *A sequence of successive cancellation decoding functions $\varphi_{i,k}^{MAC} : \{0, 1\}^{k-1} \times \{0, 1\}^{(i-1)n} \times \mathcal{Y}^n \to \{0, 1\}$ for each $i \in \mathcal{I}$ and each $k \in \{1, 2, \ldots, n\}$ such that the recursively generated $(\hat{U}_{1,1}, \ldots, \hat{U}_{1,n})$,*

$(\hat{U}_{2,1},\ldots,\hat{U}_{2,n}),\ldots,(\hat{U}_{N,1},\ldots,\hat{U}_{N,n})$ and $(\hat{X}_{1,1},\ldots,\hat{X}_{1,n}),(\hat{X}_{2,1},\ldots,\hat{X}_{2,n}),\ldots,(\hat{X}_{N,1},\ldots,\hat{X}_{N,n})$ are produced as follows. For each $i \in \mathcal{I}$ and each $k = 1, 2, \ldots, n$, given that $\hat{U}_i^{k-1}$, $\hat{U}_{[i-1]}^n$ and $\hat{X}_{[i-1]}^n$ have been constructed before the construction of $\hat{U}_{i,k}$, node d constructs the estimate of $U_{i,k}$ through computing

$$\hat{U}_{i,k} \triangleq \varphi_{i,k}^{MAC}(\hat{U}_i^{k-1}, \hat{X}_{[i-1]}^n, Y^n) \tag{29}$$

where

$$\hat{u}_{i,k} \triangleq \varphi_{i,k}^{MAC}(\hat{u}_i^{k-1}, \hat{x}_{[i-1]}^n, y^n)$$

$$= \begin{cases} 0 & \text{if } k \in \mathcal{J}_i \text{ and } p_{U_{i,k}|U_i^{k-1},X_{[i-1]}^n,Y^n}(0|\hat{u}_i^{k-1}, \hat{x}_{[i-1]}^n, y^n) \geq p_{U_{i,k}|U_i^{k-1},X_{[i-1]}^n,Y^n}(1|\hat{u}_i^{k-1}, \hat{x}_{[i-1]}^n, y^n), \\ 1 & \text{if } k \in \mathcal{J}_i \text{ and } p_{U_{i,k}|U_i^{k-1},X_{[i-1]}^n,Y^n}(0|\hat{u}_i^{k-1}, \hat{x}_{[i-1]}^n, y^n) < p_{U_{i,k}|U_i^{k-1},X_{[i-1]}^n,Y^n}(1|\hat{u}_i^{k-1}, \hat{x}_{[i-1]}^n, y^n), \\ b_{i,k} & \text{if } k \in \mathcal{J}_i^c. \end{cases} \tag{30}$$

After obtaining $\hat{U}_i^n$, node d constructs the estimate of X_i^n through computing

$$[\hat{X}_{i,1} \ \hat{X}_{i,2} \ \ldots \ \hat{X}_{i,n}] \triangleq [\hat{U}_{i,1} \ \hat{U}_{i,2} \ \ldots \ \hat{U}_{i,n}]G_n^{-1} \tag{31}$$

and declares that

$$\hat{W}_i \triangleq \left(f_i^{MAC}\right)^{-1}(\hat{U}_{i,\mathcal{J}_i}) \tag{32}$$

is the transmitted W_i where $\left(f_i^{MAC}\right)^{-1}$ denotes the inverse function of f_i^{MAC}.

Remark 1. *By inspecting Definition 4 and Definition 8, we see that every $(n, \mathcal{J}_\mathcal{I}, b_{\mathcal{I},\mathcal{J}_\mathcal{I}^c})$ polar code is also an $\left(n, (2^{|\mathcal{J}_1|}, 2^{|\mathcal{J}_2|}, \ldots, 2^{|\mathcal{J}_N|})\right)$ code.*

Definition 9. *The uniform-input $(n, \mathcal{J}_\mathcal{I})$ polar code is defined as an $(n, \mathcal{J}_\mathcal{I}, B_{\mathcal{I},\mathcal{J}_\mathcal{I}^c})$ polar code where $B_{\mathcal{I},\mathcal{J}_\mathcal{I}^c}$ consists of i.i.d. uniform bits that are independent of the message $W_\mathcal{I}$.*

Definition 10. *For the uniform-input $(n, \mathcal{J}_\mathcal{I})$ polar code defined for the MAC, the probability of decoding error is defined as*

$$\mathbb{P}\{\hat{W}_\mathcal{I} \neq W_\mathcal{I}\} = \mathbb{P}\left\{ \bigcup_{i \in \mathcal{I}} \{U_{i,\mathcal{J}_i} \neq \hat{U}_{i,\mathcal{J}_i}\} \right\}$$

where the error is averaged over the random messages and the frozen bits. The code is also called a uniform-input $(n, \mathcal{J}_\mathcal{I}, \varepsilon)$ polar code if the probability of decoding error is no larger than ε.

The following proposition bounds the error probability in terms of Bhattacharyya parameters, and it is a generalization of the well-known result for the special case $N = 1$ (e.g., see [3], Proposition 2). The proof of Proposition 1 can be deduced from Section IV of [14], and is contained in Appendix A for completeness.

Proposition 1. *For the uniform-input $(n, \mathcal{J}_\mathcal{I})$ polar code defined for the MAC $q_{Y|X_\mathcal{I}}$, we have*

$$\mathbb{P}\{\hat{W}_\mathcal{I} \neq W_\mathcal{I}\} \leq \sum_{i=1}^N \sum_{k \in \mathcal{J}_i} Z^{[p_{X_\mathcal{I}}; q_{Y|X_\mathcal{I}}]}(U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n). \tag{33}$$

The following proposition follows from combining Lemma 2, Definition 7 and Proposition 1.

Proposition 2. *There exists a universal constant $t > 0$ such that the following holds. Fix any N-source binary-input MAC characterized by $q_{Y|X_\mathcal{I}}$. Fix any $m \in \mathbb{N}$, let $n = 2^m$ and define*

$$\mathcal{J}_i^{SE} \triangleq \left\{ k \in \{1, 2, \ldots, n\} \,\middle|\, Z^{[p_{X_\mathcal{I}}^*; q_{Y|X_\mathcal{I}}]}(U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n) \leq \frac{1}{n^4} \right\} \tag{34}$$

for each $i \in \mathcal{I}$, where $p_{X_\mathcal{I}}^*$ is the uniform distribution as defined in (24) and the superscript "SE" stands for "scaling exponent". Then, the corresponding uniform-input $(n, \mathcal{J}_\mathcal{I}^{SE})$ polar code satisfies

$$\frac{\sum_{i=1}^N |\mathcal{J}_i^{SE}|}{n} \geq C_{sum} - \frac{tN}{n^{1/\beta}} \tag{35}$$

and

$$\mathbb{P}\{\hat{W}_\mathcal{I} \neq W_\mathcal{I}\} \leq \frac{N}{n^3}. \tag{36}$$

Proof. Let $t > 0$ be the universal constant specified in Lemma 2 and fix an n. For each $i \in \mathcal{I}$, it follows from Lemma 2 and Proposition 1 that

$$\frac{|\mathcal{J}_i^{SE}|}{n} \geq I_{p_{X_\mathcal{I}}^* q_{Y|X_\mathcal{I}}}(X_i; X_{[i-1]}, Y) - \frac{t}{n^{1/\beta}} \tag{37}$$

and

$$\mathbb{P}\{\hat{W}_\mathcal{I} \neq W_\mathcal{I}\} \leq \frac{N}{n^3} \tag{38}$$

for the uniform-input $(n, \mathcal{J}_\mathcal{I}^{SE})$ polar code. Since $p_{X_\mathcal{I}}^* = \prod_{i=1}^N p_{X_i}^*$, it follows that

$$I_{p_{X_\mathcal{I}}^* q_{Y|X_\mathcal{I}}}(X_i; X_{[i-1]}, Y) = I_{p_{X_\mathcal{I}}^* q_{Y|X_\mathcal{I}}}(X_i; Y|X_{[i-1]}) \tag{39}$$

holds for each $i \in \mathcal{I}$, which implies that

$$\sum_{i=1}^N I_{p_{X_\mathcal{I}}^* q_{Y|X_\mathcal{I}}}(X_i; X_{[i-1]}, Y) = I_{p_{X_\mathcal{I}}^* q_{Y|X_\mathcal{I}}}(X_\mathcal{I}; Y). \tag{40}$$

