Finite-Length Scaling for Polar Codes

S. Hamed Hassani, Kasra Alishahi, and Rüdiger Urbanke

Abstract—Consider a binary-input memoryless output-symmetric channel \( W \). Such a channel has a capacity, call it \( I(W) \), and for any \( R < I(W) \) and strictly positive constant \( P_e \) we know that we can construct a coding scheme that allows transmission at rate \( R \) with an error probability not exceeding \( P_e \). Assume now that we let the rate \( R \) tend to \( I(W) \) and we ask how we have to “scale” the blocklength \( N \) in order to keep the error probability fixed to \( P_e \). We refer to this as the “finite-length scaling” behavior. This question was addressed by Strassen as well as Polyanisky, Poor and Verdu, and the result is that \( N \) must grow at least as the square of the reciprocal of \( I(W) - R \).

Polar codes are optimal in the sense that they achieve capacity. This however is a formidable task. It is slightly easier to fix one of the parameters and then to describe the relationship (scaling) of the remaining two.

For example, assume that we fix the rate and consider the relationship between the error probability and the block-length. This is the study of the classical error exponent. For instance, for random codes a closer look shows that \( P_e = e^{-N E(R; W) + o(N)} \), where \( E(R; W) \) is the so-called random coding error exponent \( 2 \) of the channel \( W \). For polar codes, Arikan and Telatar \( 3 \) showed that when \( W \) is a BMS channel, for any fixed rate \( R < I(W) \) the block error probability of polar codes with the successive cancellation (SC) decoder is upper bounded by \( 2^{-N^{\beta}} \) for any \( \beta < \frac{1}{2} \) and \( N \) large enough.

This result was refined later in \( 4 \) to be dependent on \( R \), i.e. for polar codes with the SC decoder

\[
P_e = 2^{-2^{\frac{1}{2\sqrt{\log N}}} \left( (I(W) - R) + o(\sqrt{\log N}) \right)}
\]

Another option is to fix the error probability and to consider the relationship between the block-length and the rate. In other words, given a code and a desired (fixed) error probability \( P_e \), what is the block-length \( N \) required, in terms of the rate \( R \), so that the code has error probability less than \( P_e \)? This scaling is arguably more relevant (than the error exponent) from a practical point of view since we typically have a certain requirement on the error probability and then are interested in using the shortest code possible to transmit at a certain rate.

As a benchmark, let us mention what is the shortest block-length that we can hope for. Some thought clarifies that the random variations of the channel itself require \( R \leq I(W) - \Theta(\sqrt{N}) \) or equivalently \( N \geq \Theta(\frac{1}{(I(W) - R)^2}) \). Indeed, a sequence of works starting from \( 5 \), then \( 6 \), and finally \( 7 \) showed that the minimum possible block-length \( N \) required to achieve a rate \( R \) with a fixed error probability \( P_e \) is roughly equal to

\[
N \approx \frac{V(Q^{-1}(P_e))^2}{(I(W) - R)^2},
\]

where \( V \) is a characteristic of the channel referred to as channel dispersion. In other words, the best codes require a block-length of order \( \Theta(\frac{1}{(I(W) - R)^2}) \).

The main objective of this paper is to characterize similar types of relations for polar codes with the SC decoder. We argue in this paper that this problem is fundamentally related to the dynamics of channel polarization and especially the speed of which the polarization phenomenon is taking place. We then provide analytical bounds on the speed of polarization for BMS channels. Finally, by using these bounds we derive scaling laws between the block-length and the rate (given a fixed error probability) that hold universally for all BMS

I. INTRODUCTION

Polar coding schemes \( 1 \) provably achieve the capacity of a wide class of channels including binary-input memoryless output-symmetric (BMS) channels.

