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

Marco Mondelli and Rüdiger Urbanke
School of Computer and Communication Sciences
EPFL, CH-1015 Lausanne, Switzerland
Email: {marco.mondelli, ruediger.urbanke}@epfl.ch

S. Hamed Hassani
Computer Science Department
ETH Zürich, Switzerland
Email: hamed@inf.ethz.ch

***Abstract*—Consider transmission of a polar code of block length N and rate R over a binary memoryless symmetric channel W with capacity $I(W)$ and Bhattacharyya parameter $Z(W)$ and let $P_{\rm e}$ be the error probability under successive cancellation decoding. Recall that in the *error exponent* regime, the channel W and $R < I(W)$ are fixed, while $P_{\rm e}$ scales roughly as $2^{-\sqrt{N}}$. In the *scaling exponent* regime, the channel W and $P_{\rm e}$ are fixed, while the gap to capacity $I(W) - R$ scales as $N^{-1/\mu}$, with $3.579 \leq \mu \leq 5.702$ for any W. We develop a unified framework to characterize the relationship between R, N, $P_{\rm e}$, and W. First, we provide the tighter upper bound $\mu \leq 4.714$, valid for any W. Furthermore, when W is a binary erasure channel, we obtain an upper bound approaching very closely the value which was previously derived in a heuristic manner. Secondly, we consider a *moderate deviations* regime and we study how fast both the gap to capacity $I(W) - R$ and the error probability $P_{\rm e}$ simultaneously go to 0 as N goes large. 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 let the channel W vary, and we show that $P_{\rm e}$ scales roughly as $Z(W)^{\sqrt{N}}$.**

***Keywords*—Polar codes, error exponent, scaling exponent, moderate deviations, error floor.**

I. Introduction

The exact characterization of the relationship between the rate R, the block length N, the block error probability $P_{\rm e}$, and the quality of the transmission 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 them and by considering the relationship among the remaining ones. 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 $P_{\rm e}$. 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 as measured by, e.g., $Z(W)$ or $1 - I(W)$. Let us assume that there exists a threshold z^* such that, if $z < z^*$, then $P_{\rm e}$ tends to 0 as N grows large, while if $z > z^*$, then $P_{\rm e}$ tends to 1 as N grows large. For example, if the family of codes $\mathcal{C}$ is capacity 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 $z^* = 0.5$.

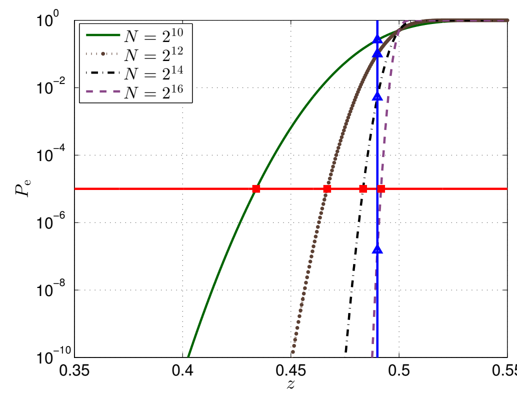

Figure 1. Performance of the family of codes $\mathcal{C}$ with rate $R = 0.5$ transmitted over the family of channels $\mathcal{W}$ with threshold $z^* = 0.5$.

The oldest approach to analyze the performance of the family $\mathcal{C}$ is known under the name of *error exponent*. We fix a channel parameter $z < z^*$ and we compute how fast the error probability tends to 0 as the block length goes large. This corresponds to consider the blue vertical cut in Figure 1. The best possible scaling is given by $P_{\rm e} = e^{-NE(R,W)+o(N)}$, where $E(R, W)$ is the so-called error exponent [1]. Another approach is known under the name of *scaling exponent*. We fix a target error probability $P_{\rm e}$ and we compute how fast the gap to the threshold $z^* - z$ tends to 0 as the block length goes large. This corresponds to consider the red horizontal cut in Figure 1. As a benchmark, the smallest possible block length N required to achieve a gap to the threshold $z^* - z$ with a fixed error probability $P_{\rm e}$ is s.t. [2], [3]