Consequently, (35) follows from (37), (40) and Definition 7, and (36) follows from (38). □

Remark 2. *Proposition 2 shows that the sum-capacity of a binary-input MAC with N sources can be achieved within a gap of $O(Nn^{-1/\beta})$ by using a superposition of N binary-input polar codes.*

4. Problem Formulation of the AWGN Channel and New Polarization Results

4.1. The AWGN Channel

It is well known that appropriately designed polar codes are capacity-achieving for the AWGN channel [13]. The main contribution of this paper is proving an upper bound on the scaling exponent of polar codes for the AWGN channel by using uniform-input polar codes for binary-input MACs described in Definition 8. The following two definitions formally define the AWGN channel and length-*n* codes for the channel.

Definition 11. *An* (n, M, P) *code is an* (n, M) *code described in Definition 1 subject to the additional assumptions that* $\mathcal{X} = \mathbb{R}$ *and the peak power constraint*

$$\mathbb{P}\left\{\frac{1}{n}\sum_{k=1}^{n} X_k^2 \leq P\right\} = 1 \tag{41}$$

is satisfied.

Definition 12. *The* AWGN channel *is a point-to-point memoryless channel described in Definition 2 subject to the additional assumption that* $\mathcal{Y} = \mathbb{R}$ *and* $q_{Y|X}(y|x) = \mathcal{N}(y; x, 1)$ *for all* $x \in \mathbb{R}$ *and* $y \in \mathbb{R}$.

Definition 13. *For an* (n, M, P) *code defined on the AWGN channel, we can calculate according to (9) the average probability of error defined as* $\mathbb{P}\{\hat{W} \neq W\}$. *We call an* (n, M, P) *code with average probability of error no larger than* ε *an* (n, M, P, ε) *code.*

4.2. Uniform-Input Polar Codes for the AWGN Channel

Recall that we would like to use uniform-input polar codes for binary-input MACs described in Definition 8 to achieve the capacity of the AWGN channel, i.e., $C(P)$ in (3). The following definition describes the basic structure of such uniform-input polar codes.

Definition 14. *Fix an* $n = 2^m$ *where* $m \in \mathbb{N}$. *An* $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{avg}$ *polar code with average power P and input alphabet* $\mathcal{A}$ *consists of the following:*

1. *An input alphabet* $\mathcal{A} \subset \mathbb{R} \cup \{0^-\}$ *with* $|\mathcal{A}| = n$ *such that*

$$\frac{1}{n}\sum_{a \in \mathcal{A} \setminus \{0^-\}} a^2 \leq P, \tag{42}$$

where $\mathbb{R} \cup \{0^-\}$ *can be viewed as a line with* 2 *origins. Introducing the symbol* 0^- *allows us to create a set of cardinality n which consists of* $n - 2$ *non-zero real numbers and* 2 *origins* 0 *and* 0^-. *We index each element of* $\mathcal{A}$ *by a unique length-m binary tuple*

$$x_{\mathcal{I}}^{MAC} \triangleq (x_1^{MAC}, x_2^{MAC}, \ldots, x_m^{MAC}), \tag{43}$$

and let $\rho : \{0, 1\}^m \to \mathcal{A}$ *be the bijection that maps the indices to the elements of* $\mathcal{A}$ *such that* $\rho(x_{\mathcal{I}}^{MAC})$ *denotes the element in* $\mathcal{A}$ *indexed by* $x_{\mathcal{I}}^{MAC}$.

2. *A binary-input MAC* $q_{Y|X_{\mathcal{I}}^{MAC}}$ *induced by* $\mathcal{A}$ *as defined through Definitions 5 and 6 with the identifications* $N = m$ *and* $\mathcal{I} = \{1, 2, \ldots, m\}$.

3. *A message set* $\mathcal{W}_i \triangleq \{1, 2, \ldots, 2^{|\mathcal{J}_i|}\}$ *for each* $i \in \{1, 2, \ldots, m\}$, *where* $\mathcal{W}_{\mathcal{I}}$ *is the message alphabet of the uniform-input* $(n, \mathcal{J}_{\mathcal{I}})$ *polar code for the binary-input MAC* $q_{Y|X_{\mathcal{I}}^{MAC}}$ *as defined through Definitions 8 and 9 such that*

$$|\mathcal{W}_{\mathcal{I}}| = \prod_{i=1}^{m} |\mathcal{W}_i| = 2^{\sum_{i=1}^{m} |\mathcal{J}_i|}. \tag{44}$$

In addition, $W_{\mathcal{I}}$ *is uniform on* $\mathcal{W}_{\mathcal{I}}$. *We view the uniform-input* $(n, \mathcal{J}_{\mathcal{I}})$ *polar code as an* $(n, (2^{|\mathcal{J}_1|}, 2^{|\mathcal{J}_2|}, \ldots, 2^{|\mathcal{J}_N|}))$ *code (cf. Remark 1) and let* $\{f_{i,k}^{MAC} \mid i \in \mathcal{I}, k \in \{1, 2, \ldots, n\}\}$ *and* φ^{MAC} *denote the corresponding set of encoding functions and the decoding function respectively (cf. Definition 4).*

4. *An encoding function* $f_k : \mathcal{W}_{\mathcal{I}} \to \mathcal{A}$ *defined as*

$$f_k(W_{\mathcal{I}}) \triangleq \rho(f_{1,k}^{MAC}(W_1), f_{2,k}^{MAC}(W_2), \ldots, f_{m,k}^{MAC}(W_m)) \tag{45}$$

for each $k \in \{1, 2, \ldots, n\}$, where f_k is used for encoding $W_\mathcal{I}$ into X_k such that

$$X_k = \begin{cases} f_k(W_\mathcal{I}) & \text{if } f_k(W_\mathcal{I}) \neq 0^-, \\ 0 & \text{if } f_k(W_\mathcal{I}) = 0^-. \end{cases} \tag{46}$$

Note that both the encoded symbols 0 and 0^- in $\mathcal{A}$ result in the same transmitted symbol $0 \in \mathbb{R}$ according to (46). By construction, $f_1(W_\mathcal{I})$, $f_2(W_\mathcal{I})$, $\ldots$, $f_n(W_\mathcal{I})$ are i.i.d. random variables that are uniformly distributed on $\mathcal{A}$ and hence X_1, X_2, $\ldots$, X_n are i.i.d. real-valued random variables (but not necessarily uniform).

