Unified Scaling of Polar Codes: Error Exponent, Scaling Exponent, Moderate Deviations, and Error Floors

Marco Mondelli, S. Hamed Hassani, and Rüdiger Urbanke

Abstract

Consider transmission of a polar code of block length N and rate R over a binary memoryless symmetric channel W and let P_{e} be the error probability under successive cancellation decoding. In this paper, we develop new bounds that characterize the relationship among the parameters R, N, P_{e}, and the quality of the channel W quantified by its capacity $I(W)$ and its Bhattacharyya parameter $Z(W)$.

In previous works, two main regimes have been studied. In the *error exponent* regime, the channel W and the rate $R < I(W)$ are fixed, and it has been proved that the error probability P_{e} scales roughly as $2^{-\sqrt{N}}$. In the *scaling exponent* approach, the channel W and the error probability P_{e} are fixed and it has been proved that the gap to capacity $I(W) - R$ scales as $N^{-1/\mu}$. Here, μ is called *scaling exponent* and this scaling exponent depends on the channel W. A heuristic computation for the binary erasure channel (BEC) gives $\mu = 3.627$ and it has been shown that, for any channel W, $3.579 \leq \mu \leq 5.702$.

The original contributions of this paper are as follows. First, we provide the tigher upper bound $\mu \leq 4.714$ valid for any W. With the same technique, we obtain an upper bound for the scaling exponent of the BEC which very closely approaches its heuristically derived value $\mu = 3.639$.

Secondly, we develop a trade-off between the gap to capacity $I(W) - R$ and the error probability P_{e}, as functions of the block length N. In other words, we neither fix the gap to capacity (error exponent regime) nor the error probability (scaling exponent regime), but we consider a *moderate deviations* regime in which we study how fast both quantities simultaneously go to 0 as functions of the block length N.

Thirdly, we prove that polar codes are not affected by *error floors*. To do so, we fix a polar code of block length N and rate R. We then vary the channel W and study the impact of this variation on the error probability. We show that the error probability P_{e} scales as the Bhattacharyya parameter $Z(W)$ raised to a power which scales like $\sqrt{N}$. This agrees with the scaling in the error exponent regime.

I. Introduction

Performance Analysis in Different Regimes. When considering the transmission using a coding scheme over a channel W, the parameters of interest are the rate R, which represents the amount of information transmitted per channel use, the block length N, which represents the total number of channel uses, and the block error probability P_{e}. The exact characterization of the relationship between R, N, P_{e}, and the quality of the channel W (which can be quantified, e.g., by its capacity $I(W)$ or its Bhattacharyya parameter $Z(W)$) is a formidable task. It is easier to study the *scaling* of these parameters in various regimes, namely by fixing some of these parameters and by considering the relationship among the remaining parameters.

To be concrete, consider the plots in Figure 1 which represent the performance of a family of codes $\mathcal{C}$ with rate $R = 0.5$. Different curves correspond to codes of different block length N. The codes are transmitted over a family of channels $\mathcal{W}$ parameterized by z, which is represented on the horizontal axis. On the vertical axis we represent the error probability P_{e}. The error probability is an increasing function of z, which means that the channel gets "better" as z decreases. The parameter z indicates the quality of the transmission channel W and, e.g., it could be set to $Z(W)$ or to $1 - I(W)$. Let us assume that there exists a threshold z^* such that, if $z < z^*$, then P_{e} tends to 0 as N grows large, while if $z > z^*$, then P_{e} tends to 1 as N grows large. For example, if the family of codes $\mathcal{C}$ is capacity

M. Mondelli and R. Urbanke are with the School of Computer and Communication Sciences, EPFL, CH-1015 Lausanne, Switzerland (e-mail: {marco.mondelli, ruediger.urbanke}@epfl.ch).

S. H. Hassani is with the Computer Science Department, ETH Zürich, Switzerland (e-mail: hamed@inf.ethz.ch).

achieving, then we can think to the threshold z^* as the channel parameter such that $I(W) = R$. In the example of Figure 1, we have that $z^* = 0.5$.

The oldest approach to analyze the performance of the family $\mathcal{C}$ is known under the name of *error exponent*. We pick any channel parameter $z < z^*$. Then, by definition of z^*, the error probability tends to 0 as N grows large. The error exponent regime quantifies this statement and computes how the error probability varies as a function of the block length. This corresponds to consider the blue vertical cut in Figure 1. The best possible scaling is obtained by considering random codes, for which $P_{\mathrm{e}} = e^{-NE(R,W)+o(N)}$, where $E(R, W)$ is the so-called error exponent [1].

Another approach is known under the name of *scaling exponent*. We pick a target error probability P_{e}. Then, by definition of z^*, the gap between the threshold and the channel parameter $z^* - z$ tends to 0 as N grows large. The scaling exponent regime quantifies this statement and computes how the gap to the threshold varies as a function of the block length. This corresponds to consider the red horizontal cut in Figure 1. From a practical viewpoint, we are interested in such a regime, since we typically have a certain requirement on the error probability and we look for the shortest code possible to transmit over the assigned channel. As a benchmark, a sequence of works starting from [2], then [3], and finally [4] shows that the smallest possible block length N required to achieve a gap to the threshold $z^* - z$ with a fixed error probability P_{e} is s.t.

$$N \approx \frac{V(Q^{-1}(P_{\mathrm{e}}))^2}{(z^* - z)^2}, \tag{1}$$

where $Q(\cdot)$ is the tail probability of the standard normal distribution and V is referred to as channel dispersion and measures the stochastic variability of the channel relative to a deterministic channel with the same capacity. A similar asymptotic expansion is put forward in [5] by using the information spectrum method. In general, if N is $\Theta\left(\frac{1}{(z^* - z)^\mu}\right)$, then we say that the family of codes $\mathcal{C}$ has scaling exponent μ. Hence, by (1), the most favorable scaling exponent is $\mu = 2$ and it is achieved by random codes. Further, for a large class of ensembles of LDPC codes and channel models the scaling exponent is also $\mu = 2$ [6].

To sum up, in the error exponent regime we compute how fast P_{e} goes to 0 as a function of N when $z^* - z$ is fixed, while in the scaling exponent regime we compute how fast $z^* - z$ goes to 0 as a function of N when P_{e} is fixed. Then, a natural question is to ask how fast *both* P_{e} *and* $z^* - z$ go to 0 as functions of N. In other words, one can describe a trade-off between the speed of decay of the error probability and the speed of decay of the gap to capacity as functions of the block length. This intermediate approach is named *moderate deviations* regime and it is studied for random codes in [7].

The last approach we are taking into account concerns the so-called *error floor*. We pick a specific code of an assigned block length N. Then, we compute how the error probability P_{e} behaves as a function of the channel parameter z. This corresponds to taking into account one of the four curves in Figure 1. This kind of analysis was introduced for iterative coding schemes, such as turbo and LDPC codes, where two distinct regions are typically recognizable: the so-called waterfall region in which the error probability falls off sharply as a function of the channel parameter and the *error floor* region in which the curves are much more shallow. In short, when there is no error floor, the error probability steadily decreases in the form of a waterfall as the channel condition improves (this is the case of Figure 1). On the other hand, when there is a point after which the error probability curve does not fall as quickly as before, then we enter the error floor region (this is the case of Figure 2). Error floors constitute a known issue for turbo codes, where they can be partially attributed to low-weight codewords [8], and for LDPC codes, where they are caused by small weaknesses in the graph and they are related to the weight distribution [9]. For transmission over the Binary Erasure Channel (BEC), the error floor region of LDPC codes is well understood [9, Section 3.24], [10] and computational techniques that accurately predict the performance for a given LDPC code in the error floor region have been developed [11].

Existing Results for Polar Codes. Polar codes have recently attracted the interest of the scientific community, since they provably achieve the capacity of a large class of channels, including any Binary Memoryless Symmetric Channel (BMSC), with low encoding and decoding complexity. Since their introduction in the seminal paper [12], the performance of polar codes has been extensively studied in different regimes.

As concerns the *error exponent* regime, in [13] it is proved that the block error probability under Successive Cancellation (SC) decoding behaves roughly as $2^{-\sqrt{N}}$. This result is further refined in [14], where it is shown that

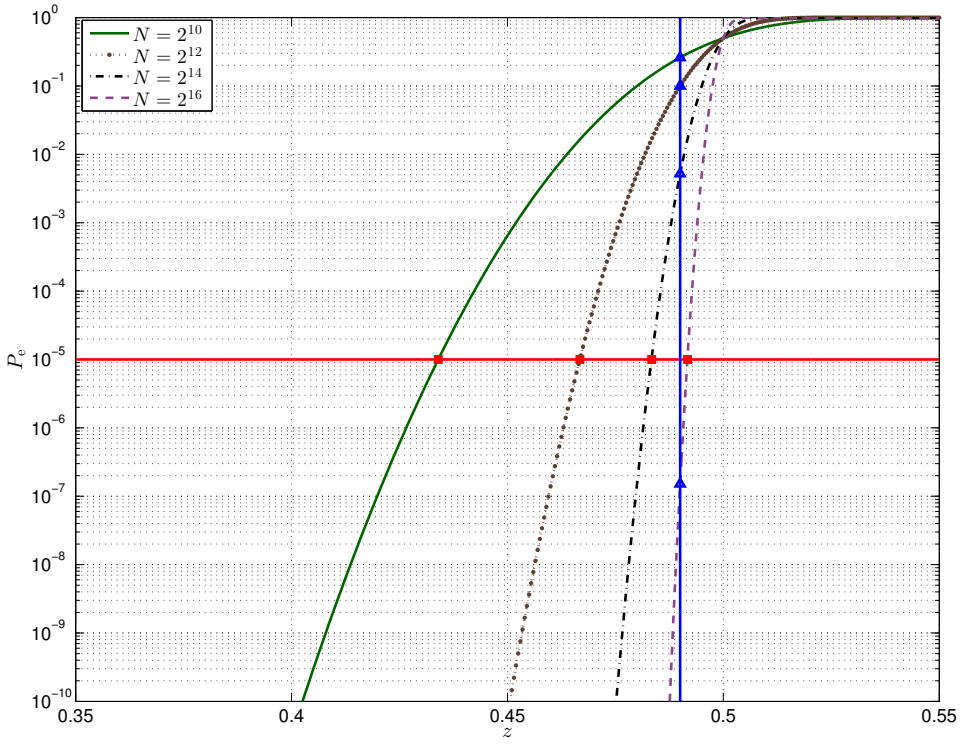

Figure 1. Performance of the family of codes $\mathcal{C}$ with rate $R = 0.5$ transmitted over the family of channels $\mathcal{W}$ with threshold $z^* = 0.5$. Each curve corresponds to a code of an assigned block length N, on the x-axis it is represented the channel parameter z, and on the y-axis the error probability P_{e}. The error exponent regime captures the behavior of the blue vertical cuts of fixed channel parameter z (or, equivalently, of fixed gap to threshold $z^* - z$). The scaling exponent regime captures the behavior of the red horizontal cuts of fixed error probability P_{e}. The error floor regime captures the behavior of a single curve of fixed block length N.

$\log_2(-\log_2 P_{\mathrm{e}})$ scales as $\frac{\log_2 N}{2} + \frac{\sqrt{\log_2 N}}{2} \cdot Q^{-1}\left(\frac{R}{C}\right) + o(\sqrt{\log_2 N})$. This last result holds both under SC decoding and under optimal MAP decoding.

As concerns the *scaling exponent* regime, the value of μ depends on the particular channel taken into account. A heuristic method to compute the scaling exponent for transmission over the BEC under SC decoding is provided in [15], and it yields $\mu \approx 3.627$. Universal bounds on μ valid for any BMSC under SC decoding are presented in [16]: the scaling exponent is lower bounded by 3.579 and it is upper bounded by 6. Furthermore, it is conjectured that the lower bound on μ can be increased up to 3.627, i.e., up to the value heuristically computed for the BEC. The upper bound on μ is further refined to 5.702 in [17]. Since a significant performance gain is obtained by using a Successive Cancellation List (SCL) decoder [18], the scaling exponent of list decoders has also been studied. However, in [19] it is proved that the value of μ does not change by adding any finite list size to the MAP decoder, and, in addition, the scaling exponent stays the same also under genie-aided SC decoding for any finite number of helps from the genie when transmission takes place over the BEC.