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

where $Q(\cdot)$ is the tail probability of the standard normal distribution and V is referred to as channel dispersion. In general, if N is $\Theta\left(1/(z^* - z)^{\mu}\right)$, we say that the family of codes $\mathcal{C}$ has scaling exponent μ. From (1) we deduce that the most favorable scaling exponent is $\mu = 2$ and it is achieved by random codes. To sum up, the error exponent regime studies $P_{\rm e}$ as a function of N when $z^* - z$ is fixed, while the scaling

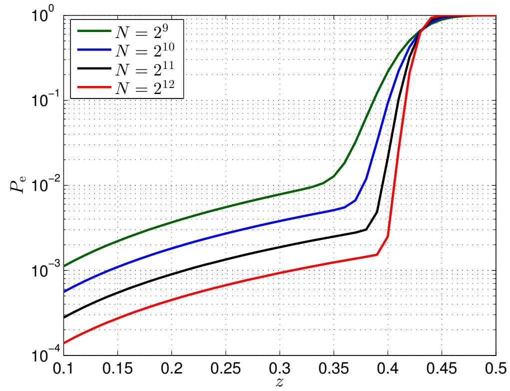

Figure 2. Performance of the family of $(3, 6)$-regular LDPC codes transmitted over the binary erasure channel with erasure probability z.

exponent regime studies $z^* - z$ as a function of N when $P_{\rm e}$ is fixed. Then, a natural question is to ask how fast *both* $P_{\rm e}$ *and* $z^* - z$ scale as functions of N. This intermediate approach is named *moderate deviations* regime and it is studied for random codes in [4]. The last scaling approach we consider concerns the so-called *error floor*. We fix a code of assigned block length N and rate R. Then, we compute how the error probability $P_{\rm e}$ behaves as a function of the channel parameter z. This is a notion that became important when iterative coding schemes were introduced [5]. For such schemes it was observed that frequently the individual curves $P_{\rm e}(z)$ show an abrupt change of slope from very steep to very shallow when going from bad channels to good channels (see, e.g., Figure 2). The region where the slope is very shallow was dubbed the error floor region. In this paper, we are going to adopt an alternative characterization that is easier to pin down mathematically: in the error floor region the slope of $P_{\rm e}(z)$ does not depend on N. We will show that this phenomenon does not happen for polar codes. In this sense we will say that polar codes do not have an error floor.

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 [6], the performance of polar codes has been extensively studied in different regimes. As concerns the *error exponent* regime, in [7] 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 [8], where it is shown that $\log_2(-\log_2 P_{\rm e})$ scales as $\log_2 N/2 + \sqrt{\log_2 N}/2 \cdot Q^{-1}(R/C) + o(\sqrt{\log_2 N})$. This last result holds both under SC decoding and under optimal MAP decoding. As concerns the *scaling exponent* regime[1], the value of μ depends on the particular channel taken into account. Universal bounds on μ valid for any BMSC under SC decoding are presented in [10]: the scaling exponent is lower bounded by 3.579 and it is upper bounded by 6. It is also conjectured that the lower bound on μ can be increased up to 3.627, i.e., up to the value heuristically computed in [9] for the binary erasure channel (BEC) . The upper bound on μ is further refined to 5.702 in [11]. In addition, the scaling exponent of list decoders is considered in [12] and in [13] it is proved that there exists a scaling exponent even if we let the error probability scale as $2^{-N^{0.49}}$. As concerns the *error floor* regime, in [14] 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 for the BEC, the existing results cannot rigorously exclude the existence of an error floor region.

This paper provides a unified view on the performance analysis of polar codes and it 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 $P_{\rm e}$, 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. First, we derive a new universal upper bound on the scaling exponent: we show that $\mu \le 4.714$ for any BMSC and that $\mu \le 3.639$ for the BEC. Basically, this result improves by 1 the previous upper bound valid for any BMSC and it approaches closely the value 3.627 which was heuristically computed in [9] for the BEC. Secondly, we consider the joint scaling of error probability and gap to capacity: we describe a trade-off between the speed of decay of $P_{\rm e}$ and the speed of decay of $I(W) - R$ as functions of N. In the limit in which the gap to capacity is arbitrarily small but independent of N, this trade-off recovers the result of [7] for the error exponent regime. Thirdly, we prove that polar codes are not affected by error floors, i.e., that the slope of the error probability as a function of the Bhattacharyya parameter of the transmission channel increases in modulus with N. To do so, we fix 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 code over other channels W which are "better" that W' and we study the error probability $P_{\rm e}$ as a function of the Bhattacharyya parameter $Z(W)$. 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 $P_{\rm e}$ scales roughly as $Z(W)^{\sqrt{N}}$, in accordance with the error exponent regime. This suffices to conclude that polar codes are not affected by error floors, because the slope of $P_{\rm e}(Z(W))$ increases in modulus as N gets larger. The rest of the paper is organized as follows. After the review of some preliminary notions in Section II, 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 presenting some directions for future research. Due to space limitation, we only focus on the statement of our original results. All the proofs, together with a more detailed presentation, are provided in the longer version [15].