5.   A decoding function $\varphi : \mathbb{R}^n \to \mathcal{W}_\mathcal{I}$ defined as

$$\varphi \triangleq \varphi^{MAC} \tag{47}$$

such that

$$\hat{W}_\mathcal{I} = \varphi(Y^n). \tag{48}$$

Remark 3. *For an $(n, \mathcal{J}_\mathcal{I}, P, \mathcal{A})_{avg}$ polar code, the flexibility of allowing $\mathcal{A}$ to contain 2 origins is crucial to proving the main result of this paper. This is because the input distribution which we will use to establish scaling results for the AWGN channel in Theorem 1 can be viewed as the uniform distribution over some set that contains 2 origins, although the input distribution in the real domain as specified in (52) to follow is not uniform.*

Proposition 3. *There exists a universal constant $t > 0$ such that the following holds. Suppose we are given an $(n, \mathcal{J}_\mathcal{I}^{SE}, P, \mathcal{A})_{avg}$ polar code defined for the AWGN channel $q_{Y|X}$ with a 2-origin $\mathcal{A}$ (i.e., $\mathcal{A} \supseteq \{0, 0^-\}$). Define $\mathcal{X} \triangleq \mathcal{A} \setminus \{0^-\} \subset \mathbb{R}$ where $\mathcal{X}$ contains 1 origin and $n - 2$ non-zero real numbers. Then, the $(n, \mathcal{J}_\mathcal{I}^{SE}, P, \mathcal{A})_{avg}$ polar code is an (n, M)-code (cf. Definition 1) which satisfies*

$$\frac{1}{n} \log M \geq I_{p'_X q_{Y|X}}(X; Y) - \frac{t \log n}{n^{1/\beta}}, \tag{49}$$

$$\mathbb{P}\{\hat{W}_\mathcal{I} \neq W_\mathcal{I}\} \leq \frac{\log n}{n^3}, \tag{50}$$

and

$$\mathbb{P}\{X^n = x^n\} = \prod_{k=1}^{n} \mathbb{P}\{X_k = x_k\} = \prod_{k=1}^{n} p'_X(x_k) \tag{51}$$

for all $x^n \in \mathcal{X}^n$ where p'_X is the distribution on $\mathcal{X}$ defined as

$$p'_X(a) = \begin{cases} \frac{1}{n} & \text{if } a \neq 0, \\ \frac{2}{n} & \text{if } a = 0. \end{cases} \tag{52}$$

Proof. The proposition follows from inspecting Proposition 2 and Definition 14 with the identifications $N = \log n$ and $\log M = \sum_{i=1}^{m} |\mathcal{J}_i|$. $\square$

The following lemma, a strengthened version of Theorem 6 of [13], provides a construction of a good $\mathcal{A}$ which leads to a controllable gap between $C(P)$ and $I_{p'_X q_{Y|X}}(X; Y)$ for the corresponding $(n, \mathcal{J}_\mathcal{I}^{SE}, P, \mathcal{A})_{avg}$ polar code. Although the following lemma is intuitive, the proof is technical and hence relegated to Appendix B.

Lemma 3. *Let $q_{Y|X}$ be the conditional distribution that characterizes the AWGN channel, and fix any $\gamma \in [0,1)$. For each $n = 2^m$ where $m \in \mathbb{N}$, define*

$$\mathcal{D}_n \triangleq \left\{ \frac{\ell}{n} \;\middle|\; \ell \in \left\{1, 2, \ldots, \frac{n}{2}, \ldots n-1\right\} \right\}, \tag{53}$$

define

$$s_X(x) \triangleq \mathcal{N}\left(x; 0, P\left(1 - \frac{1}{n^{(1-\gamma)/\beta}}\right)\right) \tag{54}$$

for all $x \in \mathbb{R}$, define Φ_X to be the cumulative distribution function (cdf) of s_X, and define

$$\mathcal{X} \triangleq \Phi_X^{-1}(\mathcal{D}_n). \tag{55}$$

Note that $\mathcal{X}$ contains 1 origin and $n-2$ non-zero real numbers, and we let p'_X be the distribution on $\mathcal{X}$ as defined in (52). In addition, define the distribution $p'_{X^n}(x^n) \triangleq \prod_{k=1}^n p'_X(x_k)$. Then, there exists a constant $t' > 0$ that depends on P and γ but not n such that the following statements hold for each $n \in \mathbb{N}$:

$$\mathbb{E}_{p'_X}[X^2] \leq P\left(1 - \frac{1}{n^{(1-\gamma)/\beta}}\right), \tag{56}$$

$$\mathbb{P}_{p'_{X^n}}\left\{ \frac{1}{n} \sum_{k=1}^n X_k^2 > P \right\} \leq \frac{e^3}{e^{n^{\frac{1}{2} - \frac{1-\gamma}{\beta}}}}, \tag{57}$$

and

$$C(P) - I_{p'_X q_{Y|X}}(X; Y) \leq \frac{t'\sqrt{\log n}}{n^{(1-\gamma)/\beta}}. \tag{58}$$

A shortcoming of Proposition 3 is that the $(n, \mathcal{J}_{\mathcal{I}}^{\text{SE}}, P, \mathcal{A})_{\text{avg}}$ polar code may not satisfy the peak power constraint (41) and hence it may not qualify as an $(n, 2^{\sum_{i=1}^N |\mathcal{J}_i^{\text{SE}}|}, P)$ code (cf. Definition 11). Therefore, we describe in the following definition a slight modification of an $(n, \mathcal{J}_{\mathcal{I}}^{\text{SE}}, P, \mathcal{A})_{\text{avg}}$ polar code so that the modified polar code always satisfies the peak power constraint (41).

Definition 15. *The 0-power-outage version of an $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{avg}$ polar code is an $(n, 2^{\sum_{i=1}^N |\mathcal{J}_i|})$ code which follows identical encoding and decoding operations of the $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{avg}$ polar code except that the source will modify the input symbol in a time slot k if the following scenario occurs: Let X^n be the desired codeword generated by the source according to the encoding operation of the $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{avg}$ polar code, where the randomness of X^n originates from the information bits $(U_{i, \mathcal{J}_i} : i \in \mathcal{I})$ and the frozen bits $B_{\mathcal{I}, \mathcal{J}_{\mathcal{I}}^c}$. If transmitting the desired symbol X_k at time k results in violating the power constraint $\frac{1}{n}\sum_{\ell=1}^k X_\ell^2 > P$, the source will transmit the symbol 0 at time k instead. An $(n, 2^{\sum_{i=1}^N |\mathcal{J}_i|}, \varepsilon)$ code is called an $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A}, \varepsilon)_{peak}$ polar code if it is the 0-power-outage version of some $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{avg}$ polar code.*

By Definition 15, every $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A}, \varepsilon)_{\text{peak}}$ polar code satisfies the peak power constraint (41) and hence achieves zero power outage, i.e., $\mathbb{P}\{\frac{1}{n}\sum_{k=1}^n X_k^2 > P\} = 0$. Using Definition 15, we obtain the following corollary which states the obvious fact that the probability of power outage of an $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{\text{avg}}$ polar code can be viewed as part of the probability of error of the 0-power-outage version of the code.

Corollary 1. *Given an $(n, \mathcal{J}_{\mathcal{I}}, P, \mathcal{A})_{avg}$-polar code, define*

$$\varepsilon_1 \triangleq \mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \tag{59}$$

and

$$\varepsilon_2 \triangleq \mathbb{P}\left\{\frac{1}{n}\sum_{k=1}^n X_k^2 > P\right\}. \tag{60}$$

Then, the 0-power-outage version of the $(n, \mathcal{J}_\mathcal{I}, P, \mathcal{A})_{avg}$ polar code is an $(n, \mathcal{J}_\mathcal{I}, P, \mathcal{A}, \varepsilon_1 + \varepsilon_2)_{peak}$ polar code that satisfies the peak power constraint (41).

5. Scaling Exponents and Main Result

5.1. Scaling Exponent of Uniform-Input Polar Codes for MACs

We define scaling exponent of uniform-input polar codes for the binary-input MAC as follows.