As concerns the *error floor* regime, in [20] it is proved that the stopping distance of polar codes scales as $\sqrt{N}$, which implies good error floor performance under Belief Propagation (BP) decoding, and simulation results have shown no sign of error floors for transmission over the BEC and over the Binary Additive White Gaussian Noise Channel (BAWGNC). However, even if we restrict to the simpler case of the transmission over the BEC, the existing results cannot rigorously exclude the existence of an error floor region for polar codes.

Contribution of the Present Work. This paper provides a unified view on the performance analysis of polar codes and presents several results about the scaling of the parameters of interest, namely, the rate R, the block length N, the error probability under SC decoding P_{e}, and the quality of the channel W. In particular, the contributions of this work concern the *scaling exponent*, the *moderate deviations*, and the *error floor* regimes, and they can be

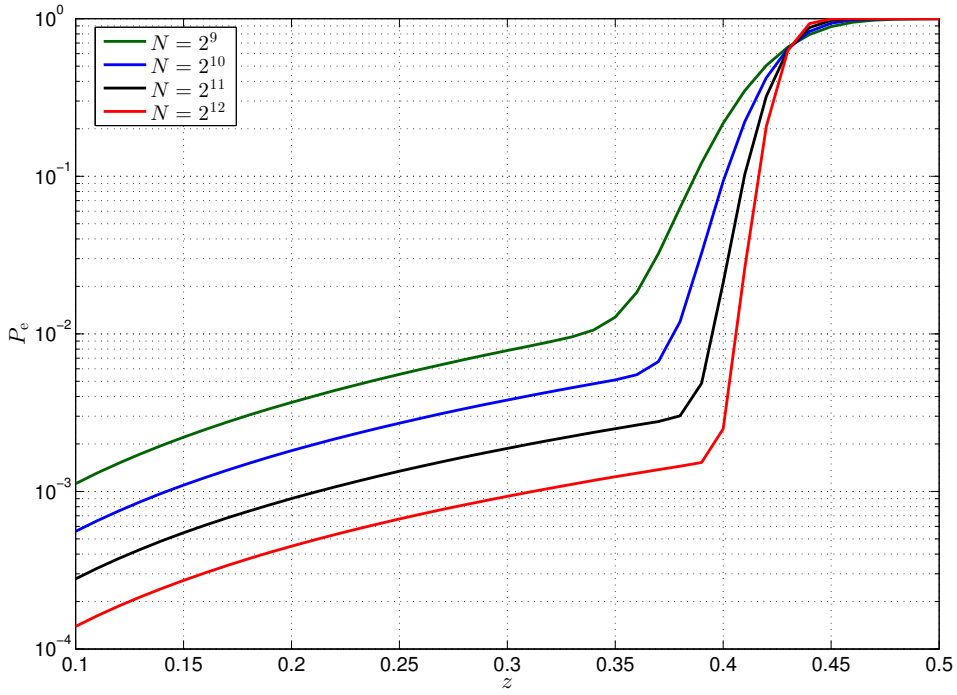

Figure 2. Performance of the family of $(3, 6)$-regular LDPC codes transmitted over the Binary Erasure Channel with erasure probability z. The waterfall region in which the error probability decreases sharply is clearly distinguishable from the error floor region in which the decay is much slower.

summarized as follows.

1) *New universal upper bound on the scaling exponent* μ. We show that $\mu \leq 4.714$ for any BMSC and that $\mu \leq 3.639$ for the BEC. Basically, this result improves by 1 the previous upper bound valid for any BMSC and approaches closely the value 3.627 which has been heuristically computed for the BEC. The proof technique consists in relating the scaling exponent to the sup of some function and, then, in describing an interpolation algorithm to obtain a provable upper bound on this sup. The values 4.714 for any BMSC and 3.639 for the BEC have been obtained for a particular number of samples used by the algorithm and they can be slightly improved simply by running the algorithm with a larger number of samples.
2) *Moderate deviations: joint scaling of error probability and gap to capacity*. We unify the two perspectives of the error exponent and the scaling exponent by allowing both the gap to capacity $I(W) - R$ and the error probability P_{e} to go to 0 as functions of the block length N. In particular, we describe a trade-off between the speed of decay of P_{e} and the speed of decay of $I(W) - R$. In the limit in which the gap to capacity is arbitrary small but independent of N, this trade-off recovers the result of [13] where it is shown that P_{e} scales roughly as $2^{-\sqrt{N}}$.
3) *Absence of error floors*. We study the dependency of the error probability on the quality of the channel over which the transmission takes place. To do so, we assign a polar code of block length N and rate R designed for transmission over a channel W'. Then, we look at the performance of this fixed code over other channels W which are "better" that W'. In particular, we study the error probability P_{e} as a function of the Bhattacharyya parameter $Z(W)$ of the transmission channel W. Note that the code is fixed and the channel varies, which means that we do not choose the optimal polar indices for W. In particular, we prove that P_{e} scales as $Z(W)$ raised to some power which depends on N and, in addition, the exponent behaves roughly as $\sqrt{N}$ in accordance to the error exponent regime. As a result, we conclude that polar codes are not affected by error floors.

The rest of the paper is organized as follows. Section II reviews some preliminary notions about polar coding. The successive three sections contain the original contributions of the paper: Section III presents the new upper bound on the scaling exponent, Section IV concerns the moderate deviations regime, and Section V proves that polar codes

are not affected by error floors. Section VI concludes the paper with some final remarks.

II. Preliminaries

Let W be a BMSC, and let $\mathcal{X} = \{0,1\}$ denote its input alphabet, $\mathcal{Y}$ the output alphabet, and $\{W(y \mid x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ the transition probabilities. Denote by $I(W) \in [0,1]$ the mutual information between the input and output of W with uniform distribution on the input. Then, $I(W)$ is also equal to the capacity of W. Denote by $Z(W) \in [0,1]$ the Bhattacharyya parameter of W, which is defined as

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

and it is related to the capacity $I(W)$ via

$$Z(W) + I(W) \geq 1, \tag{2}$$

$$Z(W)^2 + I(W)^2 \leq 1, \tag{3}$$

both proved in [12].

The basis of channel polarization consists in mapping two identical copies of the channel $W : \mathcal{X} \to \mathcal{Y}$ into the pair of channels $W^0 : \mathcal{X} \to \mathcal{Y}^2$ and $W^1 : \mathcal{X} \to \mathcal{X} \times \mathcal{Y}^2$, defined as [12, Section I-B], [16, Section I-B],

$$\begin{aligned} W^0(y_1, y_2 \mid x_1) &= \sum_{x_2 \in \mathcal{X}} \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2), \\ W^1(y_1, y_2, x_1 \mid x_2) &= \frac{1}{2} W(y_1 \mid x_1 \oplus x_2) W(y_2 \mid x_2). \end{aligned} \tag{4}$$

Then, the idea is that W^0 is a "worse" channel and W^1 is a "better" channel than W. This statement can be quantified by computing the relations among the Bhattacharyya parameters of W, W^0 and W^1:

$$Z(W)\sqrt{2 - Z(W)^2} \leq Z(W^0) \leq 2Z(W) - Z(W)^2, \tag{5}$$

$$Z(W^1) = Z(W)^2, \tag{6}$$

which follow from Proposition 5 of [12] and from Exercise 4.62 of [9]. In addition, when W is a BEC, we have that W^0 and W^1 are also BECs and, by Proposition 5 of [12],

$$Z(W^0) = 2Z(W) - Z(W)^2. \tag{7}$$

By repeating n times this operation, we map 2^n identical copies of W into the synthetic channels $W_n^{(i)}$ ($i \in \{1, \cdots, 2^n\}$), defined as

$$W_n^{(i)} = (((W^{b_1^{(i)}})^{b_2^{(i)}})^{\cdots})^{b_n^{(i)}}, \tag{8}$$

where $(b_1^{(i)}, \cdots, b_n^{(i)})$ is the binary representation of the integer $i-1$ over n bits.

Given a BMSC W, for $n \in \mathbb{N}$, define a random sequence of channels W_n, as $W_0 = W$, and

$$W_n = \begin{cases} W_{n-1}^0, & \text{w.p. } 1/2, \\ W_{n-1}^1, & \text{w.p. } 1/2. \end{cases} \tag{9}$$

Let $Z_n(W) = Z(W_n)$ be the random process that tracks the Bhattacharyya parameter of W_n. Then, from (5) and (6) we deduce that, for $n \geq 1$,

$$Z_n \begin{cases} \in \left[Z_{n-1}\sqrt{2 - Z_{n-1}^2},\, 2Z_{n-1} - Z_{n-1}^2 \right], & \text{w.p. } 1/2, \\ = Z_{n-1}^2, & \text{w.p. } 1/2. \end{cases} \tag{10}$$

When W is a BEC with erasure probability z, then the process Z_n has a simple closed form. It starts with $Z_0 = z$, and, by using (5) and (6), we deduce that, for $n \geq 1$,

$$Z_n = \begin{cases} 2Z_{n-1} - Z_{n-1}^2, & \text{w.p. } 1/2, \\ Z_{n-1}^2, & \text{w.p. } 1/2. \end{cases} \tag{11}$$

Consider transmission over W of a polar code of block length $N = 2^n$ and rate R and let P_{e} denote the block error probability under SC decoding. Then, by Proposition 2 of [12],

$$P_{\mathrm{e}} \leq \sum_{i \in \mathcal{I}} Z_n^{(i)}, \tag{12}$$

where $Z_n^{(i)}$ denotes the Bhattacharyya parameter of $W_n^{(i)}$ and $\mathcal{I}$ denotes the information set, i.e., the set containing the positions of the information bits.

III. New Universal Upper Bound on the Scaling Exponent

In this section, we present an improved upper bound on the scaling exponent which is valid for transmission over any BMSC W. First of all, we relate the value of the scaling exponent μ to the $\sup$ of some function. Secondly, we provide a provable bound on this $\sup$, which gives us a provably valid choice for μ, i.e., $\mu = 4.714$ for any BMSC and $\mu = 3.639$ for the BEC. More specifically, Section III-A contains the statement and the discussion of these two main theorems. Sections III-B and III-C contain the proof of the first and of the second result, respectively.

A. Main Result: Statement and Discussion

Theorem 1 (From fixed function to scaling exponent): Assume that there exists a *fixed function* $h(x) : [0,1] \to [0,1]$ s.t. $h(0) = h(1) = 0$, $h(x) > 0$ for any $x \in (0,1)$, and, for some $\mu > 2$,

$$\sup_{x \in (0,1), y \in [x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2) + h(y)}{2h(x)} < 2^{-1/\mu}. \tag{13}$$

Consider transmission over a BMSC W with capacity $I(W)$ using a polar code of rate $R < I(W)$. Fix $p_{\mathrm{e}} \in (0,1)$ and assume that the block error probability under successive cancellation decoding is *at most* p_{e}. Then, it suffices to have a block length N s.t.

$$N \leq \frac{\beta_1}{(I(W) - R)^{\mu}}, \tag{14}$$

where β_1 is a constant which depends only on p_{e}. If W is a BEC, a less stringent hypothesis on μ is required for (14) to hold: the condition (13) is replaced by

$$\sup_{x \in (0,1)} \frac{h(x^2) + h(2x - x^2)}{2h(x)} < 2^{-1/\mu}. \tag{15}$$

Theorem 2 (Valid choice for scaling exponent): Consider transmission over a BMSC W with capacity $I(W)$ using a polar code of rate $R < I(W)$. Fix $p_{\mathrm{e}} \in (0,1)$ and assume that the block error probability under successive cancellation decoding is *at most* p_{e}. Then, it suffices to have a block length N upper bounded by (14) with $\mu = 4.714$ Furthermore, if W is a BEC, then (14) holds with $\mu = 3.639$.

Before proceeding with the proofs, it is useful to discuss two points. The first remark focuses on the role of the fixed function $h(x)$ and shows a heuristic reason why the value of the scaling exponent is linked to the existence of a function that fulfills condition (13) (condition (15) for the BEC). The second remark points out that we can allow the error probability to tend to 0 polynomially fast in N while maintaining the same scaling between gap to capacity and block length.