In coding, the three most important parameters are: rate \( (R) \), block-length \( (N) \), and block error probability \( (P_e) \). Ideally, given a family of codes such as the family of polar codes, one would like to be able to describe the exact relationship between these three parameters. This however is a formidable

This work was supported in part by grant No. 200020_146832/1 and grant No. 200021_121903 of the Swiss National Science Foundation. This paper was presented in part in \( 12 \), \( 13 \). S. H. Hassani is with the Department of Computer Science, ETHZ, 8092 Zurich, Switzerland (e-mail: hamed@inf.ethz.ch).

K. Alishahi is with the department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran (email: alishahi@sharif.edu).

R. Urbanke is with the School of Computer and Communication Science, EPFL, 1015 Lausanne, Switzerland (e-mail: rudiger.urbanke@epfl.ch).

Footnote 1: In this paper all the logarithms are in base 2.
channels. To state things in a more convenient language, let us begin by reviewing some conventional definitions, settings, and results regarding polarization and polar codes.

A. Preliminaries

Let \( W : \mathcal{X} \to \mathcal{Y} \) be a BMS channel, with input alphabet \( \mathcal{X} = \{0, 1\} \), output alphabet \( \mathcal{Y} \), and the transition probabilities \( \{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\} \). We consider the following three parameters for the channel \( W \):

\[
H(W) = \sum_{y \in \mathcal{Y}} W(y | 1) \log \frac{W(y | 1)}{W(y | 1)},
\]

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

\[
E(W) = \sum_{y \in \mathcal{Y}} W(y | 1) \left( \mathbb{1}_{\{W(y | 0) > W(y | 1)\}} + \frac{1}{2} \mathbb{1}_{\{W(y | 0) = W(y | 1)\}} \right),
\]

where \( \mathbb{1}_{\{A\}} \) is equal to 1 if \( A \) is true and 0 otherwise. The parameter \( H(W) \) is equal to the entropy of the input of \( W \) given its output when we assume uniform distribution on the inputs, i.e., \( H(W) = H(X | Y) \). Hence, we call the parameter \( H(W) \) the entropy of the channel \( W \). Also note that the capacity of \( W \), which we denote by \( I(W) \), is given by \( I(W) = 1 - H(W) \). The parameter \( Z(W) \) is called the Bhattacharyya parameter of \( W \) and \( E(W) \) is called the error probability of \( W \). It can be shown that \( E(W) \) is equal to the error probability in estimating the channel input \( x \) on the basis of the channel output \( y \) via the maximum-likelihood decoding of \( W(y|x) \) (with the further assumption that the input has uniform distribution). The following relations hold between these parameters (see for e.g., [1] and [14] Chapter 4):

\[
0 \leq 2E(W) \leq H(W) \leq Z(W) \leq 1,
\]

\[
H(W) \leq h_2(E(W)),
\]

\[
Z(W) \leq \sqrt{1 - (1 - H(W))^2},
\]

\[
2E(W) \geq 1 - \sqrt{1 - Z(W)^2},
\]

where \( h_2(\cdot) \) denotes the binary entropy function, i.e.,

\[
h_2(x) = -x \log(x) - (1 - x) \log(1 - x).
\]

B. Channel Transform

Let \( W \) denote the set of all BMS channels and consider a transform \( W \to (W^0, W^1) \) that maps \( W \) to \( \mathcal{W}^2 \) in the following manner. Having the channel \( W : \{0, 1\} \to \mathcal{Y} \), the channels \( W^0 : \{0, 1\} \to \mathcal{Y}^2 \) and \( W^1 : \{0, 1\} \to \{(0, 1) \times \mathcal{Y}^2 \} \) are defined as

\[
W^0(y_1, y_2 | x_1) = \sum_{x_2 \in \{0, 1\}} \frac{1}{2} W(y_1 | x_1 \oplus x_2) W(y_2 | x_2)
\]

\[
W^1(y_1, y_2, x_1 | x_2) = \frac{1}{2} W(y_1 | x_1 \oplus x_2) W(y_2 | x_2),
\]

The transform \( W \to (W^0, W^1) \) is also known as the channel splitting transform. A direct consequence of the chain rule of entropy yields

\[
\frac{H(W^0) + H(W^1)}{2} = H(W).
\]