Definition 16. *Fix an $\varepsilon \in (0, 1)$ and an N-source binary-input MAC $q_{Y|X_\mathcal{I}}$ with symmetric sum-capacity C_{sum} (cf. Definition 7). The* scaling exponent of uniform-input polar codes for the MAC *is defined as*

$$\mu_\varepsilon^{\text{PC-MAC}} \triangleq$$
$$\liminf_{m\to\infty} \inf_{\mathcal{J}_\mathcal{I}} \left\{ \left. \frac{-\log n}{\log\left|C_{\text{sum}} - \frac{\sum_{i=1}^N |\mathcal{J}_i|}{n}\right|} \right| n = 2^m, \text{ there exists a uniform-input } (n, \mathcal{J}_\mathcal{I}, \varepsilon)\text{-polar code on } q_{Y|X_\mathcal{I}} \right\}. \tag{61}$$

Definition 16 formalizes the notion that we are seeking the smallest $\mu \geq 0$ such that $|C_{\text{sum}} - R_n^{\text{MAC}}| = O(n^{-\frac{1}{\mu}})$ holds where $R_n^{\text{MAC}} \triangleq \frac{|\mathcal{J}|}{n}$ denotes the rate of an $(n, \mathcal{J}_\mathcal{I}, \varepsilon)$ polar code. Using the existing results in Section IV-C of [5] and Theorem 2 of [4], we know that

$$3.579 \leq \mu_\varepsilon^{\text{PC-BMSC}} \leq \beta = 4.714 \qquad \forall \varepsilon \in (0, 1) \tag{62}$$

for the special case $N = 1$ where the binary-input MAC is a BMSC. We note from Theorem 48 of [17] (and also [18,19]) that the *optimal* scaling exponent (optimized over all codes) for any non-degenerate discrete memoryless channel (DMC) as well as BMC is equal to 2 for all $\varepsilon \in (0, 1/2)$.

Using Proposition 1 and Definition 16, we obtain the following corollary, which shows that 4.714, the upper bound on $\mu_\varepsilon^{\text{PC-BMSC}}$ in (62) for BMSCs, remains a valid upper bound on the scaling exponent for binary-input MACs.

Corollary 2. *Fix any $\varepsilon \in (0, 1)$ and any binary-input MAC $q_{Y|X_\mathcal{I}}$. Then,*

$$\mu_\varepsilon^{\text{PC-MAC}} \leq \beta = 4.714.$$

5.2. Scaling Exponent of Uniform-Input Polar Codes for the AWGN Channel

Definition 17. *Fix a $P > 0$ and an $\varepsilon \in (0, 1)$. The* scaling exponent of uniform-input polar codes for the AWGN channel *is defined as*

$$\mu_{P,\varepsilon}^{\text{PC-AWGN}} \triangleq$$
$$\liminf_{m\to\infty} \inf_{\mathcal{J}_\mathcal{I}, \mathcal{A}} \left\{ \left. \frac{-\log n}{\log\left|C(P) - \frac{\sum_{i=1}^N |\mathcal{J}_i|}{n}\right|} \right| n = 2^m, \text{ there exists a uniform-input } (n, \mathcal{J}_\mathcal{I}, P, \mathcal{A}, \varepsilon)_{peak}\text{-polar code} \right\}. \tag{63}$$

Definition 17 formalizes the notion that we are seeking the smallest $\mu \geq 0$ such that $|C(P) - R_n^{\text{AWGN}}| = O(n^{-\frac{1}{\mu}})$ holds where $R_n^{\text{AWGN}} \triangleq \frac{\sum_{i=1}^N |\mathcal{J}_i|}{n}$ denotes the rate of an $(n, \mathcal{J}_\mathcal{I}, P, \varepsilon)_{peak}$ polar code. We note from Theorem 54 of [17] and Theorem 5 of [19] that that the *optimal* scaling exponent of the optimal code for the AWGN channel is equal to 2 for any $\varepsilon \in (0, 1/2)$. The following theorem is the main result of this paper, which shows that 4.714 is a valid upper bound on the scaling exponent of polar codes for the AWGN channel.

Theorem 1. *Fix any $P > 0$ and any $\varepsilon \in (0, 1)$. There exists a constant $t^* > 0$ that does not depend on n such that the following holds. For any $n = 2^m$ where $m \in \mathbb{N}$, there exists an $\mathcal{A}$ such that the corresponding $(n, \mathcal{J}_{\mathcal{I}}^{SE}, P, \mathcal{A})_{peak}$ polar code defined for the AWGN channel $q_{Y|X}$ satisfies*

$$\frac{1}{n} \sum_{i=1}^{N} |\mathcal{J}_i^{SE}| \geq \mathrm{C}(P) - \frac{t^* \log n}{n^{1/\beta}} \tag{64}$$

and

$$\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \leq \frac{\log n}{n^3} + \frac{e^3}{e^{n^{\frac{1}{2} - \frac{1}{\beta}}}} . \tag{65}$$

In particular, we have,

$$\mu_{P,\varepsilon}^{PC\text{-}AWGN} \leq \beta = 4.714. \tag{66}$$

Proof. Fix a $P > 0$, an $\varepsilon \in (0, 1)$ and an $n = 2^m$ where $m \in \mathbb{N}$. Combining Proposition 3 and Lemma 3, we conclude that there exists a constant $t^* > 0$ that does not depend on n and an $\mathcal{A}$ such that the corresponding $(n, \mathcal{J}_{\mathcal{I}}^{SE}, P, \mathcal{A})_{avg}$ polar code defined for the AWGN channel $q_{Y|X}$ satisfies (64),

$$\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \leq \frac{\log n}{n^3} , \tag{67}$$

and

$$\mathbb{P}\left\{ \frac{1}{n} \sum_{k=1}^{n} X_k^2 > P \right\} \leq \frac{e^3}{e^{n^{\frac{1}{2} - \frac{1}{\beta}}}} . \tag{68}$$

Using (67), (68) and Corollary 1, we conclude that the $(n, \mathcal{J}_{\mathcal{I}}^{SE}, P, \mathcal{A})_{avg}$ polar code is an $(n, \mathcal{J}_{\mathcal{I}}^{SE}, P, \mathcal{A})_{peak}$ polar code that satisfies (64) and (65). Since

$$\frac{\log n}{n^3} + \frac{e^3}{e^{n^{\frac{1}{2} - \frac{1}{\beta}}}} \leq \varepsilon \tag{69}$$

for all sufficiently large n, it follows from (64), (65) and Definition 17 that (66) holds. $\square$

6. Moderate Deviations Regime

6.1. Polar Codes That Achieve the Symmetric Capacity of a BMC

The following result is based on Section IV of [4], which developed a tradeoff between the gap to capacity and the decay rate of the error probability for a BMC under the moderate deviations regime [15] where both the gap to capacity and the error probability vanish as n grows.

Lemma 4. *([4], Section IV) There exists a universal constant $t_{MD} > 0$ such that the following holds. Fix any $\gamma \in \left(\frac{1}{1+\beta}, 1 \right)$ and any BMC characterized by $q_{Y|X}$. Recall that p_X^* denotes the uniform distribution on $\{0, 1\}$. Then for any $n = 2^m$ where $m \in \mathbb{N}$, we have*

$$\frac{1}{n} \left| \left\{ k \in \{1, 2, \ldots, n\} \, \left| \, Z^{[p_X^*; q_{Y|X}]}(U_k | U^{k-1}, Y^n) \leq 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma \beta + \gamma - 1}{\gamma \beta} \right)}} \right. \right\} \right| \geq I_{p_X^* q_{Y|X}}(X; Y) - \frac{t_{MD}}{n^{(1-\gamma)/\beta}} \tag{70}$$

where $h_2 : [0, 1/2] \to [0, 1]$ denotes the binary entropy function.

6.2. Polar Codes that Achieve the Symmetric Sum-Capacity of a Binary-Input MAC

The following lemma, whose proof is omitted because it is analogous to the proof of Lemma 2, is a direct consequence of Lemma 4.