Remark 3 (Heuristic interpretation of fixed function): First, let W be a BEC and consider the operator T_{BEC} defined as

$$T_{\mathrm{BEC}}(g) = \frac{g(z^2) + g(2z - z^2)}{2}, \tag{16}$$

where $g(z)$ is a bounded and real valued function over $[0,1]$. The relation between the Bhattacharyya process Z_n and the operator T_{BEC} is given by

$$\mathbb{E}\left[g(Z_n) \mid Z_0 = z\right] = \overbrace{T_{\mathrm{BEC}} \circ T_{\mathrm{BEC}} \circ \cdots \circ T_{\mathrm{BEC}}(g)}^{n \text{ times}} = T_{\mathrm{BEC}}^n(g), \tag{17}$$

where the formula comes from a straightforward application of (11). A detailed explanation of the dynamics of the functions $T_{\mathrm{BEC}}^n(g)$ is provided in Section III of [16]. In short, a simple check shows that $\lambda = 1$ is an eigenvalue of the operator T_{BEC} with eigenfunctions $v_0(z) = 1$ and $v_1(z) = z$. Let λ^* be the largest eigenvalue of T_{BEC} other than $\lambda = 1$ and define μ^* as $\mu^* = -\dfrac{1}{\log_2 \lambda^*}$. Then, the heuristic discussion of [16] leads to the fact that μ^* is the largest candidate that we could plug in (15). For this choice, the *fixed function* $h(x)$ represents the eigenfunction associated to the eigenvalue λ^*, namely,

$$\frac{h(x^2) + h(2x - x^2)}{2} = 2^{-1/\mu^*} h(x). \tag{18}$$

A numerical method for the calculation of this second eigenvalue was originally proposed in [15] and it yields $\mu^* = 3.627$. Furthermore, in Section III of [16] it is also heuristically explained how $\mu^* = 3.627$ gives a lower bound to the scaling exponent of the BEC.

Now, let W be a BMSC and consider the operator T_{BMSC} defined as

$$T_{\mathrm{BMSC}}(g) = \sup_{y \in [x\sqrt{2-x^2}, 2x-x^2]} \frac{g(z^2) + g(y)}{2}. \tag{19}$$

The relation between the Bhattacharyya process Z_n and the operator T_{BMSC} is given by

$$\mathbb{E}\left[g(Z_n) \mid Z_0 = z\right] \leq T_{\mathrm{BMSC}}^n(g), \tag{20}$$

where the formula comes from a straightforward application of (10). Similarly to the case of the BEC, $\lambda = 1$ is an eigenvalue of T_{BMSC} and we write the largest eigenvalue other than $\lambda = 1$ as $2^{-1/\mu^*}$. Then, the idea is that μ^* is the largest candidate that we could plug in (13) and, for this choice, the *fixed function* $h(x)$ represents the eigenfunction associated to the eigenvalue $2^{-1/\mu^*}$, namely,

$$\sup_{y \in [x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2) + h(y)}{2} = 2^{-1/\mu^*} h(x). \tag{21}$$

In Section IV of [16] it is proved that the scaling exponent μ is upper bounded by 6: this result is obtained by showing that the eigenvalue is at least $2^{-1/5}$, i.e. $\mu^* \leq 5$ and, then, that $\mu^* + 1$ is an upper bound on the scaling exponent μ. Furthermore, it is conjectured that μ^* is a tighter upper bound on the scaling exponent μ. In [17], a more refined computation of μ^* is presented, which yields $\mu^* \leq 4.702$, and, therefore, $\mu \leq 5.702$. In this paper, we solve the conjecture of [16] by proving that, indeed, μ^* is an upper bound on the scaling exponent μ. In addition, we show an algorithm which guarantees a *provable* bound on the eigenvalue, thus obtaining $\mu \leq 4.714$ for any BMSC and $\mu \leq 3.639$ for the BEC. We finally note from (20) that T_{BMSC} provides only an upper bound on the (expected) evolution of Z_n. As a result, although $\mu \leq 4.714$ holds universally for any channel, this bound is certainly not tight if we consider a specific BMSC.

Remark 4 (Polynomial decay of P_{e}): With some more work, it is possible to prove the following generalization of Theorem 1. Assume that there exists $h(x)$ as in Theorem 1 and consider transmission over a BMSC W with capacity $I(W)$ using a polar code of rate $R < I(W)$. Then, for any $\nu > 0$, the block length N and the block error probability under successive cancellation decoding P_{e} are s.t.

$$\begin{aligned} P_{\mathrm{e}} &\leq \frac{1}{N^{\nu}}, \\ N &\leq \frac{\beta_2}{(I(W) - R)^{\mu}}, \end{aligned} \tag{22}$$

where β_2 is a constant. A sketch of the proof of this statement is given at the end of Section III-B. The result (22) is a generalization of Theorem 1 in the sense that, instead of being an assigned constant, the error probability goes to 0 polynomially fast in $1/N$, while the scaling between block length and gap to capacity, i.e., the value of μ, stays the same. On the other hand, as described in Section IV, if the error probability is $O(2^{-N^{\beta}})$ for some $\beta \in (0, 1/2)$, then the scaling between block length and gap to capacity changes and depends on the exponent β.

B. From Fixed Function to Scaling Exponent: Proof of Theorem 1

The proof of Theorem 1 relies on the following two auxiliary results: Lemma 5, which is proved in Appendix A, relates the number of synthetic channels with Bhattacharyya parameter small enough to an expected value over the Bhattacharyya process; Lemma 6, which is proved in Appendix B, relates the expected value over the Bhattacharyya process to the *fixed function* $h(x)$.

Lemma 5 (From expectation to scaling exponent): Let $Z_n(W)$ be the Bhattacharyya process associated to the channel W. Pick any $\alpha \in (0,1)$ and assume that, for $n \geq 1$ and for some $\rho \leq 1/2$,

$$\mathbb{E}\left[(Z_n(1-Z_n))^{\alpha}\right] \leq c_1 2^{-n\rho}, \tag{23}$$

where c_1 is a constant which does not depend on n. Then,

$$\mathbb{P}\left(Z_n \leq p_{\mathrm{e}} 2^{-n}\right) \geq I(W) - c_2 2^{-n(\rho-\alpha)}, \tag{24}$$

where $c_2 = \sqrt{2p_{\mathrm{e}}} + 2c_1 p_{\mathrm{e}}^{-\alpha}$.

Lemma 6 (From fixed function to expectation): Let $h(x) : [0,1] \to [0,1]$ s.t. $h(0) = h(1) = 0$, $h(x) > 0$ for any $x \in (0,1)$, and

$$\sup_{x\in(0,1), y\in[x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2)+h(y)}{2h(x)} \leq 2^{-\rho_1}. \tag{25}$$

for some $\rho_1 \leq 1/2$. Let $Z_n(W)$ be the Bhattacharyya process associated to the channel W. Pick any $\alpha \in (0,1)$. Then, for any $\delta \in (0,1)$, and for $n \in \mathbb{N}$,

$$\mathbb{E}\left[(Z_n(1-Z_n))^{\alpha}\right] \leq \frac{1}{\delta}\left(2^{-\rho_1} + \sqrt{2}\frac{\delta}{1-\delta}c_3\right)^n, \tag{26}$$

where c_3 is defined as

$$c_3 = \sup_{x\in(\epsilon_1(\alpha), 1-\epsilon_2(\alpha))} \frac{(x(1-x))^{\alpha}}{h(x)}, \tag{27}$$

and $\epsilon_1(\alpha)$, $\epsilon_2(\alpha)$ denote the only two solutions in $[0,1]$ of the equation

$$\frac{1}{2}\left(\big(x(1+x)\big)^{\alpha} + \big((2-x)(1-x)^{1/3}\big)^{\alpha}\right) = 2^{-\rho_1}. \tag{28}$$

If W is a BEC, a less stringent hypothesis on ρ_1 is required for (26) to hold: the condition (25) is replaced by

$$\sup_{x\in(0,1)} \frac{h(x^2)+h(2x-x^2)}{2h(x)} \leq 2^{-\rho_1}. \tag{29}$$

At this point, we are ready to put everything together and prove Theorem 1.

Proof of Theorem 1: Let us define

$$\rho_1 = \min\left(\frac{1}{2}, -\log_2 \sup_{x\in(0,1), y\in[x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2)+h(y)}{2h(x)}\right), \tag{30}$$

where $h(x)$ is the fixed function of the hypothesis.

Set

$$\alpha = \log_2\left(1 + \frac{2^{-1/\mu} - 2^{-\rho_1}}{2^{-1/\mu} + 2^{-\rho_1}}\right). \tag{31}$$

By using (13) and the fact that $\mu > 2$, we immediately realize that $2^{-1/\mu} - 2^{-\rho_1} > 0$, and, therefore, that $\alpha > 0$. In addition, one easily checks that $\alpha < 1$.

Set

$$\delta = \frac{2^{-1/\mu} - 2^{-\rho_1}}{2\sqrt{2}c_3 + 2^{-1/\mu} - 2^{-\rho_1}}, \tag{32}$$

where c_3 is defined as in (27). Since $2^{-1/\mu} - 2^{-\rho_1} > 0$, we have that $\delta \in (0,1)$.

In addition, $\rho_1 \leq 1/2$ and the condition (25) clearly follows from the definition (30). Consequently, we can apply Lemma 6 which yields formula (26).

Set

$$\rho = -\log_2\left(2^{-\rho_1} + \sqrt{2}\frac{\delta}{1-\delta}c_3\right). \tag{33}$$

Then, $\rho \leq \rho_1 \leq 1/2$, and we can apply Lemma 5 with $c_1 = 1/\delta$ which yields

$$\mathbb{P}\left(Z_n \leq p_{\mathrm{e}} 2^{-n}\right) \geq I(W) - c_2 2^{-n(\rho-\alpha)} = I(W) - c_2 2^{-n/\mu}, \tag{34}$$

where $c_2 = \sqrt{2p_{\mathrm{e}}} + 2p_{\mathrm{e}}^{-\alpha}/\delta$ and the last equality uses the definitions (33), (31) and (32).

Consider transmission of a polar code of block length $N = 2^n$ and rate $R = I(W) - c_2 2^{-n/\mu}$ over W. Then, by combining (12) and (34), we have that the error probability under successive cancellation decoding is upper bounded by p_{e}. Therefore, the result (14) follows with $\beta_1 = c_2^{\mu}$.

A similar proof holds for the specific case in which W is a BEC.

■

Eventually, let us briefly sketch how to prove the result stated in Remark 4. First, we need to generalize Lemma 5 by showing that, under the same hypothesis (23), we have that, for any $\nu > 0$,

$$\mathbb{P}\left(Z_n \leq 2^{-n(\nu+1)}\right) \geq I(W) - c_4 2^{-n(\rho-(\nu+1)\alpha)}, \tag{35}$$

where $c_4 = \sqrt{2} + 2c_1$. Then, we simply follow the procedure described in the proof of Theorem 1 with the difference that α is a factor $1+\nu$ smaller than in (31).

C. Valid Choice for Scaling Exponent: Proof of Theorem 2

Let W be a BMSC. The proof of Theorem 2 consists in providing a good candidate for the *fixed function*, i.e., in finding a function $h(x) : [0,1] \to [0,1]$ s.t. $h(0) = h(1) = 0$, $h(x) > 0$ for any $x \in (0,1)$ and (13) is satisfied with a value of μ as small as possible. In particular, we will prove that $\mu = 4.714$ is a valid choice.

The idea is to apply repeatedly the operator T_{BMSC} defined in (19) until we converge to the fixed function $h(x)$. Hence, let us define $h_k(x)$ recursively for any $k \geq 1$ as

$$h_k(x) = \frac{f_k(x)}{\sup_{y\in(0,1)} f_k(y)}, \tag{36}$$

$$f_k(x) = \sup_{y\in[x\sqrt{2-x^2}, 2x-x^2]} \frac{h_{k-1}(x^2) + h_{k-1}(y)}{2}, \tag{37}$$

with some initial condition $h_0(x)$ s.t. $h_0(0) = h_0(1) = 0$ and $h_0(x) > 0$ for any $x \in (0,1)$. Note that the normalization step (36) ensures that the function $h_k(x)$ does not tend to the constant function 0 in the interval $[0,1]$.