Regarding the other parameters, we have (see [1] and [14] Chapter 4):

\[
Z(W) \sqrt{2 - Z(W)^2} \leq Z(W^0) \leq 1 - (1 - Z(W))^2;
\]

\[
Z(W^1) = Z(W)^2;
\]

and (see [14] Chapter 4):

\[
E(W^0) = 1 - (1 - E(W))^2;
\]

\[
E(W^2) \leq E(W^1) \leq E(W).
\]

C. Channel Polarization

Consider an infinite binary tree with the root node placed at the top. In this tree each vertex has 2 children and there are \( 2^n \) vertices at level \( n \). Assume that we label these vertices from left to right from 0 to \( 2^n - 1 \). Here, we intend to assign to each vertex of the tree a BMS channel. We do this by a recursive procedure. Assign to the root node the channel \( W \) itself. Now consider the channel splitting transform \( W \to (W^0, W^1) \) and from left to right, assign \( W^0 \) and \( W^1 \) to the children of the root node. In general, if \( Q \) is the channel that is assigned to vertex \( v \), we assign \( Q^0 \) to the “left” child of \( v \) and \( Q^1 \) to the “right” child of \( v \). In this way, we recursively assign a channel to all the vertices of the tree. Figure 1 shows the first 2 levels of the binary tree. Assuming \( N = 2^n \), we let \( W_N^{(i)} \) denote the channel that is assigned to a vertex with label \( i \) at level \( n \) of the tree, \( 0 \leq i \leq N - 1 \). As a result, one can equivalently relate the channel \( W_N^{(i)} \) to \( W \) via the following procedure: let the
binary representation of $i$ be $b_1 b_2 \ldots b_n$, where $b_1$ is the most significant digit. Then we have

$$W^{(i)}_N = \left( ((W^{b_1})^{b_2})^{\ldots}^{b_n} \right).$$

As an example, assuming $i = 6$, $n = 3$ we have $W^{(6)}_3 = ((W^1)^1)^0$. We now proceed with defining a stochastic process called the polarization process. This process can be considered as a stochastic representation of the channels associated to different levels of the infinite binary tree.

### D. Polarization Process

Let $\{B_n, n \geq 1\}$ be a sequence of independent and identically distributed (iid) Bernoulli($\frac{1}{2}$) random variables. Denote by $(\mathcal{F}, \Omega, \mathbb{P}_1)$ the probability space generated by this sequence and let $(\mathcal{F}_n, \Omega_n, \mathbb{P}_{BEC})$ be the probability space generated by $(B_1, \ldots, B_n)$. For a BMS channel $W$, define a random sequence of channels $W_n$, $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, as $W_0 = W$ and

$$W_n = \begin{cases} W_{n-1}^0 & \text{if } B_n = 0, \\ W_{n-1}^1 & \text{if } B_n = 1, \end{cases}$$

(17)

where the random processes $\{W_n\}_{n \geq 1}$, $\{I_n\}_{n \geq 1}$, $\{Z_n\}_{n \geq 1}$ and $\{E_n\}_{n \geq 1}$ as $H_n = H(W_n)$, $I_n = I(W_n) = 1 - H(W_n)$, $Z_n = Z(W_n)$ and $E_n = E(W_n)$.

**Example 1:** By a straightforward calculation one can show that for $W = \text{BEC}(z)$ we have

$$W^0 = \text{BEC}(1-(1-z)^2) \quad \text{(18)}$$

$$W^1 = \text{BEC}(z^2). \quad \text{(19)}$$

Hence, when $W = \text{BEC}(z)$, the channel $W_n$ is always a BEC. Furthermore, the processes $H_n, I_n, Z_n$ and $E_n$ admit simple closed form recursions as follows. We have $H_0 = z$ and for $n \geq 1$

$$H_n = \begin{cases} 1 - (1 - H_{n-1})^2, & \text{w.p. } \frac{1}{2}, \\ H_{n-1}^2, & \text{w.p. } \frac{1}{2}. \end{cases}$$