II. Preliminaries

Let W be a BMSC with capacity $I(W)$ and Bhattacharyya parameter $Z(W)$. The basis of channel polarization consists in mapping two identical copies of W into (W^0, W^1) s.t. W^0 is

[1] In [9], the scaling exponent is defined as the value of μ s.t. $\lim_{N\to\infty, N^{1/\mu}(C-R)=z} P_{\rm e}(N, R, C) = f(z)$ for some function $f(z)$. However, it is an open question to prove that such a limit exists.

a "worse" channel and W^1 is a "better" channel than W:

$$Z(W)\sqrt{2-Z(W)^2} \leq Z(W^0) \leq 2Z(W) - Z(W)^2, \quad (2)$$
$$Z(W^1) = Z(W)^2, \quad (3)$$

which follow from Proposition 5 of [6] and from Exercise 4.62 of [5]. In addition, when W is a BEC, we have that W^0 and W^1 are also BECs and $Z(W^0) = 2Z(W) - Z(W)^2$ [6]. By repeating n times this operation, we map 2^n identical copies of W into the synthetic channels $W_n^{(i)}$ ($i \in \{1, \cdots, 2^n\}$).

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

$$W_n = \begin{cases} W_{n-1}^0, & \text{w.p. } 1/2, \\ W_{n-1}^1, & \text{w.p. } 1/2. \end{cases} \quad (4)$$

Let $Z_n(W) = Z(W_n)$ be the random process that tracks the Bhattacharyya parameter of W_n. Then, from (2) and (3) we deduce that, for $n \geq 1$,

$$Z_n \begin{cases} \in \left[Z_{n-1}\sqrt{2-Z_{n-1}^2},\, 2Z_{n-1} - Z_{n-1}^2\right], & \text{w.p. } 1/2, \\ = Z_{n-1}^2, & \text{w.p. } 1/2. \end{cases} \quad (5)$$

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, for $n \geq 1$,

$$Z_n = \begin{cases} 2Z_{n-1} - Z_{n-1}^2, & \text{w.p. } 1/2, \\ Z_{n-1}^2, & \text{w.p. } 1/2. \end{cases} \quad (6)$$

Consider transmission over W of a polar code of block length $N = 2^n$ and rate R and let $P_{\rm e}$ denote the block error probability under SC decoding. Then, by Proposition 2 of [6],

$$P_{\rm e} \leq \sum_{i \in \mathcal{I}} Z_n^{(i)}, \quad (7)$$

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

First, we relate the value of the scaling exponent μ to the sup of some function. Then, 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.

Theorem 1 (From function to scaling exponent): Assume that there exists 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, for some $\mu > 2$,

$$\sup_{x\in(0,1),y\in[x\sqrt{2-x^2},2x-x^2]} \frac{h(x^2)+h(y)}{2h(x)} < 2^{-1/\mu}. \quad (8)$$

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

$$N \leq \frac{\beta_1}{(I(W)-R)^\mu}, \quad (9)$$

where β_1 is a universal constant which does not depend on W, but it depends only on $p_{\rm e}$. If W is a BEC, a less stringent hypothesis on μ is required for (9) to hold: the condition (8) is replaced by

$$\sup_{x\in(0,1)} \frac{h(x^2)+h(2x-x^2)}{2h(x)} < 2^{-1/\mu}. \quad (10)$$

Theorem 2 (Valid choice for scaling exponent): Consider transmission over a BMSC W with capacity $I(W)$ using a polar code of rate $R < I(W)$ and assume that the block error probability under SC decoding is *at most* $p_{\rm e} \in (0,1)$. Then, (9) holds with $\mu = 4.714$ and, if W is a BEC, it holds with $\mu = 3.639$.