However, even if we choose some simple initial condition $h_0(x)$, the sequence of functions $\{h_k(x)\}_{k\in\mathbb{N}}$ is analytically intractable. Hence, we need to resort to numerical methods, keeping in mind that we require a *provable* upper bound for any $x \in (0,1)$ on the function

$$r(x) = \sup_{y\in[x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2) + h(y)}{2h(x)}. \tag{38}$$

To do so, first we construct an adequate candidate for the fixed function $h(x)$. This function is going to depend on some auxiliary parameters. Then, we describe an algorithm to analyze this candidate and we present a choice of the parameters that gives $\mu = 4.714$. Let us underline that, despite the procedure is numerical, the resulting upper bound and, therefore, the value of μ are rigorously *provable*.

As concerns the construction part, we observe numerically that, when k is large enough, the function $h_k(x)$ depends weakly on the initial condition $h_0(x)$ and it does not change much after one more iteration, i.e., $h_{k+1}(x) \approx h_k(x)$. In addition, let us point out that the goal is *not* to obtain an exact approximation of the sequence of functions $\{h_k(x)\}_{k\in\mathbb{N}}$ defined in (36)-(37). Indeed, the actual goal is to obtain a candidate $h(x)$ which satisfies (13) with a value of μ as low as possible.

Pick a large integer $N_{\rm s}$ and let us define the sequence of functions $\{\hat{h}_k(x)\}_{k\in\mathbb{N}}$ as follows. For any $k\in\mathbb{N}$, $\hat{h}_k(x)$ is the piece-wise linear function obtained by linear interpolation from the samples $\hat{h}_k(x_i)$, where $x_i = i/N_{\rm s}$ for $i\in\{0,1,\cdots,N_{\rm s}\}$. The samples $\hat{h}_k(x_i)$ are given by

$$\begin{aligned}\hat{h}_k(x_i) &= \frac{\hat{f}_k(x_i)}{\max_{j\in\{0,1,\cdots,N_{\rm s}\}}\hat{f}_k(x_j)},\\ \hat{f}_k(x_i) &= \frac{\hat{h}_{k-1}((x_i)^2)+\max_{j\in\{0,1,\cdots,M_{\rm s}\}}\hat{h}_{k-1}(y_{i,j})}{2},\end{aligned} \tag{39}$$

where $M_{\rm s}$ is a large integer, and, for $j\in\{0,1,\cdots,M_{\rm s}\}$, $y_{i,j}$ is defined as

$$y_{i,j} = x_i\sqrt{2-x_i^2}+\frac{j}{M_{\rm s}}x_i\left(2-x_i-\sqrt{2-x_i^2}\right). \tag{40}$$

The initial samples $\hat{h}_0(x_i)$ are obtained by evaluating at the points $\{x_i\}_{i=0}^{N_{\rm s}}$ some function $h_0(x)$ s.t. $h_0(0)=h_0(1)=0$ and $h_0(x)>0$ for any $x\in(0,1)$.

It is clear that, by increasing $N_{\rm s}$ and $M_{\rm s}$, we obtain a better approximation of the sequence of functions (36)-(37). In addition, by increasing k we get closer to the limiting function $\lim_{k\to\infty}\hat{h}_k(x)$. Set

$$\hat{r}_k = \max_{i\in\{1,\cdots,N_{\rm s}-1\}}\frac{\hat{h}_k((x_i)^2)+\max_{j\in\{0,1,\cdots,M_{\rm s}\}}\hat{h}_k(y_{i,j})}{2\hat{h}_k(x_i)}. \tag{41}$$

We observe from numerical simulations that, when k increases, the sequence $\hat{r}_k$ tends to the limiting value 0.86275. Furthermore, this limit depends very weakly on the particular choice of the initial conditions $\{\hat{h}_0(x_i)\}_{i=0}^{N_{\rm s}}$.

Note that $\hat{r}_k$ gives an indication of what is the smallest value of μ that we could hope for by using the samples $\{\hat{h}_k(x_i)\}_{i=0}^{N_{\rm s}}$, i.e., $\mu = -\dfrac{1}{\log_2 0.86275} = 4.695$. Indeed, if we obtain $h(x)$ by interpolating the samples $\{\hat{h}_k(x_i)\}_{i=0}^{N_{\rm s}}$, then $\hat{r}_k = \max_{i\in\{1,\cdots,N_{\rm s}-1\}} r(i/N_{\rm s})$, where $r(x)$ is defined in (38). Therefore, $\hat{r}_k \le \sup_{x\in(0,1)} r(x)$, i.e., $\hat{r}_k$ is a lower bound on the desired sup, while we are looking for an upper bound to that quantity.

Fix a large integer $\bar{k}$ and, before computing a provable upper bound on $\sup_{x\in(0,1)} r(x)$, let us describe the interpolation method to obtain the candidate $h(x)$ from the samples $\{\hat{h}_{\bar{k}}(x_i)\}_{i=0}^{N_{\rm s}}$.

For x close to 0 and for x close to 1, linear interpolation does not yield a good candidate $h(x)$. Indeed, assume that $h(x)=\hat{h}_{\bar{k}}(x)$ for $x\in\left[0,\dfrac{1}{N_{\rm s}}\right]$. Then, $\lim_{x\to 0^+} r(x) = 1$, and, therefore, $\sup_{x\in(0,1)} r(x)\ge 1$. Similarly, if $h(x)=\hat{h}_{\bar{k}}(x)$ for $x\in\left[1-\dfrac{1}{N_{\rm s}},1\right]$, then $\lim_{x\to 1^-} r(x) = 1$. On the other hand, if $h(x)$ grows as x^η in a neighborhood of 0 for $\eta\in(0,1)$, then, it is easy to see that $\lim_{x\to 0^+} r(x) = 2^{\eta-1}$. Similarly, if $h(x)$ grows as $(1-x)^\eta$ in a neighborhood of 1 for $\eta\in(0,1)$, then $\lim_{x\to 1^-} r(x) = 2^{\eta-1}$. Consequently, the idea is to choose η slightly smaller than $1-\dfrac{1}{4.695}$, where 4.695 constitutes a good approximation to the target value of μ that we are going to achieve. Motivated by this observation, set

$$b_0(x) = \hat{h}_{\bar{k}}\left(\frac{\bar{m}}{N_{\rm s}}\right)\left(\frac{\bar{m}}{N_{\rm s}}\right)^{-\eta}x^\eta, \tag{42}$$

$$b_1(x) = \hat{h}_{\bar{k}}\left(1-\frac{\bar{m}}{N_{\rm s}}\right)\left(\frac{\bar{m}}{N_{\rm s}}\right)^{-\eta}(1-x)^\eta, \tag{43}$$

for some integer $\bar{m}\ge 2$. Then, sample $b_0(x)$ for $x\in\left[\dfrac{1}{N_{\rm s}},\dfrac{\bar{m}}{N_{\rm s}}\right]$, sample $\hat{h}_{\bar{k}}(x)$ for $x\in\left[\dfrac{\bar{m}}{N_{\rm s}},1-\dfrac{\bar{m}}{N_{\rm s}}\right]$, and sample $b_1(x)$ for $x\in\left[1-\dfrac{\bar{m}}{N_{\rm s}},1-\dfrac{1}{N_{\rm s}}\right]$. Note that it is better not to have a uniform sampling, but to choose the number of samples according to the following rule: pick some $\delta_{\rm s}$ small enough; then, for each couple of consecutive samples, the bigger one has to be at most a factor $1+\delta_{\rm s}$ larger than the smaller one. Let $\{x_i'\}_{i=1}^{N_{\rm s}'}$ denote the set of sampling positions and $\{\hat{h}_i\}_{i=1}^{N_{\rm s}'}$ denote the set of samples obtained with this procedure, where $N_{\rm s}'$ is the number of such

samples. Eventually, we define the candidate $h(x)$ as

$$\begin{aligned} h(x) &= b_0(x), && \text{for } x \in \left[0, \frac{1}{N_\mathrm{s}}\right], \\ h(x) &= b_1(x), && \text{for } x \in \left[1 - \frac{1}{N_\mathrm{s}}, 1\right], \end{aligned} \tag{44}$$

and, for $x \in \left[\frac{1}{N_\mathrm{s}}, 1 - \frac{1}{N_\mathrm{s}}\right]$, $h(x)$ is obtained by linear interpolation from the samples $\{\hat{h}_i\}$.

As concerns the analysis of $h(x)$, let us remind that the goal is to find a provable upper bound on $\sup_{x\in(0,1)} r(x)$. First, consider the values of x in a neighborhood of 0. The following chain of inequalities holds for any $x \in \left[0, \frac{1}{N_\mathrm{s}}\right]$,

$$\begin{aligned} r(x) &\overset{\text{(a)}}{\leq} \frac{h(x^2) + h(2x)}{2h(x)} \\ &\overset{\text{(b)}}{\leq} \frac{b_0(x^2) + b_0(2x)}{2b_0(x)} \\ &\overset{\text{(c)}}{=} \frac{x^\eta}{2} + 2^{\eta-1} \\ &\leq H_0 \triangleq \frac{(N_\mathrm{s})^{-\eta}}{2} + 2^{\eta-1}, \end{aligned} \tag{45}$$

where (a) uses that $h(y) \leq h(2x)$ for any $y \in [x\sqrt{2-x^2}, 2x - x^2]$ since $h(x)$ is increasing for $x \in \left[0, \frac{2}{N_\mathrm{s}}\right]$, (b) uses that $h(x) = b_0(x)$ for $x \in \left[0, \frac{1}{N_\mathrm{s}}\right]$ and $h(x) \leq b_0(x)$ for $x \in \left[\frac{1}{N_\mathrm{s}}, \frac{2}{N_\mathrm{s}}\right]$ since in that interval $h(x)$ is the linear interpolation of samples taken from $b_0(x)$ and $b_0(x)$ is concave for any $\eta \in (0,1)$, and (c) uses the definition (42) of $b_0(x)$.

Secondly, consider the values of x is a neighborhood of 1. The following chain of inequalities holds for any $x \in \left[1 - \frac{1}{N_\mathrm{s}}, 1\right]$,

$$\begin{aligned} r(x) &\overset{\text{(a)}}{\leq} \frac{h(x^2) + h(x\sqrt{2-x^2})}{2h(x)} \\ &\overset{\text{(b)}}{\leq} \frac{b_1(x^2) + b_1(x\sqrt{2-x^2})}{2b_1(x)} \\ &\overset{\text{(c)}}{=} \frac{(1+x)^\eta}{2} + \frac{1}{2}\left(\frac{1 - x\sqrt{2-x^2}}{1-x}\right)^\eta \\ &\overset{\text{(d)}}{\leq} H_1 \triangleq 2^{\eta-1} + \frac{1}{2}\left(N_\mathrm{s} - (N_\mathrm{s}-1)\sqrt{1 + \frac{2}{N_\mathrm{s}} - \frac{1}{(N_\mathrm{s})^2}}\right)^\eta, \end{aligned} \tag{46}$$

where (a) uses that $h(y) \leq h(x\sqrt{2-x^2})$ for any $y \in [x\sqrt{2-x^2}, 2x - x^2]$ since $h(x)$ is decreasing for $x \in \left[1 - \frac{1}{N_\mathrm{s}}, 1\right]$, (b) uses that $h(x) = b_1(x)$ for $x \in \left[1 - \frac{1}{N_\mathrm{s}}, 1\right]$ and $h(x) \leq b_1(x)$ for $x \in \left[\frac{1}{N_\mathrm{s}}, \frac{2}{N_\mathrm{s}}\right]$ since in that interval $h(x)$ is the linear interpolation of samples taken from $b_1(x)$ and $b_1(x)$ is concave for any $\eta \in (0,1)$, (c) uses the definition (43) of $b_1(x)$, and (d) uses that $\frac{1 - x\sqrt{2-x^2}}{1-x}$ is decreasing for any $x \in (0,1)$.