Lemma 5. *There exists a universal constant $t_{MD} > 0$ such that the following holds. Fix any $\gamma \in \left(\frac{1}{1+\beta}, 1\right)$ and any binary-input MAC characterized by $q_{Y|X_{\mathcal{I}}}$. Recall that $p^*_{X_{\mathcal{I}}} = \prod_{i \in \mathcal{I}} p^*_{X_i}$. Then for any $n = 2^m$ where $m \in \mathbb{N}$, we have*

$$
\frac{1}{n}\left|\left\{k \in \{1, 2, \ldots, n\} \left| Z^{[p^*_{X_{\mathcal{I}}}; q_{Y|X_{\mathcal{I}}}]}(U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n) \leq 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}}\right.\right\}\right|
$$
$$
\geq I_{p^*_{X_{\mathcal{I}}} q_{Y|X_{\mathcal{I}}}}(X_i; X_{[i-1]}, Y) - \frac{t_{MD}}{n^{(1-\gamma)/\beta}} \tag{71}
$$

for each $i \in \mathcal{I}$.

Combining Lemma 5, Definition 7 and Proposition 1, we obtain the following proposition, whose proof is analogous to the proof of Proposition 2 and hence omitted.

Proposition 4. *There exists a universal constant $t_{MD} > 0$ such that the following holds. Fix any $\gamma \in \left(\frac{1}{1+\beta}, 1\right)$ and any N-source binary-input MAC characterized by $q_{Y|X_{\mathcal{I}}}$. In addition, fix any $m \in \mathbb{N}$, let $n = 2^m$ and define*

$$
\mathcal{J}_i^{MD} \triangleq \left\{k \in \{1, 2, \ldots, n\} \left| Z^{[p_{X_{\mathcal{I}}}; q_{Y|X_{\mathcal{I}}}]}(U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n) \leq 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}}\right.\right\} \tag{72}
$$

for each $i \in \mathcal{I}$ where the superscript "MD" stands for "moderate deviations". Then, the corresponding uniform-input $(n, \mathcal{J}_{\mathcal{I}}^{MD})$ polar code described in Definition 9 satisfies

$$
\frac{\sum_{i=1}^{N} \left|\mathcal{J}_i^{MD}\right|}{n} \geq C_{sum} - \frac{t_{MD}N}{n^{(1-\gamma)/\beta}} \tag{73}
$$

and

$$
\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \leq Nn\, 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}}. \tag{74}
$$

6.3. Uniform-Input Polar Codes for the AWGN Channel

Proposition 5. *There exists a universal constant $t_{MD} > 0$ such that the following holds. Fix any $\gamma \in \left(\frac{1}{1+\beta}, 1\right)$. Suppose we are given an $(n, \mathcal{J}_{\mathcal{I}}^{MD}, P, \mathcal{A})_{avg}$ polar code (cf. Definition 14) defined for the AWGN channel $q_{Y|X}$ with a 2-origin $\mathcal{A}$ (i.e., $\mathcal{A} \supseteq \{0, 0^-\}$). Define $\mathcal{X} \triangleq \mathcal{A} \setminus \{0^-\} \subset \mathbb{R}$ where $\mathcal{X}$ contains 1 origin and $n - 2$ non-zero real numbers. Then, the $(n, \mathcal{J}_{\mathcal{I}}^{MD}, P, \mathcal{A})_{avg}$ polar code is an (n, M)-code (cf. Definition 1) which satisfies*

$$
\frac{1}{n}\log M \geq I_{p'_X q_{Y|X}}(X; Y) - \frac{t_{MD}\log n}{n^{(1-\gamma)/\beta}}, \tag{75}
$$

$$
\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \leq (\log n)n\, 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}}, \tag{76}
$$

and

$$\mathbb{P}\left\{X^n = x^n\right\} = \prod_{k=1}^{n} \mathbb{P}\left\{X_k = x_k\right\} = \prod_{k=1}^{n} p'_X(x_k) \tag{77}$$

for all $x^n \in \mathcal{X}^n$ where p'_X is the distribution on $\mathcal{X}$ as defined in (52).

Proof. The proposition follows from inspecting Proposition 4 and Definition 14 with the identifications $N = \log n$ and $\log M = \sum_{i=1}^{m} |\mathcal{J}_i^{\text{MD}}|$. □

The following theorem develops the tradeoff between the gap to capacity and the decay rate of the error probability for $(n, \mathcal{J}_{\mathcal{I}}^{\text{MD}}, P, \mathcal{A})_{\text{peak}}$ polar codes defined for the AWGN channel.

Theorem 2. *Fix a $\gamma \in \left(\frac{1}{1+\beta}, 1\right)$. There exists a constant $t_{MD}^* > 0$ that depends on P and γ but not n such that the following holds for any $n = 2^m$ where $m \in \mathbb{N}$. There exists an $(n, \mathcal{J}_{\mathcal{I}}^{MD}, P, \mathcal{A}, \varepsilon)_{peak}$ polar code defined for the AWGN channel $q_{Y|X}$ that satisfies,*

$$\frac{1}{n}\sum_{i=1}^{N} |\mathcal{J}_i^{MD}| \geq C(P) - \frac{t_{MD}^* \log n}{n^{(1-\gamma)/\beta}}, \tag{78}$$

and,

$$\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \leq (n\log n + e^3)2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}}. \tag{79}$$

Proof. By Proposition 5 and Lemma 3, there exists a constant $t_{\text{MD}}^* > 0$ that depends on P and γ but not n such that for any $n = 2^m$ where $m \in \mathbb{N}$, there exists an $\mathcal{A}$ and the corresponding $(n, \mathcal{J}_{\mathcal{I}}^{\text{MD}}, P, \mathcal{A})_{\text{avg}}$ polar code that satisfies (78),

$$\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} \leq (\log n)n\, 2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}}, \tag{80}$$

and

$$\mathbb{P}\left\{\frac{1}{n}\sum_{k=1}^{n} X_k^2 > P\right\} \leq \frac{e^3}{e^{n^{\frac{1}{2} - \frac{1-\gamma}{\beta}}}}. \tag{81}$$

Equation (79) remains to be shown. Using (80), (81) and Corollary 1, we conclude that the $(n, \mathcal{J}_{\mathcal{I}}^{\text{MD}}, P, \mathcal{A})_{\text{avg}}$ polar code is an $(n, \mathcal{J}_{\mathcal{I}}^{\text{MD}}, P, \mathcal{A}, \varepsilon)_{\text{peak}}$ polar code that satisfies,

$$\varepsilon \triangleq (n\log n)2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}} + \frac{e^3}{e^{n^{\frac{1}{2} - \frac{1-\gamma}{\beta}}}} \tag{82}$$

$$\leq (n\log n + e^3)2^{-n^{\gamma h_2^{-1}\left(\frac{\gamma\beta+\gamma-1}{\gamma\beta}\right)}} \tag{83}$$

where the inequality follows from the fact that $h_2(x) \geq 2x$ for all $x \in [0, 1/2]$. This concludes the proof. □

Remark 4. *A candidate of $\mathcal{A}$ in Theorem 2 can been explicitly constructed according to Lemma 3 with the identification $\mathcal{A} \equiv \mathcal{X} \cup \{0^-\}$.*

7. Concluding Remarks

In this paper, we provided an upper bound on the scaling exponent of polar codes for the AWGN channel (Theorem 1). In addition, in Theorem 2 we have shown a moderate deviations result—namely, the existence of polar codes which obey a certain tradeoff between the gap to capacity and the decay rate of the error probability for the AWGN channel.

Since the encoding and decoding complexities of the binary-input polar code for a BMC are $O(n \log n)$ as long as we allow pseudorandom numbers to be shared between the encoder and the decoder for encoding and decoding the randomized frozen bits (e.g., see [3], Section IX), the encoding and decoding complexities of the polar codes for the AWGN channel defined in Definition 14 and Definition 15 are $O(n \log n) \times \log n = O(n \log^2 n)$. By a standard probabilistic argument, there must exist a deterministic encoder for the frozen bits such that the decoding error of the polar code for the AWGN channel with the deterministic encoder is no worse than the polar code with randomized frozen bits. In the future, it may be fruitful to develop low-complexity algorithms for finding a good deterministic encoder for encoding the frozen bits. Another interesting direction for future research is to compare the empirical performance between our polar codes in Definitions 14 and 15 and the state-of-the-art polar codes. One may also explore various techniques (e.g., list decoding, cyclic redundancy check (CRC), etc.) to improve the empirical performance of the polar codes constructed herein.