The proof of these results is contained in Section III-B and III-C of the longer version [15]. Note that Theorem 1 provides an upper bound on the scaling exponent which depends on the choice of $h(x)$. The proof of Theorem 2 consists in the derivation and the analysis of a good choice of $h(x)$. In the remainder of this section, we focus on a heuristic interpretation of the function $h(x)$. First, let W be a BEC and consider the operator $T_{\rm BEC}$ defined as

$$T_{\rm BEC}(g) = \frac{g(z^2)+g(2z-z^2)}{2}, \quad (11)$$

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

$$\mathbb{E}\left[g(Z_n)|Z_0=z\right] = \overbrace{T_{\rm BEC}\circ T_{\rm BEC}\circ\cdots\circ T_{\rm BEC}(g)}^{n \text{ times}} = T_{\rm BEC}^n(g), \quad (12)$$

where the formula comes from a straightforward application of (6). A detailed explanation of the dynamics of the functions $T_{\rm BEC}^n(g)$ is provided in Section III of [10]. In short, a simple check shows that $\lambda = 1$ is an eigenvalue of the operator $T_{\rm BEC}$ with eigenfunctions $v_0(z) = 1$ and $v_1(z) = z$. Let λ^* be the largest eigenvalue of $T_{\rm BEC}$ other than $\lambda = 1$ and define μ^* as $\mu^* = -1/\log_2 \lambda^*$. Then, the heuristic discussion of [10] leads to the fact that μ^* is the largest candidate that we could plug in (10). For this choice, the function $h(x)$ represents the eigenfunction associated to the eigenvalue λ^*, namely,

$$\frac{h(x^2)+h(2x-x^2)}{2} = 2^{-1/\mu^*}h(x). \quad (13)$$

A numerical method for the calculation of this second eigenvalue was originally proposed in [9] and it yields $\mu^* = 3.627$. Furthermore, in Section III of [10] 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 $T_{\rm BMSC}$ defined as

$$T_{\rm BMSC}(g) = \sup_{y\in[x\sqrt{2-x^2},2x-x^2]} \frac{g(z^2)+g(y)}{2}. \quad (14)$$

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

$$\mathbb{E}\left[g(Z_n) \mid Z_0 = z\right] \leq T_{\rm BMSC}^n(g), \quad (15)$$

where the formula comes from a straightforward application of (5). Similarly to the case of the BEC, $\lambda = 1$ is an eigenvalue of $T_{\rm BMSC}$ 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 (8) and, for this choice, the function $h(x)$ represents the eigenfunction associated to the eigenvalue $2^{-1/\mu^*}$, namely,

$$\sup_{y \in [x\sqrt{2-x^2}, 2x-x^2]} \frac{h(x^2)+h(y)}{2} = 2^{-1/\mu^*} h(x). \qquad (16)$$

In Section IV of [10] it is proved that the scaling exponent μ is upper bounded by 6: this result is obtained by showing that the eigenvalue is at most $2^{-1/5}$, i.e., $\mu^* \le 5$ and, then, that μ^*+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. A more refined computation of μ^* is proposed in [11], which yields $\mu^* \le 4.702$, and, therefore, $\mu \le 5.702$. Theorem 1 solves the conjecture of [10] by proving that 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 $\mu \le 4.714$ for any BMSC and $\mu \le 3.639$ for the BEC.

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

In the scaling exponent regime, $P_{\rm e}$ is fixed and we study how fast $I(W)-R$ tends to 0 as a function of N. In Section III-A of [15], we also point out that $P_{\rm e}$ can go to 0 polynomially fast in N, without changing the scaling between $I(W)-R$ and N. The following theorem, whose proof is in Section IV-B of [15], shows that, by allowing a less favorable scaling between $I(W)-R$ and N, $P_{\rm e}$ goes to 0 sub-exponentially fast in N.