Finally, consider the values of x in the interval $\left[\frac{1}{N_\mathrm{s}}, 1 - \frac{1}{N_\mathrm{s}}\right]$. For any $i \in \{1, \cdots, N'_\mathrm{s} - 1\}$, define

$$\begin{aligned} J_i^+ &= \{j : x'_j \in [(x'_i)^2, (x'_{i+1})^2]\}, \\ J_i^- &= \{j : x'_j \in [x'_i\sqrt{2 - (x'_i)^2}, 2x'_{i+1} - (x'_{i+1})^2]\}. \end{aligned}$$

Then, since $h(x)$ is piece-wise linear in the interval $\left[\frac{1}{N_{\mathrm{s}}}, 1-\frac{1}{N_{\mathrm{s}}}\right]$, we have that, for any $x \in [x'_i, x'_{i+1}]$,

$$\begin{aligned} h(x) &\geq \min\left(h(x'_i), h(x'_{i+1})\right), \\ h(x^2) &\leq h_i^+ \triangleq \max\left(h\left((x'_i)^2\right), h\left((x'_{i+1})^2\right), \max_{j\in J_i^+}\left(h(x'_j)\right)\right), \\ \sup_{y\in[x\sqrt{2-x^2}, 2x-x^2]} h(y) &\leq h_i^- \triangleq \max\left(h\left(x'_i\sqrt{2-(x'_i)^2}\right), h\left(2x'_{i+1}-(x'_{i+1})^2\right), \max_{j\in J_i^-}\left(h(x'_j)\right)\right), \end{aligned}$$

which implies that, for any $x \in [x'_i, x'_{i+1}]$,

$$r(x) \leq \frac{h_i^+ + h_i^-}{2\min\left(h(x'_i), h(x'_{i+1})\right)}. \tag{47}$$

As a result, by combining (45), (46), and (47), we conclude that

$$\sup_{x\in(0,1)} r(x) \leq \max\left(H_0, H_1, \max_{i\in\{1,\cdots,N'_{\mathrm{s}}-1\}} \frac{h_i^+ + h_i^-}{2\min\left(h(x'_i), h(x'_{i+1})\right)}\right), \tag{48}$$

which implies that (13) holds for any μ s.t. $2^{-1/\mu}$ is an upper bound on the RHS of (48).

Let us choose δ_{s}, η, the sampling positions $\{x'_i\}_{i=1}^{N'_{\mathrm{s}}}$, and the samples $\{\hat{h}_i\}_{i=1}^{N'_{\mathrm{s}}}$ to be rational numbers. Then, the RHS of (48) is the maximum of either rational numbers or sums of rational powers of rational numbers. Consequently, we can provide a provable upper bound on the RHS of (48), and, therefore, on μ. In particular, by setting $N_{\mathrm{s}} = 10^6$, $M_{\mathrm{s}} = 10^4$, $h_0(x) = (x(1-x))^{3/4}$, $k = 100$, $\delta_{\mathrm{s}} = 10^{-4}$, $\eta = 78/100$, and $\bar{m} = 13$, we obtain $\mu = 4.714$.

For the BEC the idea is to apply repeatedly the operator T_{BEC} defined in (16). Hence, by adapting the procedure described above and by setting $N_{\mathrm{s}} = 10^6$, $M_{\mathrm{s}} = 10^4$, $h_0(x) = (x(1-x))^{2/3}$, $k = 100$, $\delta_{\mathrm{s}} = 10^{-4}$, $\eta = 72/100$, and $\bar{m} = 5$, we obtain $\mu = 3.639$.

IV. Moderate Deviations: Joint Scaling of Error Probability and Gap to Capacity

The scaling exponent describes how fast the gap to capacity tends to 0 as a function of the block length, when the error probability is fixed. Hence, it is natural to ask how fast the gap to capacity tends to 0 as a function of the block length, when the error probability tends to 0 at a certain speed. The discussion of Remark 4 in Section III-A points out that we can allow the error probability to go to 0 polynomially fast in N, while maintaining the same scaling exponent. In this section, we show that, if we allow a less favorable scaling between gap to capacity and block length, i.e. a larger scaling exponent, then the error probability goes to 0 sub-exponentially fast in N. More specifically, Section IV-A contains the exact statement of this result together with some remarks, and Section IV-B contains the proof.

A. Main Result: Statement and Discussion

Theorem 7 (Joint scaling: exponential decay of P_{e}): Assume that there exists a *fixed function* $h(x)$ which satisfies the hypotheses of Theorem 1 for some $\mu > 2$. Consider transmission over a BMSC W with capacity $I(W)$ using a polar code of rate $R < I(W)$. Then, for any $\gamma \in \left(\frac{1}{1+\mu}, 1\right)$, the block length N and the block error probability under successive cancellation decoding P_{e} are s.t.

$$\begin{aligned} P_{\mathrm{e}} &\leq N \cdot 2^{-N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}, \\ N &\leq \frac{\beta_3}{(I(W)-R)^{\mu/(1-\gamma)}}, \end{aligned} \tag{49}$$

where β_3 is a constant, and $h_2^{(-1)}$ is the inverse of the binary entropy function defined as $h_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$ for any $x \in [0, 1/2]$. If W is a BEC, the less stringent hypothesis (15) on μ is required for (49) to hold.

In short, formula (49) describes a trade-off between gap to capacity and error probability as functions of the block length N. Indeed, let γ go from $\dfrac{1}{1+\mu}$ to 1: on the one hand, the error probability goes faster and faster to 0, since the exponent $\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)$ is increasing in γ; on the other hand, the gap to capacity goes slower to 0, since the exponent $\dfrac{\mu}{1-\gamma}$ is increasing in γ. Before proceeding with the proof, it is useful to discuss three points. The first remark concerns the possible choices for μ in (49). The second remark shows how to recover from Theorem 7 the result [13] concerning the error exponent regime. The third remark adds the Bhattacharyya parameter $Z(W)$ to the picture outlined in Theorem 7 and, in particular, it focuses on the dependency between P_{e} and $Z(W)$.

Remark 8 (Valid choice for μ in (49)*):* By constructing a fixed function $h(x)$ as in the proof of Theorem 2 contained in Section III-C, we immediately have that valid choices of μ in (49) are $\mu = 4.714$ for any BMSC and $\mu = 3.637$ for the special case of the BEC.

Remark 9 (Error exponent regime and Theorem 7): By picking γ close to 1, we recover the result [13] concerning the error exponent regime: if we allow the gap to capacity to be arbitrary small but independent of N, then P_{e} is $O(2^{-N^{\beta}})$ for any $\beta \in (0, 1/2)$[1]. On the other hand, note that it is not possible to recover from Theorem 7 the result of Theorem 1 concerning the scaling exponent regime. Indeed, pick γ close to $\dfrac{1}{1+\mu}$. Then, the exponent $\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)$ tends to 0, i.e., we approach a regime in which the error probability is independent of N, but N is $O\left(\dfrac{1}{(I(W)-R)^{\mu+1}}\right)$ instead of $O\left(\dfrac{1}{(I(W)-R)^{\mu}}\right)$ as in (14). We believe that this is only an artifact of the proof technique used to show Theorem 7 and that it might be possible to find a joint scaling which contains as special cases the error exponent and the scaling exponent regimes.

Remark 10 (Dependency between P_{e} and $Z(W)$): Consider transmission over a BMSC W with Bhattacharyya parameter $Z(W)$. Then, under the hypotheses of Theorem 7, it is possible to prove that

$$\begin{aligned} P_{\mathrm{e}} &\leq N \cdot Z(W)^{\frac{1}{2}\cdot N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}, \\ N &\leq \frac{\beta_4}{(I(W)-R)^{\mu/(1-\gamma)}}, \end{aligned} \tag{50}$$

where β_4 is a constant. A sketch of the proof of this statement is given in Appendix C. In short, the error probability scales as $Z(W)$ raised to some power of N, where the exponent follows the trade-off of Theorem 7. To see that this is a meaningful bound, consider the case of transmission over the BEC in the error exponent regime. On the one hand, formula (50) gives that P_{e} scales roughly as $Z(W)^{\sqrt{N}}$. On the other hand, $P_{\mathrm{e}} \geq \max_{i\in\mathcal{I}} Z_n^{(i)}$, where $\mathcal{I}$ denotes the set of information positions and $Z_n^{(i)}$ is a polynomial in $Z(W)$ with minimum degree which scales roughly as $\sqrt{N}$[2]. The scaling between the error probability and the Bhattacharyya parameter will be further explored in Section V.

B. Proof of Theorem 7

Proof: Let $Z_n(W)$ be the Bhattacharyya process associated to the channel W. Then, by following the same procedure that gives (34), we have that, for any $n_0 \in \mathbb{N}$,

$$\mathbb{P}\left(Z_{n_0} \leq 2^{-n_0}\right) \geq I(W) - c_5\, 2^{-n_0/\mu}, \tag{51}$$

where c_5 is a constant which does not depend on n and it is given by $c_5 = \sqrt{2} + 2/\delta$, with δ defined as in (32).

Let $\{B_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with distribution Bernoulli$\left(\frac{1}{2}\right)$. Then, by using (10), it is clear that, for $n \geq 1$,

$$Z_{n_0+n} \leq \begin{cases} Z_{n_0+n-1}^2, & \text{if } B_n = 1, \\ 2Z_{n_0+n-1}, & \text{if } B_n = 0. \end{cases}$$

[1]Theorem 7 contains as a particular case also the stronger result in [21], where the authors prove that the block length scales polynomially fast with the inverse of the gap to capacity and, at the same time, the error probability is upper bounded by $2^{-N^{0.49}}$.

[2]To see this, note that the minimum degree of $Z_n^{(i)}$ seen as a polynomial in $Z(W)$ is equal to the minimum distance of the code, which scales roughly as $\sqrt{N}$ according to Lemma 4 of [22].

Therefore, by applying Lemma 22 of [16], we obtain that, for $n_1 \geq 1$,

$$\mathbb{P}\left(Z_{n_0+n_1} \leq 2^{-2^{\sum_{i=1}^{n_1} B_i}} \mid Z_{n_0} = x\right) \geq 1 - c_6\, x(1-\log_2 x), \tag{52}$$

with $c_6 = \dfrac{2}{(\sqrt{2}-1)^2}$.

Consequently, we have that

$$\begin{aligned}
\mathbb{P}\left(Z_{n_0+n_1} \leq 2^{-2^{\sum_{i=1}^{n_1} B_i}}\right) &= \mathbb{P}\left(Z_{n_0} \leq 2^{-n_0}\right) \cdot \mathbb{P}\left(Z_{n_0+n_1} \leq 2^{-2^{\sum_{i=1}^{n_1} B_i}} \mid Z_{n_0} \leq 2^{-n_0}\right) \\
&\overset{\text{(a)}}{\geq} \mathbb{P}\left(Z_{n_0} \leq 2^{-n_0}\right) \cdot \left(1 - c_6\, 2^{-n_0}(1+n_0)\right) \\
&\overset{\text{(b)}}{\geq} \left(I(W) - c_5\, 2^{-n_0/\mu}\right) \cdot \left(1 - c_6 \frac{\sqrt{2}}{\ln 2}\, 2^{-n_0/2}\right) \\
&\overset{\text{(c)}}{\geq} I(W) - \left(c_5 + c_6 \frac{\sqrt{2}}{\ln 2}\right) 2^{-n_0/\mu},
\end{aligned} \tag{53}$$

where (a) uses (52) and the fact that $1 - c_6\, x(1-\log_2 x)$ is decreasing in x for any $x \leq 2^{-n_0} \leq 1/2$, (b) uses (51) and that $1 - c_6\, 2^{-n_0}(1+n_0) \geq 1 - c_6 \dfrac{\sqrt{2}}{\ln 2}\, 2^{-n_0/2}$ for any $n_0 \in \mathbb{N}$, and (c) uses that $\mu > 2$.

Let $h_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$ denote the binary entropy function. Then, for any $\epsilon \in (0, 1/2)$

$$\begin{aligned}
\mathbb{P}\left(2^{-2^{\sum_{i=1}^{n_1} B_i}} > 2^{-2^{n_1\epsilon}}\right) &= \mathbb{P}\left(\sum_{i=1}^{n_1} B_i < n_1\epsilon\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{n_1} B_i \leq \lfloor n_1\epsilon \rfloor\right) \\
&= \sum_{k=0}^{\lfloor n_1\epsilon \rfloor} \binom{n_1}{k} \left(\frac{1}{2}\right)^{n_1} \\
&\overset{\text{(a)}}{\leq} \left(\frac{1}{2}\right)^{n_1} 2^{n_1 h_2(\lfloor n_1\epsilon \rfloor / n_1)} \\
&\overset{\text{(b)}}{\leq} 2^{-n_1(1-h_2(\epsilon))},
\end{aligned} \tag{54}$$

where (a) uses formula (1.59) of [9], and (b) we uses that $h_2(x)$ is increasing for any $x \leq 1/2$.