(20)

Also, we have

$$2E_n = H_n = 1 - I_n = Z_n.$$

For channels other than the BEC, the channel $W_n$ gets quite complicated in the sense that the cardinality of the output alphabet of the channel $W_n$ is doubly exponential in $n$ (or exponential in $N$). Thus, tracking the exact outcome of $W_n$ seems to be a difficult task (for more details see [16], [17]). Instead, as we will see in the sequel, one can prove many interesting properties regarding the processes $H_n, Z_n$ and $E_n$.

Let us quickly review the limiting properties of the above mentioned processes [1], [3]. From (12) and (17), one can write for $n \geq 1$

$$\mathbb{E}[H(W_n) | W_{n-1}] = \frac{H(W^0_{n-1}) + H(W^1_{n-1})}{2} = H(W_{n-1}).$$

(21)

Hence, the process $H_n$ is a martingale. Furthermore, since $H_n$ is also bounded (see [3]), by Doob’s martingale convergence theorem, the process $H_n$ converges almost surely to a limit random variable $H_\infty$. As $H_n$ is also bounded, we have for $n \to \infty$

$$\mathbb{E}[H_n - H_{n-1}] = \mathbb{E}[H(W_n^0) - H(W_n^1)] \to 0.$$

As a result, we must have that $H(W_n^0) - H(W_n^1)$ converges to 0 almost surely (a.s.). We will shortly prove that for a channel $P$, in order to have $H(P^0) = H(P)$ we must either have $H(P) \approx 0$ (i.e., $P$ is the noiseless channel) or $H(P) \approx 1$ (i.e., $P$ is the completely noisy channel). By this claim and the fact that $H_n$ converges a.s. to $H_\infty$, we conclude that $H_\infty$ takes its values in the set $\{0, 1\}$. Also, as $\mathbb{E}[H_n] = \mathbb{E}[H_\infty] = H(W)$, we obtain

$$H_\infty = \begin{cases} 0 & \text{w.p. } 1 - H(W), \\ 1 & \text{w.p. } H(W). \end{cases}$$

(22)

It remains to prove the claimed mentioned above. It is sufficient to show that for a channel $P$, in order to have $H(P^0) = H(P)$ we must have $H(P) \in \{0, 1\}$. We use the so called extremes of information combining inequalities [14, Theorem 4.141]: Let $P$ be an arbitrary BMS channel. To simplify notation, let $h \equiv H(P)$ and also let $\epsilon \in [0, \frac{1}{2}]$ be such that $h_2(\epsilon) = h$ (in this way, the two channels BEC$(h)$ and BSC$(\epsilon)$ have the same capacity). We have

$$h \leq H(\text{BSC}(\epsilon)^0) \leq H(\epsilon) \leq H(\text{BEC}(\epsilon)^0).$$

(23)

$$\frac{1 - (1 - h)_2}{h^2 - 2h_2(2(1 - h))} \leq H(\text{BEC}(\epsilon)^1) \leq H(1 - \epsilon) \leq h.$$\(\frac{1 - (1 - h)_2}{h^2 - 2h_2(2(1 - h))}\)

(24)

Now, to prove the claim, assume that $P$ is such that $H(P^0) = H(P) = h$. Using (23) we obtain $H(\text{BSC}(\epsilon)^0) = H(P)$ or equivalently $h_2(2(1 - \epsilon)) = h = h_2(\epsilon)$. As a result, $\epsilon$ must be a solution of the equation $\epsilon = 2(1 - \epsilon)$ which yields $\epsilon \in [0, \frac{1}{2}]$. Also, as $H(P) = h_2(\epsilon)$, then $H(P)$ can either be 0 or 1 and hence the claim is justified. Using the bounds (23)–(24) it is clear that the processes $Z_n$ and $E_n$ converge a.s. to $H_\infty$ and $\frac{1}{2}H_\infty$, respectively.