Theorem 3 (Joint scaling: exponential decay of $P_{\rm e}$): Assume that there exists a 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 $\gamma \in (1/(1+\mu), 1)$, the block length N and the block error probability under successive cancellation decoding $P_{\rm e}$ are s.t.

$$\begin{aligned} P_{\rm e} &\le N \cdot 2^{-N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}, \\ N &\le \frac{\beta_3}{(I(W)-R)^{\mu/(1-\gamma)}}, \end{aligned} \qquad (17)$$

where β_3 is a universal constant which does not depend on W or on γ, 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 (10) on μ is required for (17) to hold.

In short, formula (17) describes a trade-off between gap to capacity and error probability as functions of the block length N. Indeed, let γ go from $1/(1+\mu)$ 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 $\mu/(1-\gamma)$ is increasing in γ.

As concerns the possible choices for μ in (17), by constructing a function $h(x)$ as in the proof of Theorem 2, we immediately obtain the valid choices $\mu = 4.714$ for any BMSC and $\mu = 3.637$ for the special case of the BEC.

Note that by picking γ close to 1, we recover the result [7] concerning the error exponent regime: if the gap to capacity is fixed and independent of N, then $P_{\rm e}$ is $O(2^{-N^\beta})$ for any $\beta \in (0, 1/2)$[2]. On the other hand, it is not possible to recover from Theorem 3 the result of Theorem 1 concerning the scaling exponent regime. Indeed, pick γ close to $1/(1+\mu)$. Then, $\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)$ tends to 0, i.e., the error probability is fixed and independent of N, but N is $O\left(1/(I(W)-R)^{\mu+1}\right)$ instead of $O\left(1/(I(W)-R)^{\mu}\right)$ as in (9).

Finally, let us add the dependency on the Bhattacharyya parameter $Z(W)$ to this picture: under the hypotheses of Theorem 3, it is possible to prove that

$$\begin{aligned} P_{\rm e} &\le N \cdot Z(W)^{\frac{1}{2}\cdot N^{\gamma \cdot h_2^{(-1)}\left(\frac{\gamma(\mu+1)-1}{\gamma\mu}\right)}}, \\ N &\le \frac{\beta_4}{(I(W)-R)^{\mu/(1-\gamma)}}, \end{aligned} \qquad (18)$$

where β_4 is a constant which does not depend on W or on γ. In short, $P_{\rm e}$ scales as $Z(W)$ raised to some power of N, where the exponent follows the trade-off of Theorem 3. 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 (18) gives that $P_{\rm e}$ is roughly upper bounded by $Z(W)^{\sqrt{N}}$. On the other hand, $P_{\rm e} \ge \max_{i \in \mathcal{I}} Z_n^{(i)}$, where $Z_n^{(i)}$ is a polynomial in $Z(W)$ whose minimum degree scales roughly as $\sqrt{N}$ because of minimum distance considerations.

V. Absence of Error Floors

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 $P_{\rm e}$ and the Bhattacharyya parameter $Z(W)$. Note that when we have analyzed the dependency between $P_{\rm e}$ and $Z(W)$ at the end of the previous section, the polar code changed according to the channel. More specifically, Theorem 4 relates the Bhattacharyya parameter $Z_n^{(i)}(W)$ of the i-th synthetic channel obtained by polarizing W to the Bhattacharyya parameter $Z_n^{(i)}(W')$ obtained by polarizing W'. From this, Corollary 5 relates the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W, i.e., $\tilde{P}_{\rm e}(W) \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W)$, to the sum of Bhattacharyya parameters obtained by polarizing W', i.e., $\tilde{P}_{\rm e}(W') \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W')$. The indices of the information positions are the same in both sums, since the set $\mathcal{I}$ is fixed. The proof of Theorem 4 is in Section V-B of [15].

Theorem 4 (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

[2] Theorem 3 contains as a particular case also the result in [13], where the authors prove that N scales polynomially fast in $1/(I(W)-R)$ and, at the same time, $P_{\rm e}$ is upper bounded by $2^{-N^{0.49}}$.

obtained from W'. If $Z(W) \leq Z(W')^2$, then

$$Z_n^{(i)}(W) \leq Z_n^{(i)}(W')^{\frac{\log_2 Z(W)}{\log_2 Z(W')}}. \quad (19)$$

If W and W' are BECs, then (19) holds if $Z(W) \leq Z(W')$.