Note that, for any two events A and B, $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$. Hence, by combining (53) and (54), we obtain that

$$\mathbb{P}\left(Z_{n_0+n_1} \leq 2^{-2^{n_1\epsilon}}\right) \geq I(W) - \left(c_5 + c_6 \frac{\sqrt{2}}{\ln 2}\right) 2^{-n_0/\mu} - 2^{-n_1(1-h_2(\epsilon))}. \tag{55}$$

Let $n \geq 1$. Set $n_1 = \lceil \gamma n \rceil$, $n_0 = n - \lceil \gamma n \rceil$, and $\epsilon = h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)$, where $h_2^{(-1)}(\cdot)$ is the inverse of $h_2(x)$ for any $x \in [0, 1/2]$. Note that if $\gamma \in \left(\dfrac{1}{1+\mu}, 1\right)$, then $\epsilon \in (0, 1/2)$. Consequently, formula (55) can be rewritten as

$$\mathbb{P}\left(Z_{n_0+n_1} \leq 2^{-2^{n\gamma h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}\right) \geq I(W) - c_7\, 2^{-n\frac{1-\gamma}{\mu}}, \tag{56}$$

with $c_7 = 1 + \sqrt{2}\left(c_5 + c_6 \dfrac{\sqrt{2}}{\ln 2}\right)$.

Consider transmission of a polar code of block length $N = 2^n$ and rate R given by the RHS of (56). Then, the result (49) holds with $\beta_3 = c_7^{\mu}$.

■

V. Absence of Error Floors

The discussion of Remark 10 in Section IV-A studies the dependency between the error probability and the Bhattacharyya parameter and considers a setting in which, as the channel varies, the polar code used for transmission changes accordingly. In this section, we consider a different scenario in which the polar code stays fixed as the channel varies and we prove a result about the speed of decay of the error probability as a function of the Bhattacharyya parameter of the channel. By doing so, we conclude that polar codes are not affected by error floors. More specifically, Section V-A formalizes and discusses this result, and Section V-B contains the proof.

A. Main Result: Statement and Discussion

Let $\mathcal{C}$ be the polar code with information set $\mathcal{I}$ designed for transmission over the BMSC W' with Bhattacharyya parameter $Z(W')$. Then, the actual channel over which transmission takes place is the BMSC W with Bhattacharyya parameter $Z(W)$. In the error floor regime, the code $\mathcal{C}$ is fixed and W varies, and we study the scaling between the error probability under SC decoding and the Bhattacharyya parameter $Z(W)$.

The main result is presented in Theorem 11 and it relates the Bhattacharyya parameter $Z_n^{(i)}(W)$ obtained by polarizing W to the Bhattacharyya parameter $Z_n^{(i)}(W')$ at the same position obtained by polarizing W'. From this, in Corollary 12 we relate the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W, i.e., $\tilde{P}_{\mathrm{e}}(W) \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W)$, to the sum of Bhattacharyya parameters obtained by polarizing W', i.e., $\tilde{P}_{\mathrm{e}}(W') \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W')$. Note that the indices of the information positions are the same in both sums, since the information set $\mathcal{I}$ is fixed. The proof of Theorem 11 is in Section V-B, and the proof of Corollary 12 easily follows.

Theorem 11 (Scaling of $Z_n^{(i)}(W)$): Consider two BMSCs W and W' with Bhattacharyya parameter $Z(W)$ and $Z(W')$, respectively. For $n \in \mathbb{N}$ and $i \in \{1, \cdots, 2^n\}$, let $Z_n^{(i)}(W)$ be the Bhattacharyya parameter of the channel $W_n^{(i)}$ obtained from W via channel polarization and let $Z_n^{(i)}(W')$ be similarly obtained from W'. If $Z(W) \leq Z(W')^2$, then

$$Z_n^{(i)}(W) \leq Z_n^{(i)}(W')^{\frac{\log_2 Z(W)}{\log_2 Z(W')}}. \tag{57}$$

If W and W' are BECs, then (57) holds if $Z(W) \leq Z(W')$.

Corollary 12 (Scaling of $\tilde{P}_{\mathrm{e}}(W)$): Let W' be a BMSC with Bhattacharyya parameter $Z(W')$ and let $\mathcal{C}$ be the polar code of block length $N = 2^n$ and rate R for transmission over W'. Denote by $\tilde{P}_{\mathrm{e}}(W')$ the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W', i.e., $\tilde{P}_{\mathrm{e}}(W') \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W')$, where $\mathcal{I}$ is the information set of the polar code $\mathcal{C}$. Now, consider transmission over the BMSC W with Bhattacharyya parameter $Z(W)$ using the polar code $\mathcal{C}$ and let $\tilde{P}_{\mathrm{e}}(W)$ be the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W, i.e., $\tilde{P}_{\mathrm{e}}(W) \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W)$. If $Z(W) \leq Z(W')^2$, then

$$\tilde{P}_{\mathrm{e}}(W) \leq \tilde{P}_{\mathrm{e}}(W')^{\frac{\log_2 Z(W)}{\log_2 Z(W')}}. \tag{58}$$

If W and W' are BECs, then (58) holds if $Z(W) \leq Z(W')$.

Now, let us discuss how the results above imply that polar codes are not affected by error floors. Denote by $P_{\mathrm{e}}(W)$ the error probability under SC decoding for transmission of $\mathcal{C}$ over W and recall from (12) that $P_{\mathrm{e}}(W) \leq \tilde{P}_{\mathrm{e}}(W)$. Hence, formula (58) implies that

$$P_{\mathrm{e}}(W) \leq Z(W)^{\frac{\log_2 \tilde{P}_{\mathrm{e}}(W')}{\log_2 Z(W')}}. \tag{59}$$

Note that the upper bound (50) on P_{e} comes from an identical upper bound on the sum of the Bhattacharyya parameters $\tilde{P}_{\mathrm{e}}$. Thus, $\tilde{P}_{\mathrm{e}}(W')$ scales roughly as $Z(W')^{\sqrt{N}}$ and, therefore, from (59) we conclude that $P_{\mathrm{e}}(W)$ scales roughly as $Z(W)^{\sqrt{N}}$, which excludes the existence of an error floor region.

Recall that in the error floor regime we fix a polar code and we let the transmission channel vary. From the discussion above, it follows that the dependency between the error probability and the Bhattacharyya parameter of the channel is essentially the same as if we designed the polar code for the actual transmission channel. As a result, in terms of this particular scaling, nothing is lost by considering a "mismatched" code. On the other hand, considering a "mismatched" code yields a loss in rate. Indeed, if W and W' are BECs, then (2) holds with equality and $Z(W) \leq Z(W')$ implies that $I(W) \geq I(W')$. If W and W' can be any BMSC, by using (2) and (3) we

easily deduce that $Z(W) \leq Z(W')^2$ implies $I(W) \geq I(W')$. Consequently, the rate of a polar code for W' is s.t. $R < I(W')$, while by constructing a polar code for W we have the less restrictive condition $R < I(W)$.

Before proceeding with the proof of Theorem 11, let us make a brief remark concerning the case $Z(W) \in \big(Z(W')^2, Z(W')\big]$.

Remark 13 (The case $Z(W) \in (Z(W')^2, Z(W')]$): If W and W' are BECs, then (57) and (58) hold for any $Z(W) \leq Z(W')$, i.e., for the whole range of parameters of interest as we think to W as a "better" channel than W'. On the other hand, if W and W' can be any BMSC, we require that $Z(W) \leq Z(W')^2$. If there is no additional hypothesis on W and W', the main result (57) cannot hold in the case $Z(W) \in (Z(W')^2, Z(W')]$. Indeed, if $Z(W) = Z(W')$, we can choose W and W' s.t. $I(W) < I(W')$. If $I(W) < I(W')$, then the number of indices i_1 s.t. $\lim_{n\to\infty} Z_n^{(i_1)}(W) = 0$ is smaller than the number of indices i_2 s.t. $\lim_{n\to\infty} Z_n^{(i_2)}(W') = 0$. Hence, (57) cannot hold for any $i \in \{1, \cdots, 2^n\}$. A natural additional hypothesis consists in assuming that W' is degraded with respect to W, i.e., $W \succ W'$. In this case, we can at least ensure that $Z_n^{(i)}(W) \leq Z_n^{(i)}(W')$. However, it is possible to find W and W' s.t. (57) is violated for $n = 1$ when $Z(W) \in (Z(W')^2, Z(W')]$. The questions concerning whether the bound (58) is still valid or what kind of looser bound holds when $W \succ W'$ and $Z(W) \in (Z(W')^2, Z(W')]$ remain open.

B. Proof of Theorem 11

Proof: Assume that for any $j \in \{1, \cdots, 2^{n-1}\}$ and for some $\eta \in \mathbb{R}^+$,

$$Z_{n-1}^{(j)}(W) \leq Z_{n-1}^{(j)}(W')^{\eta}. \tag{60}$$

Then, let us study for what values of η we have that (60) implies that, for any $i \in \{1, \cdots, 2^n\}$,

$$Z_n^{(i)}(W) \leq Z_n^{(i)}(W')^{\eta}. \tag{61}$$

Recall from Section II that $(b_1^{(i)}, \cdots, b_n^{(i)})$ denotes the binary representation of the integer $i-1$ over n bits. Let i be an even integer and set $i^+ = \dfrac{i}{2}$. Then, $b_n^{(i)} = 1$ and the binary representation of $i^+ - 1$ over $n-1$ bits is $(b_1^{(i)}, \cdots, b_{n-1}^{(i)})$. Hence, the following chain of inequalities holds for any BMSC W:

$$\begin{aligned} Z_n^{(i)}(W) &\overset{\text{(a)}}{=} \left(Z_{n-1}^{(i^+)}(W)\right)^2 \\ &\overset{\text{(b)}}{\leq} \left(Z_{n-1}^{(i^+)}(W')\right)^{2\eta} \\ &\overset{\text{(c)}}{=} \left(Z_n^{(i)}(W')\right)^{\eta}, \end{aligned} \tag{62}$$

where (a) uses (8) and (6), (b) uses the assumption (60) with $j = i^+$, and (c) uses again (8) and (6). Consequently, if i is even, then (61) holds for any BMSC W without any restriction on η.

Let i be an odd integer and set $i^- = \dfrac{i-1}{2}$. Then, $b_n^{(i)} = 0$ and the binary representation of $i^- - 1$ over $n-1$ bits is $(b_1^{(i)}, \cdots, b_{n-1}^{(i)})$. Hence, the following chain of inequalities holds for any BMSC W:

$$\begin{aligned} Z_n^{(i)}(W) &\overset{\text{(a)}}{\leq} Z_{n-1}^{(i^-)}(W)\left(2 - Z_{n-1}^{(i^-)}(W)\right) \\ &\overset{\text{(b)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\left(2 - \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\right) \\ &\overset{\text{(c)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\left(2 - \left(Z_{n-1}^{(i^-)}(W')\right)^{2}\right)^{\eta/2} \\ &\overset{\text{(d)}}{\leq} \left(Z_n^{(i)}(W')\right)^{\eta}, \end{aligned} \tag{63}$$

where (a) uses (8) and (5), (b) uses the assumption (60) with $j = i^-$, (c) uses that $2 - x^{\eta} \leq (2 - x^2)^{\eta/2}$ for any $x \in [0,1]$ if and only if $\eta \geq 2$, and (d) uses again (8) and (5). Consequently, if i is odd, then (61) holds for any

BMSC W provided that $\eta \geq 2$. If W is a BEC, a less restrictive condition on η is necessary. Indeed, the following chain of inequalities holds when W is a BEC:

$$\begin{aligned} Z_n^{(i)}(W) &\overset{\text{(a)}}{=} Z_{n-1}^{(i^-)}(W)\left(2 - Z_{n-1}^{(i^-)}(W)\right) \\ &\overset{\text{(b)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\left(2 - \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\right) \\ &\overset{\text{(c)}}{\leq} \left(Z_{n-1}^{(i^-)}(W')\right)^{\eta}\left(2 - Z_{n-1}^{(i^-)}(W')\right)^{\eta} \\ &\overset{\text{(d)}}{=} \left(Z_n^{(i)}(W')\right)^{\eta}, \end{aligned} \tag{64}$$

where (a) uses (8) and (7), (b) uses the assumption (60) with $j = i^-$, (c) uses that $2 - x^\eta \leq (2-x)^\eta$ for any $x \in [0,1]$ if and only if $\eta \geq 1$, and (d) uses again (8) and (7). Consequently, if i is odd and W is a BEC, then (61) holds provided that $\eta \geq 1$.