### E. Polar Codes

Given the rate $R < I(W)$, polar coding is based on selecting a set of $2^n R$ rows of the matrix $G_n = \left[ \begin{array}{c} 1 \\ 0 \end{array} \right]^{\otimes n}$ to form a $2^n R \times 2^n$ matrix which is used as the generator matrix in the encoding procedure. The way this set is selected is dependent on the channel $W$ and is briefly explained as follows: Order the the set of channels $\{W_n^{(i)}\}_{0 \leq i < N - 1}$ according to their error probability (given in [4]). Then, pick the $N$ $R$ channels which have the smallest error probability and consider the rows of $G_n$ with the same indices as these channels. E.g., if the channel $W_N^{(i)}$ is chosen, then the $i$-th row of $G_n$ is selected. In the

---

1Here, we are skipping some unnecessary details. For the sake of completeness, we note that the function $H_i()$ is a continuous function over the space of BMS channels. For more details, we refer to [14, Chapter 4].

2One can also construct polar codes by choosing the channels that have the least Bhattacharyya parameter or the least entropy (see 2 and 3). In essence, these constructions are all equivalent except that a few indices might be different. Choosing the channels that have the least error probability has the advantage of minimizing the “union”-type bounds that can be provided on the block-error probability when we use SC decoding (see e.g. the right side of [25]).
following, given $N$, we call the set of indices of $N \cdot R$ channels with the least error probability the set of good indices and denote it by $\mathcal{I}_{N,R}$. Moreover, we will frequently use the terms “the set of good indices” and $\mathcal{I}_{N,R}$ interchangeably.

We now briefly explain why such a code construction is reliable for any rate $R < I(W)$, provided that the block-length is large enough. It is proven in [1] that the block error probability of such polar coding scheme under SC decoding, denoted by $P_e$, is bounded from both sides by

$$\max_{i \in \mathcal{I}_{N,R}} E(W_N^{(i)}) \leq P_e \leq \sum_{i \in \mathcal{I}_{N,R}} E(W_N^{(i)}). \quad (25)$$

Recall from Subsection 1-D that the process $E_n = E(W_n)$ converges a.s. to a random variable $E_\infty$ such that $\Pr(E_\infty = 0) = I(W)$. Hence, it is clear from the definition of the set of good indices, $\mathcal{I}_{N,R}$, that the left side of (25) decays to 0 for any $R < I(W)$ as $n$ grows large. However, the story is not over yet as this is only a lower bound on $P_e$. Nonetheless, one can also show that the right side of (25) decays to 0. This was initially shown in [1], and later in [3] it was proven that all of the three terms in (25) behave like $2^{-z^2}$.

II. Problem Formulation

As we have seen in the previous section, the processes $H_n$ and $Z_n$ polarize in the sense that they converge a.s. to $\{0, 1\}$-valued random variables $H_\infty$ and $Z_\infty$, respectively. In other words, almost surely as $n$ grows, the value of $Z_n$ (or $H_n$) is either very close to 0 or very close to 1. Here, we investigate the dynamics of polarization. We start by noting that at each time $n$ there still exists a (small and in $n$ vanishing) probability that the process $Z_n$ (or $H_n$) takes a value far away from the endpoints of the unit interval (i.e., 0 and 1). Our primary objective is to study these small probabilities. More concretely, let $0 < a < b < 1$ be constants and consider the quantity $\Pr(Z_n \in [a, b])$. This quantity represents the fraction of sub-channels that are still un-polarized at time $n$.

An important question is how fast (in terms of $n$) the quantity $\Pr(Z_n \in [a, b])$ decays to zero. This question is intimately related to measuring the limiting properties of the sequence $\{\frac{1}{n} \log \Pr(Z_n \in [a, b])\}_{n \in \mathbb{N}}$.