Author Contributions: S. L. Fong carried out the research and wrote the paper. V. Y. F. Tan proposed the research topic, suggested and improved the flow of the presentation in the paper, and verified the correctness of the results.

Appendix A. Proof of Proposition 1

Unless specified otherwise, all the probabilities in this proof are evaluated according to the distribution induced by the uniform-input $(n, \mathcal{J}_{\mathcal{I}})$ polar code. Consider

$$\mathbb{P}\{\hat{W}_{\mathcal{I}} \neq W_{\mathcal{I}}\} = \sum_{i=1}^{N} \mathbb{P}\{\{\hat{W}_i \neq W_i\} \cap \{\hat{W}_{[i-1]} = W_{[i-1]}\}\} \tag{A1}$$

$$= \sum_{i=1}^{N} \mathbb{P}\{\{\hat{U}_i^n \neq U_i^n\} \cap \{\hat{X}_{[i-1]}^n = X_{[i-1]}^n\}\} \tag{A2}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{n} \mathbb{P}\{\{\hat{U}_{i,k} \neq U_{i,k}\} \cap \{\hat{U}_i^{k-1} = U_i^{k-1}\} \cap \{\hat{X}_{[i-1]}^n = X_{[i-1]}^n\}\} \tag{A3}$$

where (A2) is due to Definition 8 where $\{\hat{W}_i = W_i\} = \{\hat{U}_i^n = U_i^n\}$ and $\{\hat{W}_{[i-1]} = W_{[i-1]}\} = \{\hat{X}_{[i-1]}^n = X_{[i-1]}^n\}$. For each $i \in \mathcal{I}$ and each $k \in \mathcal{J}_i$, we have

$$\mathbb{P}\{\{\hat{U}_{i,k} \neq U_{i,k}\} \cap \{\hat{U}_i^{k-1} = U_i^{k-1}\} \cap \{\hat{X}_{[i-1]}^n = X_{[i-1]}^n\}\}$$

$$\leq \sum_{u_{i,k} \in \{0,1\}} \sum_{u_i^{k-1} \in \mathcal{U}_i^{k-1}} \sum_{x_{[i-1]}^n \in \mathcal{X}_{[i-1]}^n} \int_{\mathcal{Y}^n} p_{U_i^k, X_{[i-1]}^n, Y^n}(u_i^k, x_{[i-1]}^n, y^n)$$

$$\times \mathbf{1}\left\{p_{U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n}(u_{i,k}|u_i^{k-1}, x_{[i-1]}^n, y^n) \leq p_{U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n}(1 - u_{i,k}|u_i^{k-1}, x_{[i-1]}^n, y^n)\right\} \mathrm{d}y^n \tag{A4}$$

$$\leq 2 \sum_{u_i^{k-1} \in \mathcal{U}_i^{k-1}} \sum_{x_{[i-1]}^n \in \mathcal{X}_{[i-1]}^n} \int_{\mathcal{Y}^n} p_{U_i^{k-1}, X_{[i-1]}^n, Y^n}(u_i^{k-1}, x_{[i-1]}^n, y^n)$$

$$\times \sqrt{p_{U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n}(0|u_i^{k-1}, x_{[i-1]}^n, y^n) p_{U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n}(1|u_i^{k-1}, x_{[i-1]}^n, y^n)} \mathrm{d}y^n \tag{A5}$$

$$= Z^{[p_{X_{\mathcal{I}}}; q_{Y|X_{\mathcal{I}}}]}(U_{i,k}|U_i^{k-1}, X_{[i-1]}^n, Y^n) \tag{A6}$$

where (A4) follows from (30). In addition, it follows from (27) and (30) that

$$\mathbb{P}\{\{\hat{U}_{i,k} \neq U_{i,k}\} \cap \{\hat{U}_i^{k-1} = U_i^{k-1}\} \cap \{\hat{X}_{[i-1]}^n = X_{[i-1]}^n\}\} = 0 \tag{A7}$$

for each $i \in \mathcal{I}$ and each $k \in \mathcal{J}_i^c$. Combining (A3), (A6) and (A7), we obtain (33).

Appendix B. Proof of Lemma 3

Let $q_{Y|X}$ be the conditional distribution that characterizes the AWGN channel and fix a $\gamma \in [0,1)$. Recall the definitions of p'_X and s_X in (52) and (54) respectively and recall that Φ_X is the cdf of s_X. Fix a sufficiently large $n \geq 36$ that satisfies

$$\frac{1}{n^{(1-\gamma)/\beta}} \in \left(0, \frac{1+P}{2}\right] \tag{A8}$$

and

$$\Phi_X^{-1}\left(\frac{1}{n^{1-(1-\gamma)/\beta}}\right) \leq -1. \tag{A9}$$

In addition, recall the definition of $\mathcal{X}$ in (55) and let $g : \mathbb{R} \to \mathcal{X}$ be a quantization function such that,

$$g(t) \triangleq \Phi_X^{-1}\left(\frac{\ell}{n}\right) \tag{A10}$$

where $\ell \in \{1, 2, \ldots, \frac{n}{2}, \ldots n-1\}$ is the unique integer that satisfies

$$\begin{cases} \Phi_X^{-1}\left(\frac{\ell}{n}\right) \leq t < \Phi_X^{-1}\left(\frac{\ell+1}{n}\right) & \text{if } t \geq 0, \\ \Phi_X^{-1}\left(\frac{\ell-1}{n}\right) < t \leq \Phi_X^{-1}\left(\frac{\ell}{n}\right) & \text{if } t < 0. \end{cases} \tag{A11}$$

In words, g quantizes every $a \in \mathbb{R}$ to its nearest point in $\mathcal{X}$ whose magnitude is smaller than $|a|$. Let

$$\hat{X} \triangleq g(X) \tag{A12}$$

be the quantized version of X. By construction,

$$\mathbb{P}_{s_X}\left\{|\hat{X}| \leq |X|\right\} = 1, \tag{A13}$$

$$\mathbb{P}_{s_X}\left\{\left|\Phi_X(\hat{X}) - \Phi_X(X)\right| \leq \frac{1}{n}\right\} = 1 \tag{A14}$$

and

$$\mathbb{P}_{s_X}\left\{X \in \left(\Phi_X^{-1}\left(\frac{\ell}{n}\right), \Phi_X^{-1}\left(\frac{\ell+1}{n}\right)\right)\right\} = \frac{1}{n} \tag{A15}$$

for all $\ell \in \{0, 1, \ldots, n-1\}$. It follows from (A15) and the definition of p'_X in (52) that

$$\mathrm{E}_{p'_X}[X^2] = \mathrm{E}_{s_X}[\hat{X}^2] \tag{A16}$$

and

$$\mathbb{P}_{p'_{X^n}}\left\{\frac{1}{n}\sum_{k=1}^{n} X_k^2 > P\right\} = \mathbb{P}_{s_{X^n}}\left\{\frac{1}{n}\sum_{k=1}^{n} \hat{X}_k^2 > P\right\} \tag{A17}$$

where $s_{X^n}(x^n) \triangleq \prod_{k=1}^{n} s_X(x_k)$. Consequently, in order to show (56) and (57), it suffices to show

$$\mathrm{E}_{s_X}[\hat{X}^2] \leq P\left(1 - \frac{1}{n^{(1-\gamma)/\beta}}\right) \tag{A18}$$