By combining (62) and (63), we have that if (60) holds for $\eta \geq 2$ after $n-1$ steps of polarization, then the same relation holds for $\eta \geq 2$ after n steps of polarization, namely, the inequality stays preserved after one more step of polarization. Clearly, as the Bhattacharyya parameter is between 0 and 1, a smaller value of η gives a tighter bound. Since $Z_0^{(1)}(W) = Z(W)$ and $Z_0^{(1)}(W') = Z(W')$, the smallest choice for η is $\dfrac{\log_2 Z(W)}{\log_2 Z(W')}$. The condition $\eta \geq 2$ is equivalent to $Z(W) \leq Z(W')^2$ and, for the case of the BEC, the condition $\eta \geq 1$ is equivalent to $Z(W) \leq Z(W')$. Eventually, the result (57) follows easily by induction.

■

VI. Concluding Remarks

This paper presents a unified view on the scaling of polar codes by studying the relation among the fundamental parameters at play, i.e., the block length N, the rate R, the error probability under Successive Cancellation (SC) decoding P_{e}, the capacity of the transmission channel $I(W)$ and its Bhattacharyya parameter $Z(W)$. Let us summarize the main results contained in this work, along with open questions and directions for future research.

First of all, we prove a new upper bound on the scaling exponent for any BMSC W. The setting is the following: we fix the error probability P_{e} and we study how the gap to capacity $I(W) - R$ scales with the block length N. In particular, N is $O\left(\dfrac{1}{(I(W)-R)^\mu}\right)$, where μ is the so-called scaling exponent whose value depends on W, and we show a better upper bound on μ valid for any BMSC W. The proof technique consists in relating the value of μ to the $\sup$ of a function which fulfills certain constraints. Then, we upper bound the $\sup$ by constructing and analyzing a suitable candidate function. Let us underline that the proposed bound is *provable* and that the analysis of the algorithm is not affected by numerical errors, since all the computations can be reduced to computations over integers and, therefore, they can be performed exactly. The proposed proof technique yields $\mu \leq 4.714$ for any BMSC, which essentially improves by 1 the existing upper bound. If W is a BEC, we obtain $\mu \leq 3.639$, which approaches the value previously computed with heuristic methods. These bounds can be slightly tightened simply by increasing the number of samples used by the algorithm. Possibly the most interesting open question concerning the performance of polar codes consists in improving the scaling exponent, i.e., the speed of decay of the gap to capacity, by changing the construction of the code and by devising better decoding algorithms. One promising method consists in constructing a code which interpolates between a polar and a Reed-Muller code and in using the MAP decoder or even the low-complexity SCL decoder [23].

Secondly, we consider a moderate deviations regime and we prove a trade-off between the speed of decay of the error probability and that of the gap to capacity. The setting is the following: we do not fix either the error probability P_{e} or the gap to capacity $I(W)-R$, but we study how fast both P_{e} and $I(W)-R$ go to 0 at the same time as functions of the block length N. In particular, we show that, if the gap to capacity is s.t. N is $O\left(\dfrac{1}{(I(W)-R)^{\mu/(1-\gamma)}}\right)$ for $\gamma \in \left(\dfrac{1}{1+\mu}, 1\right)$, then the error probability P_{e} is $O(N \cdot 2^{-N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}})$. Note that, since the exponents $\dfrac{\mu}{1-\gamma}$ and $\gamma \cdot h_2^{(-1)}\left(\dfrac{\gamma(\mu+1)-1}{\gamma\mu}\right)$ are both increasing in γ, if the error probability decays faster, then the gap to capacity

decays slower. This trade-off recovers the existing result for the error exponent regime, but it does not match the new bound on the scaling exponent. An interesting open question consists in finding the optimal trade-off which provides the fastest possible decay of the error probability, given a certain speed of decay of the gap to capacity. Note that this optimal trade-off would match the existing results for both the error exponent and the scaling exponent regimes.

Thirdly, we prove that polar codes are not affected by error floors. The setting is the following: we fix a polar code of block length N and rate R designed for a channel W', we let the transmission channel W vary, and we study how the error probability $P_{\mathrm{e}}(W)$ scales with the Bhattacharyya parameter $Z(W)$ of the channel W. In particular, we show that $P_{\mathrm{e}}(W) \leq Z(W)^{\frac{\log_2 \tilde{P}_{\mathrm{e}}(W')}{\log_2 Z(W')}}$, where $\tilde{P}_{\mathrm{e}}(W')$ denotes the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W'. In addition, $\dfrac{\log_2 \tilde{P}_{\mathrm{e}}(W')}{\log_2 Z(W')}$ scales roughly as $\sqrt{N}$, which is the best possible scaling according to the error exponent regime. Hence, the scaling between P_{e} and $Z(W)$ would have been the same even if we "matched" the code to the channel. However, when W and W' can be any BMSC, the result holds only if $Z(W) \leq Z(W')^2$. An interesting open question is to explore further the case $Z(W) \in (Z(W')^2, Z(W')]$, in order to see whether a similar but perhaps less tight bound still holds.

Finally, let us point out that the techniques described in this paper could be useful in the analysis of polar codes with kernels larger than the 2×2 matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. The study of polar codes with general $\ell \times \ell$ kernels aims at improving the scaling behavior. As concerns the error exponent, in [24] it is proved that, as ℓ goes large, the error probability scales roughly as 2^{-N}. As concerns the scaling exponent, in [25] it is observed that μ can be reduced when $\ell \geq 8$ and in [26] the author provides evidence that the $\mu \to 2$ as $\ell \to \infty$.

APPENDIX

A. Proof of Lemma 5

Proof: First of all, we upper bound $\mathbb{P}(Z_n \in [p_{\mathrm{e}} 2^{-n}, 1 - p_{\mathrm{e}} 2^{-n}])$ as follows:

$$\begin{aligned}
\mathbb{P}\left(Z_n \in \left[p_{\mathrm{e}} 2^{-n}, 1 - p_{\mathrm{e}} 2^{-n}\right]\right) &\overset{\text{(a)}}{=} \mathbb{P}\left((Z_n(1-Z_n))^\alpha \geq (p_{\mathrm{e}} 2^{-n}(1 - p_{\mathrm{e}} 2^{-n}))^\alpha\right) \\
&\overset{\text{(b)}}{\leq} \frac{\mathbb{E}\left[(Z_n(1-Z_n))^\alpha\right]}{(p_{\mathrm{e}} 2^{-n}(1 - p_{\mathrm{e}} 2^{-n}))^\alpha} \\
&\overset{\text{(c)}}{\leq} \frac{c_1\, 2^{-n\rho}}{(p_{\mathrm{e}} 2^{-n}(1 - p_{\mathrm{e}} 2^{-n}))^\alpha} \\
&\overset{\text{(d)}}{\leq} 2\, c_1\, p_{\mathrm{e}}^{-\alpha}\, 2^{-n(\rho-\alpha)},
\end{aligned} \tag{65}$$

where (a) uses the concavity of the function $f(x) = (x(1-x))^\alpha$, (b) follows from Markov inequality, (c) uses the hypothesis $\mathbb{E}[(Z_n(1-Z_n))^\alpha] \leq c_1\, 2^{-n\rho}$, and (d) uses that $1 - p_{\mathrm{e}} 2^{-n} \geq \frac{1}{2}$ for any $n \geq 1$.

Let us define

$$\begin{aligned}
A &= \mathbb{P}\left(Z_n \in \left[0, p_{\mathrm{e}} 2^{-n}\right)\right), \\
B &= \mathbb{P}\left(Z_n \in \left[p_{\mathrm{e}} 2^{-n}, 1 - p_{\mathrm{e}} 2^{-n}\right]\right), \\
C &= \mathbb{P}\left(Z_n \in \left(1 - p_{\mathrm{e}} 2^{-n}, 1\right]\right),
\end{aligned} \tag{66}$$

and let A', B', and C' be the fraction of A, B, and C, respectively, that will go to 0 as $n \to \infty$. More formally,

$$\begin{aligned}
A' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_n \in \left[0, p_{\mathrm{e}} 2^{-n}\right), Z_{n+m} \leq 2^{-m}\right), \\
B' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_n \in \left[p_{\mathrm{e}} 2^{-n}, 1 - p_{\mathrm{e}} 2^{-n}\right], Z_{n+m} \leq 2^{-m}\right), \\
C' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_n \in \left(1 - p_{\mathrm{e}} 2^{-n}, 1\right], Z_{n+m} \leq 2^{-m}\right).
\end{aligned} \tag{67}$$

In (67) we simply require that Z_{n+m} goes to 0 as m goes large, and we do not have any requirement on the speed at which it does so. Hence, we could substitute 2^{-m} with any other function which is $O(2^{-2^{\beta m}})$ for any $\beta \in (0, 1/2)$, see [13].

It is clear that

$$A' + B' + C' = \liminf_{m\to\infty} \mathbb{P}\left(Z_{n+m} \leq 2^{-m}\right) = I(W). \tag{68}$$

In addition, from (65), we have that

$$B' \leq B \leq 2\,c_1\,p_{\mathrm{e}}^{-\alpha}\,2^{-n(\rho-\alpha)}. \tag{69}$$

In order to upper bound C', we proceed as follows:

$$\begin{aligned} C' &= \liminf_{m\to\infty} \mathbb{P}\left(Z_{n+m} \leq 2^{-m} \mid Z_n \in \left(1-p_{\mathrm{e}}\,2^{-n}, 1\right]\right) \cdot \mathbb{P}\left(Z_n \in \left(1-p_{\mathrm{e}}\,2^{-n}, 1\right]\right) \\ &\leq \liminf_{m\to\infty} \mathbb{P}\left(Z_{n+m} \leq 2^{-m} \mid Z_n \in \left(1-p_{\mathrm{e}}\,2^{-n}, 1\right]\right). \end{aligned} \tag{70}$$

The last term equals the capacity of a channel with Bhattacharyya parameter in the interval $(1-p_{\mathrm{e}}\,2^{-n}, 1]$. Using (3), we obtain that

$$C' \leq \sqrt{1-(1-p_{\mathrm{e}}\,2^{-n})^2} \leq \sqrt{2p_{\mathrm{e}}\,2^{-n}}. \tag{71}$$

As a result, we have that

$$\begin{aligned} \mathbb{P}\left(Z_n \in \left[0, p_{\mathrm{e}}\,2^{-n}\right)\right) = A &\geq A' \\ &\overset{\text{(a)}}{=} I(W) - B' - C' \\ &\overset{\text{(b)}}{\geq} I(W) - 2\,c_1\,p_{\mathrm{e}}^{-\alpha}\,2^{-n(\rho-\alpha)} - \sqrt{2p_{\mathrm{e}}\,2^{-n}}, \\ &\overset{\text{(c)}}{\geq} I(W) - \left(\sqrt{2p_{\mathrm{e}}} + 2\,c_1\,p_{\mathrm{e}}^{-\alpha}\right) 2^{-n(\rho-\alpha)}, \end{aligned}$$

where (a) uses (68), (b) uses (69) and (71) and (c) uses that $\rho \leq 1/2$. This chain of inequalities implies the desired result.