Corollary 5 (Scaling of $\tilde{P}_{\rm e}(W)$): 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 $\tilde{P}_{\rm e}(W')$ the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W', i.e., $\tilde{P}_{\rm e}(W') \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W')$, 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 $\tilde{P}_{\rm e}(W)$ be the sum of the Bhattacharyya parameters at the information positions obtained by polarizing W, i.e., $\tilde{P}_{\rm e}(W) \triangleq \sum_{i\in\mathcal{I}} Z_n^{(i)}(W)$. If $Z(W) \leq Z(W')^2$, then

$$\tilde{P}_{\rm e}(W) \leq \tilde{P}_{\rm e}(W')^{\frac{\log_2 Z(W)}{\log_2 Z(W')}}. \quad (20)$$

If W and W' are BECs, then (20) 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 $P_{\rm e}(W)$ the error probability under SC decoding for transmission of $\mathcal{C}$ over W and recall from (7) that $P_{\rm e}(W) \leq \tilde{P}_{\rm e}(W)$. Hence, formula (20) implies that

$$P_{\rm e}(W) \leq Z(W)^{\frac{\log_2 \tilde{P}_{\rm e}(W')}{\log_2 Z(W')}}. \quad (21)$$

Note that the upper bound (18) on $P_{\rm e}$ comes from an identical upper bound on the sum of the Bhattacharyya parameters $\tilde{P}_{\rm e}$. Thus, by taking $\gamma \approx 1$ in (18), we have that $\tilde{P}_{\rm e}(W')$ scales roughly as $Z(W')^{\sqrt{N}}$. Therefore, from (21) we conclude that $P_{\rm e}(W)$ scales roughly as $Z(W)^{\sqrt{N}}$. This fact excludes the existence of an error floor region, since it proves that the slope of the error probability increases in modulus with N. Furthermore, in the discussion at the end of Section IV, we pointed out that $P_{\rm e}(W)$ scales as $Z(W)^{\sqrt{N}}$ when W is fixed and, consequently, the polar code can be constructed according to the actual transmission channel. Since in the error floor regime the code cannot depend on the transmission channel, in terms of the scaling between $P_{\rm e}$ and $Z(W)$ 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 $Z(W) \leq Z(W')$ implies that $I(W) \geq I(W')$. If W and W' can be any BMSC, one can prove that $Z(W) \leq Z(W')^2$ implies that $I(W) \geq I(W')$. Recall that the rate of a polar code for W' is s.t. $R < I(W')$, and the rate of a polar code for W is s.t. $R < I(W)$. As $I(W) \geq I(W')$, by constructing a polar code for W, we can transmit reliably at larger rates.

VI. Concluding Remarks

Let us summarize the main results contained in this work, along with directions for future research. First of all, we prove a new upper bound on the scaling exponent for any BMSC W. Possibly the most interesting open question concerning the performance of polar codes consists in improving such a scaling exponent by changing the construction of the code and by devising better decoding algorithms. One promising method interpolates between a polar and a Reed-Muller code and employs a successive cancellation list decoder [16]. Secondly, we prove a trade-off between the speed of decay of the error probability and that of the gap to capacity, which recovers the existing result for the error exponent regime. Thirdly, we prove that polar codes are not affected by error floors. However, when W and W' are any BMSC, the results (19) and (20) hold only if $Z(W) \leq Z(W')^2$. It remains an open question whether similar but perhaps less tight bounds still hold for $Z(W) \in (Z(W')^2, Z(W')]$. Finally, let us point out that the techniques described in this paper can be useful in the analysis of polar codes with general $\ell \times \ell$ kernels [17].

Acknowledgment

This work was supported by grant No. 200020_146832/1 of the Swiss National Science Foundation. S. Hamed Hassani is supported by ERC Starting Grant under grant number 307036.

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

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

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

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

[6] 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.

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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.

[13] V. Guruswami and P. Xia, "Polar codes: speed of polarization and polynomial gap to capacity," *IEEE Trans. Inform. Theory*, vol. 61, no. 1, pp. 3–16, Jan. 2015.

[14] 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.

[15] M. Mondelli, S. H. Hassani, and R. Urbanke, "Unified scaling of polar codes: error exponent, scaling exponent, moderate deviations, and error floors," Jan. 2015, [Online]. Available: http://arxiv.org/pdf/1501.02444.pdf.

[16] ——, "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.

[17] 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.