and

$$\mathbb{P}_{s_{X^n}}\left\{\frac{1}{n}\sum_{k=1}^n \hat{X}_k^2 > P\right\} \le \frac{e^3}{e^{n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}}} \tag{A19}$$

respectively. Using (A13) and the definition of s_X in (54), we obtain (A18). In order to show (A19), we consider the following chain of inequalities:

$$\mathbb{P}_{s_{X^n}}\left\{\frac{1}{n}\sum_{k=1}^n \hat{X}_k^2 > P\right\}$$

$$\le \mathbb{P}_{s_{X^n}}\left\{\frac{1}{n}\sum_{k=1}^n X_k^2 > P\right\} \tag{A20}$$

$$= \mathbb{P}_{s_{X^n}}\left\{\frac{\sum_{k=1}^n X_k^2}{\sqrt{n}P\left(1-\frac{1}{n^{(1-\gamma)/\beta}}\right)} > \sqrt{n} + \frac{n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}}{1-\frac{1}{n^{(1-\gamma)/\beta}}}\right\} \tag{A21}$$

$$\le \mathbb{P}_{s_{X^n}}\left\{e^{\frac{\sum_{k=1}^n X_k^2}{\sqrt{n}P\left(1-\frac{1}{n^{(1-\gamma)/\beta}}\right)}} > e^{\left(\sqrt{n}+n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}\right)}\right\} \tag{A22}$$

$$\le \left(1-\frac{1}{\sqrt{n}/2}\right)^{-n/2} e^{-\left(\sqrt{n}+n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}\right)} \tag{A23}$$

$$= \left(1+\frac{1}{\sqrt{n}/2-1}\right)^{n/2} e^{-\left(\sqrt{n}+n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}\right)} \tag{A24}$$

$$\le e^{\frac{n}{\sqrt{n}-2}} \cdot e^{-\left(\sqrt{n}+n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}\right)} \tag{A25}$$

$$= e^{\frac{2\sqrt{n}}{\sqrt{n}-2}} \cdot e^{-n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}} \tag{A26}$$

$$\le e^3 \cdot e^{-n^{\frac{1}{2}-\frac{1-\gamma}{\beta}}} \tag{A27}$$

where

- (A20) is due to (A13).
- (A23) is due to Markov's inequality.
- (A25) is due to the fact that $(1+\frac{1}{t})^t \le e$ for all $t > 0$.
- (A27) is due to the assumption that $\sqrt{n} \ge 6$.

Equation (58) remains to be shown. To this end, we let q_Z denote the distribution of the standard normal random variable (cf. (1)) and consider

$$C\left(P - \frac{1}{n^{(1-\gamma)/\beta}}\right) - I_{p_X' q_{Y|X}}(X;Y)$$

$$= C\left(P - \frac{1}{n^{(1-\gamma)/\beta}}\right) - I_{p_X' q_Z}(X; X+Z) \tag{A28}$$

$$= I_{s_X q_Z}(X; X+Z) - I_{p_X' q_Z}(X; X+Z) \tag{A29}$$

where (A29) is due to the definition of s_X in (54). In order to simplify the right hand side of (A29), we invoke Corollary 4 of [20] and obtain

$$I_{s_X q_Z}(X; X+Z) - I_{p_X' q_Z}(X; X+Z) \le (\log e)(3\sqrt{1+P}+4\sqrt{P})W_2(s_Y, p_Y'). \tag{A30}$$

After some tedious calculations which will be elaborated after this proof, it can be shown that the Wasserstein distance in (A30) satisfies

$$W_2(s_Y, p'_Y) \leq \sqrt{\frac{\kappa \log n}{n^{\frac{2(1-\gamma)}{\beta}}}}, \tag{A31}$$

where

$$\kappa \triangleq P^2 + 4P + \frac{4P}{\log e}\left(1 + \log\sqrt{\frac{P}{2\pi}}\right). \tag{A32}$$

On the other hand, since

$$\frac{\log(1 + P - \xi)}{\log e} \geq \frac{\log(1 + P)}{\log e} - \frac{\xi}{1 + P} - \frac{2\xi^2}{(1 + P)^2} \tag{A33}$$

for each $\xi \in \left(0, \frac{1+P}{2}\right]$ by Taylor's theorem and $\frac{1}{n^{(1-\gamma)/\beta}} \in \left(0, \frac{1+P}{2}\right]$ by (A8), we have

$$C\left(P - \frac{1}{n^{\frac{1-\gamma}{\beta}}}\right) \geq C(P) - \frac{\log e}{2}\left(\frac{1}{n^{\frac{1-\gamma}{\beta}}(1+P)} + \frac{2}{n^{\frac{2(1-\gamma)}{\beta}}(1+P)^2}\right). \tag{A34}$$

Using (A29), (A30), (A31) and (A34), we obtain

$$C(P) - I_{p'_X q_{Y|X}}(X; Y)$$

$$\leq (\log e)(3\sqrt{1+P} + 4\sqrt{P})\sqrt{\frac{\kappa \log n}{n^{2(1-\gamma)/\beta}}} + \frac{\log e}{2}\left(\frac{1}{n^{\frac{1-\gamma}{\beta}}(1+P)} + \frac{2}{n^{\frac{2(1-\gamma)}{\beta}}(1+P)^2}\right). \tag{A35}$$

Consequently, (58) holds for some constant $t' > 0$ that does not depend on n.

Derivation of (A31)

Consider the distribution (coupling) $r_{X,\hat{X},Y,\hat{Y}}$ defined as

$$r_{X,\hat{X},Y,\hat{Y}}(x, \hat{x}, y, \hat{y}) = s_X(x)q_Z(y - x)\mathbf{1}\{\hat{x} = g(x)\}\mathbf{1}\{\hat{y} = g(x) + y - x\} \tag{A36}$$

and simplify the Wasserstein distance in (A30) as follows:

$$(W_2(s_Y, p'_Y))^2 \leq \int_{\mathbb{R}}\int_{\mathbb{R}} r_{Y,\hat{Y}}(y, \hat{y})(y - \hat{y})^2 \mathrm{d}\hat{y}\mathrm{d}y \tag{A37}$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} r_{X,\hat{X}}(x, \hat{x})(x - \hat{x})^2 \mathrm{d}\hat{x}\mathrm{d}x \tag{A38}$$

$$= \int_{\mathbb{R}} s_X(x)(x - g(x))^2 \mathrm{d}x \tag{A39}$$

where

- (A37) follows from the definition of W_2 in (6) and the fact due to (A36) that $r_Y = s_Y$ and $r_{\hat{Y}} = p'_Y$.
- (A38) follows from the fact due to (A36) that $\mathbb{P}_{r_{X,\hat{X},Y,\hat{Y}}}\{Y - \hat{Y} = X - \hat{X}\} = 1$.
- (A39) is due to (A36).