■

B. Proof of Lemma 6

Proof: Let $\alpha^* = \min(1/2, \rho_1/\log_2(4/3))$. As $\mathbb{E}\left[\left(Z_n(1-Z_n)\right)^{\alpha}\right]$ is decreasing in α, we can assume that $\alpha < \alpha^*$ without loss of generality. Since $h(x) \geq 0$ for any $x \in [0,1]$ and $Z_n \in [0,1]$ for any $n \in \mathbb{N}$, we have that

$$\mathbb{E}\left[\left(Z_n(1-Z_n)\right)^{\alpha}\right] \leq \frac{1}{\delta}\mathbb{E}\left[(1-\delta)h(Z_n) + \delta(Z_n(1-Z_n))^{\alpha}\right] = \frac{1}{\delta}\mathbb{E}\left[g(Z_n)\right], \tag{72}$$

with

$$g(x) = (1-\delta)h(x) + \delta(x(1-x))^{\alpha}. \tag{73}$$

Let

$$L_g = \sup_{x\in(0,1), y\in[x\sqrt{2-x^2}, 2x-x^2]} \frac{g(x^2)+g(y)}{2g(x)}.$$

Then, by definition (10) of the Bhattacharyya process Z_n, we have that

$$\mathbb{E}\left[g(Z_n) \mid Z_{n-1}\right] \leq g(Z_{n-1})L_g.$$

Consequently, by induction, one can readily prove that

$$\mathbb{E}\left[g(Z_n)\right] \leq (L_g)^n\, g(Z(W)) \leq (L_g)^n, \tag{74}$$

where the last inequality follows from the fact that $g(x) \leq 1$ for $x \in [0,1]$.

Now, by combining (72) with (74), we obtain that

$$\mathbb{E}[(Z_n(1-Z_n))^{\alpha}] \leq \frac{1}{\delta}(L_g)^n. \tag{75}$$

Hence, to conclude the proof it remains to find an upper bound on L_g (i.e., to show that $L_g \leq 2^{-\rho_1} + 2\sqrt{2}\delta c_3$). By using (25), after some calculations, we have that

$$\frac{g(x^2)+g(y)}{2g(x)} \leq \frac{(1-\delta)h(x)2^{-\rho_1} + \frac{\delta}{2}\Big(\left(x^2(1-x)(1+x)\right)^{\alpha} + \left(y(1-y)\right)^{\alpha}\Big)}{(1-\delta)h(x) + \delta(x(1-x))^{\alpha}}. \tag{76}$$

For any $y \in [x\sqrt{2-x^2}, 2x-x^2]$, we get

$$y(1-y) \leq x(2-x)(1-x\sqrt{2-x^2}). \tag{77}$$

In addition, for any $x \in (0,1)$,

$$1 - x\sqrt{2-x^2} \leq (1-x)^{4/3}. \tag{78}$$

In order to prove (78), one strategy is the following: elevate the LHS and the RHS to the third power, isolate on one side the terms which multiply $\sqrt{2-x^2}$, and square again the LHS and the RHS. In this way, we have that (78) is equivalent to

$$(1-x)^4(2+8x+3x^2+4x^3-4x^4-4x^5-x^6) \geq 0,$$

which is clearly satisfied when $x \in (0,1)$.

Therefore, by combining (76), (77), and (78), we obtain that

$$\frac{g(x^2)+g(y)}{2g(x)} \leq \frac{(1-\delta)h(x)2^{-\rho_1} + \delta(x(1-x))^\alpha t(x)}{(1-\delta)h(x)+\delta(x(1-x))^\alpha}, \tag{79}$$

with

$$t(x) = \frac{1}{2}\left(\big(x(1+x)\big)^\alpha + \big((2-x)(1-x)^{1/3}\big)^\alpha\right). \tag{80}$$

First of all, we upper bound the expression on the RHS of (79) when x is small. Clearly, $t(0) < 2^{-\rho_1}$ and $t(1/2) > 2^{-\rho_1}$, since $\rho_1 \leq 0.5$ and $\alpha < \alpha^*$. In addition, some passages of calculus show that the second derivative of $t(x)$ is given by

$$\frac{\alpha}{2}\frac{(x(1+x))^\alpha}{x^2(1+x)^2}\left(-1-2x-2x^2+\alpha(1+2x)^2\right) + \frac{\alpha}{18}\frac{\big((2-x)(1-x)^{1/3}\big)^\alpha}{(2-3x+x^2)^2}\left(-21+30x-12x^2+\alpha(5-4x)^2\right).$$

Since $\alpha < 1/2$, we have that

$$\begin{aligned} -1-2x-2x^2+\alpha(1+2x)^2 &\leq -1-2x-2x^2+\frac{(1+2x)^2}{2} < 0, \\ -21+30x-12x^2+\alpha(5-4x)^2 &\leq -1-2x-2x^2+\frac{(5-4x)^2}{2} < 0. \end{aligned} \tag{81}$$

Hence, $t(x)$ is concave for any $x \in (0,1)$ and, therefore, there exist $\epsilon_1(\alpha), \epsilon_2(\alpha) \in (0,1)$ s.t.

$$t(x) \leq 2^{-\rho_1}, \qquad \forall\, x \in [0, \epsilon_1(\alpha)] \cup [1-\epsilon_2(\alpha), 1]. \tag{82}$$

Indeed, the precise values of $\epsilon_1(\alpha)$ and $\epsilon_2(\alpha)$ can be found from (28). By combining (79) with (82), we have that, for any $x \in [0, \epsilon_1(\alpha)] \cup [1-\epsilon_2(\alpha), 1]$ and for any $y \in [x\sqrt{2-x^2}, 2x-x^2]$,

$$\frac{g(x^2)+g(y)}{2g(x)} \leq 2^{-\rho_1}. \tag{83}$$

Then, we upper bound the expression on the RHS of (79) when x is not too small, namely, $x \in (\epsilon_1(\alpha), 1-\epsilon_2(\alpha))$:

$$\begin{aligned} \frac{(1-\delta)h(x)2^{-\rho_1} + \delta(x(1-x))^\alpha t(x)}{(1-\delta)h(x)+\delta(x(1-x))^\alpha} &\overset{\text{(a)}}{\leq} \frac{(1-\delta)h(x)2^{-\rho_1} + \delta(x(1-x))^\alpha 2^\alpha}{(1-\delta)h(x)+\delta(x(1-x))^\alpha} \\ &\overset{\text{(b)}}{\leq} 2^{-\rho_1} + \delta\frac{2^\alpha}{1-\delta}\frac{(x(1-x))^\alpha}{h(x)} \\ &\overset{\text{(c)}}{\leq} 2^{-\rho_1} + \sqrt{2}\frac{\delta}{1-\delta}c_3, \end{aligned} \tag{84}$$

where (a) uses that $t(x) \leq 2^\alpha$ for any $x \in (0,1)$, (b) uses that $h(x) \geq 0$ and $(x(1-x))^\alpha \geq 0$, and (c) uses that $\alpha \leq 1/2$, and the definition of c_3 in (27). By putting (83) and (84) together, we have that

$$L_g \leq 2^{-\rho_1} + \sqrt{2}\frac{\delta}{1-\delta}c_3. \tag{85}$$

By combining (75) and (85), the result for a general BMSC follows.

Finally, consider the special case in which W is a BEC. Clearly, (72) still holds, and, by using the definition (11) of the Bhattacharyya process Z_n for the BEC, in analogy to (74), we obtain that

$$\mathbb{E}[(Z_n(1-Z_n))^\alpha] \leq \frac{1}{\delta}(L'_g)^n, \tag{86}$$

where we define

$$L'_g = \sup_{x\in(0,1)} \frac{g(x^2)+g(2x-x^2)}{2g(x)}.$$

By using (29), after some calculations, we have that

$$\frac{g_0(x^2)+g_0(2x-x^2)}{2g_0(x)} \leq \frac{(1-\delta)h(x)2^{-\rho_1}+\delta(x(1-x))^\alpha t'(x)}{(1-\delta)h(x)+\delta(x(1-x))^\alpha},$$

with

$$t'(x) = \frac{1}{2}\left(\big(x(1+x)\big)^\alpha + \big((2-x)(1-x)\big)^\alpha\right).$$

Since $(1-x) \leq (1-x)^{1/3}$ for any $x \in (0,1)$, we obtain that $t'(x) \leq t(x)$ defined in (80), and, therefore, the result for the BEC easily follows. ■

C. Sketch of the Proof of (50)

Eventually, let us briefly sketch how to prove the result stated in Remark 10. The dependency on the Bhattacharyya parameter $Z(W)$ first appears in formula (74). Hence, under the hypothesis of Lemma 6, one can easily prove that

$$\mathbb{E}\left[(Z_n(1-Z_n))^\alpha\right] \leq \frac{g(Z(W))}{\delta}\left(2^{-\rho_1}+\sqrt{2}\frac{\delta}{1-\delta}c_3\right)^n, \tag{87}$$

where $g(x)$ is defined in (73). Consequently, by following passages similar to those in the proof of Lemma 5 in Appendix A and of Theorem 1 in Section III-B, we conclude that

$$\mathbb{P}\left(Z_{n_0} \leq Z(W)\cdot 2^{-2n_0}\right) \geq I(W) - c_8\, 2^{-n_0/\mu}, \tag{88}$$

where c_8 is a constant. Note that in formula (52) $Z_{n_0+n_1}$ is upper bounded by a quantity which does not depend on x. In order to make this dependency appear, we use passages similar to those of the proof of Lemma 22 in [16], thus obtaining that

$$\mathbb{P}\left(Z_{n_0+n_1} \leq x^{\frac{1}{2}\cdot 2^{\sum_{i=1}^{n_1} B_i}} \mid Z_{n_0} = x\right) \geq 1 - c_9\sqrt{x}(1-\log_2 x), \tag{89}$$

where c_9 is a constant. By combining (88) and (89), the result follows using the arguments similar to those of the proof of Theorem 7 in Section IV-B.

ACKNOWLEDGMENT

This work was supported by grant No. 200020_146832/1 of the Swiss National Science Foundation.

REFERENCES

[1] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[2] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability*, vol. 1, 1961, pp. 211–252.

[3] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, 1962, pp. 689–723.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite block-length regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[5] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[6] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 473–498, Feb. 2009.

[7] Y. Altug and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4417–4426, Aug. 2014.

[8] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 409–428, Mar. 1996.

[9] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[10] C. Di, D. Proietti, T. Richardson, E. Telatar, and R. Urbanke, "Finite length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1570–1579, Jun. 2002.

[11] T. Richardson, "Error floors of LDPC codes," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, vol. 41, no. 3, Monticello, IL, USA, 1998, pp. 1426–1435.

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

[13] E. Arıkan and I. E. Telatar, "On the rate of channel polarization," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seoul, South Korea, July 2009, pp. 1493–1495.

[14] S. H. Hassani, R. Mori, T. Tanaka, and R. Urbanke, "Rate-dependent analysis of the asymptotic behavior of channel polarization," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2267–2276, Apr. 2013.

[15] S. B. Korada, A. Montanari, I. E. Telatar, and R. Urbanke, "An empirical scaling law for polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, June 2010, pp. 884–888.

[16] S. H. Hassani, K. Alishahi, and R. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014.

[17] D. Goldin and D. Burshtein, "Improved bounds on the finite length scaling of polar codes," *IEEE Trans. Inform. Theory*, vol. 60, no. 11, pp. 6966–6978, Nov. 2014.

[18] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Aug. 2011, pp. 1–5.

[19] M. Mondelli, S. H. Hassani, and R. Urbanke, "Scaling exponent of list decoders with applications to polar codes," Jul. 2013, [Online]. Available: http://arxiv.org/pdf/1304.5220v3.pdf.

[20] A. Eslami and H. Pishro-Nik, "On finite-length performance of polar codes: stopping sets, error floor, and concatenated design," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 919–929, Mar. 2013.

[21] V. Guruswami and P. Xia, "Polar codes: speed of polarization and polynomial gap to capacity," Nov. 2013, [Online]. Available: http://arxiv.org/pdf/1304.4321v2.pdf.

[22] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, July 2009, pp. 1488–1492.

[23] M. Mondelli, S. H. Hassani, and R. Urbanke, "From polar to Reed-Muller codes: a technique to improve the finite-length performance," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3084–3091, Sept. 2014.

[24] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6253–6264, Dec. 2010.

[25] A. Fazeli and A. Vardy, "On the scaling exponent of binary polarization kernels," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, Oct. 2014.

[26] S. H. Hassani, "Polarization and spatial coupling: two techniques to boost performance," Ph.D. dissertation, EPFL, 2013.