Following (A39), we define ξ_n to be the positive number that satisfies

$$\int_{\xi_n}^{\infty} s_X(x)\mathrm{d}x = \frac{1}{n^{1-(1-\gamma)/\beta}} \tag{A40}$$

and consider

$$
\int_{\mathbb{R}} s_X(x)(x - g(x))^2 dx
$$

$$
= \int_{-\infty}^{-\xi_n} s_X(x)(x - g(x))^2 dx + \int_{-\xi_n}^{\xi_n} s_X(x)(x - g(x))^2 dx + \int_{\xi_n}^{\infty} s_X(x)(x - g(x))^2 dx \tag{A41}
$$

$$
\leq \int_{-\infty}^{-\xi_n} s_X(x)x^2 dx + \int_{-\xi_n}^{\xi_n} s_X(x)(x - g(x))^2 dx + \int_{\xi_n}^{\infty} s_X(x)x^2 dx \tag{A42}
$$

$$
= 2\int_{\xi_n}^{\infty} s_X(x)x^2 dx + \int_{-\xi_n}^{\xi_n} s_X(x)(x - g(x))^2 dx \tag{A43}
$$

where (A42) follows from the fact due to (A10) that $x - g(x) \geq 0$ for all $x \in \mathbb{R}$. In order to bound the first term in (A43), we let

$$
P_n \triangleq P\left(1 - \frac{1}{n^{(1-\gamma)/\beta}}\right) \tag{A44}
$$

and consider

$$
\int_{\xi_n}^{\infty} s_X(x)x^2 dx = P_n\left(\int_{\xi_n}^{\infty} s_X(x)dx + \xi_n s_X(\xi_n)\right) \tag{A45}
$$

$$
< \int_{\xi_n}^{\infty} s_X(x)dx \left(2P_n + \xi_n^2\right) \tag{A46}
$$

$$
< \frac{1}{n^{1-(1-\gamma)/\beta}}\left(2P + \xi_n^2\right) \tag{A47}
$$

where

- (A45) follows from integration by parts.
- (A46) is due to the simple fact that

$$
\int_{\xi_n}^{\infty} s_X(x)dx > \frac{\xi_n^2}{P_n + \xi_n^2}\int_{\xi_n}^{\infty}\left(1 + \frac{P_n}{x^2}\right)s_X(x)dx \tag{A48}
$$

$$
= \left(\frac{P_n}{P_n + \xi_n^2}\right)\xi_n s_X(\xi_n). \tag{A49}
$$

- (A47) is due to (A40) and (A44).

  In order to bound the term in (A47), we note that

$$
\xi_n \geq 1 \tag{A50}
$$

by (A9) and (A40) and would like to obtain an upper bound on ξ_n through the following chain of inequalities:

$$
\frac{1}{n^{1-(1-\gamma)/\beta}} = \int_{\xi_n}^{\infty} s_X(x)dx \tag{A51}
$$

$$
\leq \int_{\xi_n}^{\infty} x s_X(x)dx \tag{A52}
$$

$$
= P_n s_X(\xi_n) \tag{A53}
$$

$$
= \sqrt{\frac{P_n}{2\pi}}e^{-\frac{\xi_n^2}{2P_n}} \tag{A54}
$$

$$
< \sqrt{\frac{P}{2\pi}}e^{-\frac{\xi_n^2}{2P}} \tag{A55}
$$

where

- (A51) is due to (A40).
- (A52) is due to (A50).

Since

$$\xi_n^2 \le \frac{2P}{\log e}\left(\left(1 - \frac{1-\gamma}{\beta}\right)\log n + \log \sqrt{\frac{P}{2\pi}}\right) \tag{A56}$$

by (A55) and $\beta = 4.714 > 3$, it follows from (A47) that

$$\int_{\xi_n}^{\infty} s_X(x)x^2 \mathrm{d}x < \frac{1}{n^{1-\frac{1-\gamma}{\beta}}}\left(2P + \frac{2P}{\log e}\left(\left(1 - \frac{1-\gamma}{\beta}\right)\log n + \log \sqrt{\frac{P}{2\pi}}\right)\right) \tag{A57}$$

$$< \frac{\log n}{n^{\frac{2(1-\gamma)}{\beta}}}\left(2P + \frac{2P}{\log e}\left(1 + \log \sqrt{\frac{P}{2\pi}}\right)\right). \tag{A58}$$

In order to bound the second term in (A43), we consider

$$\int_{-\xi_n}^{\xi_n} s_X(x)(x - g(x))^2 \mathrm{d}x$$

$$= \int_{-\xi_n}^{\xi_n} s_X(x)\left(\Phi^{-1}\big(\Phi(x)\big) - \Phi^{-1}\big(\Phi(g(x))\big)\right)^2 \mathrm{d}x \tag{A59}$$

$$\le \int_{-\xi_n}^{\xi_n} s_X(x)\left(\frac{1}{n} \times \frac{1}{s_X(\xi_n)}\right)^2 \mathrm{d}x \tag{A60}$$

$$\le \int_{-\xi_n}^{\xi_n} s_X(x)\left(\frac{P_n^2}{n^{\frac{2(1-\gamma)}{\beta}}}\right) \mathrm{d}x \tag{A61}$$

$$< \frac{P^2}{n^{\frac{2(1-\gamma)}{\beta}}} \tag{A62}$$

where

- (A60) is due to (A14), the mean value theorem and the fact that the derivative of Φ is always positive and uniformly bounded below by $s_X(\xi_n)$ on the interval $[-\xi_n, \xi_n]$.
- (A61) is due to (A53).

  Combining (A39), (A43), (A58) and (A62) and recalling the definition of κ in (A32), we obtain (A31).

References

1. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley and Sons Inc.: Hoboken, NJ, USA, 2006.
2. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
3. Arıkan, E. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory* **2009**, *55*, 1–23.
4. Mondelli, M.; Hassani, S.H.; Urbanke, R. Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors. *IEEE Trans. Inf. Theory* **2016**, *62*, 6698–6712.
5. Hassani, S.H.; Alishahi, K.; Urbanke, R. Finite-length scaling for polar codes. *IEEE Trans. Inf. Theory* **2014**, *60*, 5875–5898.
6. Goldin, D.; Burshtein, D. Improved Bounds on the Finite Length Scaling of Polar Codes. *IEEE Trans. Inf. Theory* **2014**, *60*, 6966–6978.
7. Fong, S.L.; Tan, V.Y.F. On the Scaling Exponent of Polar Codes for Binary-Input Energy-Harvesting Channels. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 3540–3551.
8. Mahdavifar, H. Fast Polarization and Finite-Length Scaling for Non-Stationary Channels. *arXiv* **2016**, arXiv:1611.04203.

9.   Şaşoğlu, E.; Telatar, I.; Arıkan, E. Polarization of arbitrary discrete memoryless channels. In Proceedings of the 2009 IEEE Information Theory Workshop, Taormina, Italy, 11–16 October 2009; pp. 114–118.

10.   Sutter, D.; Renes, J.M.; Dupuis, F.; Renner, R. Achieving the capacity of any DMC using only polar codes. In Proceedings of the 2012 IEEE Information Theory Workshop, Lausanne, Switzerland, 3–7 September 2012; pp. 114–118.

11.   Honda, J.; Yamamoto, H. Polar coding without alphabet extension for asymmetric models. *IEEE Trans. Inf. Theory* **2013**, *59*, 7829–7838.

12.   Mondelli, M.; Urbanke, R.; Hassani, S.H. How to Achieve the Capacity of Asymmetric Channels. In Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 30 September–3 October 2014; pp. 789–796.

13.   Abbe, E.; Barron, A. Polar coding schemes for the AWGN channel. In Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings, St. Petersburg, Russia, 31 July–5 August 2011; pp. 194–198.

14.   Abbe, E.; Telatar, E. Polar Codes for the *m*-User Multiple Access Channel. *IEEE Trans. Inf. Theory* **2012**, *58*, 5437–5448.

15.   Altuǧ, Y.; Wagner, A.B. Moderate Deviations in Channel Coding. *IEEE Trans. Inf. Theory* **2014**, *20*, 4417–4426.

16.   Mahdavifar, H.; El-Khamy, M.; Lee, J.; Kang, I. Achieving the Uniform Rate Region of General Multiple Access Channels by Polar Coding. *IEEE Trans. Commun.* **2016**, *64*, 467–478.

17.   Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359.

18.   Strassen, V. Asymptotische Abschätzungen in Shannons Informationstheorie. *Trans. Third Prague Conf. Inf. Theory* **1962**, 689–723.

19.   Hayashi, M. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. Inf. Theory* **2009**, *55*, 4947–4966.

20.   Polyanskiy, Y.; Wu, Y. Wasserstein Continuity of Entropy and Outer Bounds for Interference Channels. *IEEE Trans. Inf. Theory* **2016**, *62*, 3992–4002.