**Example 2:** Assume $W$ is BEC($z$). In this case the process $Z_n$ has a simple closed form recursion as $Z_0 = z$ and

$$Z_{n+1} = \begin{cases} Z_n^2, & \text{w.p. } \frac{1}{2}, \\ 1 - (1 - Z_n)^2, & \text{w.p. } \frac{1}{2}. \end{cases} \quad (26)$$

Hence, it is straightforward to compute the value $\Pr(Z_n \in [a, b])$ numerically. Let $a = 1 - b = 0.1$. Figure 2 shows the value $\frac{1}{n} \log \Pr(Z_n \in [a, b])$ in terms of $n$ for $z = 0.5, 0.6, 0.7$. This figure suggests that the sequence $\{\frac{1}{n} \log \Pr(Z_n \in [a, b])\}_{n \in \mathbb{N}}$ converges to a limiting value that is somewhere between $-0.27$ and $-0.28$. Note that for different values of $z$, the limiting values are very close to each other.

For other BMS channels, the process $Z_n$ does not have a simple closed form recursion as for the BEC, and hence we need to use approximation methods (for more details see [16], [17]). Using such methods, we have plotted in Figure 3 the value of $\Pr(Z_n \in [a, b])$ for the channel families BSC($\epsilon$) and BAWGN($\sigma$) with different parameter values.

The above numerical evidence suggests that the quantity $\Pr(Z_n \in [a, b])$ decays to zero exponentially fast in $n$. Further, we observe that the limiting value of this sequence is dependent on the starting channel $W$ (e.g., from the figures it is clear that the channels BEC, BSC and BAWGN have different limiting values). Let us now be concrete and rephrase the above speculations as follows.

**Question 1:** Does the quantity $\Pr(Z_n \in [a, b])$ decay exponentially in $n$? If yes, what is the limiting value of $\frac{1}{n} \log \Pr(Z_n \in [a, b])$ and how is this limit related to the starting channel $W$ and the choice of $a$ and $b$?

From Figures 2 and 3, we observe that the value of $\frac{1}{n} \log \Pr(Z_n \in [a, b])$ is the least when $W$ is a BEC and this suggests that the channel BEC polarizes faster than the other BMS channels. This is intuitively justified as follows: Fix a value $z \in (0, 1)$ and assume that $W$ is a BMS channel with Bhattacharyya parameter $Z(W) = z$. Now, consider the values $Z(W^0)$ and $Z(W^1)$. Using relations (13) and (14), it is clear that the values $Z(W^0)$ and $Z(W^1)$ are closest to the end points of the unit interval if $W$ is a BEC. In other words, at the channel splitting transform, the channel BEC($z$) polarizes faster than the other BMS channels.

\footnotetext{Note here that by \[ the error probability of a BMS channel is less than its Bhattacharyya value. Hence, the right side of (25) is a better upper bound for the block error probability than the sum of Bhattacharyya values.}
Question 2: For which set of channels does the quantity $\Pr(Z_n \in [a, b])$ decay the fastest or the slowest?

Let us now be more ambitious and aim for the ultimate goal. Question 3: Can we characterize the exact behavior of $\Pr(Z_n \in [a, b])$ as a function of $n$, $a$, $b$ and $W$?

Finally, we ask how the answers to the above questions will guide us through the understanding of the finite-length scaling behavior of polar codes. An immediate relation stems from the fact that the quantity $\Pr(Z_n \in [a, b])$ indicates the portion of the sub-channels that have not polarized at time $n$. In particular, all the channels in this set have a large Bhattacharyya value (and hence a large error probability). Consequently, if any of such un-polarized channels (or equivalently indices) are included in the set of good indices then the error probability would not be small (see (25)). Thus, the maximum reliable rate that we can achieve is restricted by the portion of these yet un-polarized channels. The answers to Questions 1 and 2 posed above will therefore be crucial in finding answers to the following question.

Question 4: Fix the channel $W$ and a target block error probability $P_e$. To have a polar code with error probability less than $P_e$, how does the required block-length $N$ scale with the rate $R$?

Finding a suitable answer to the above questions is an easier task when the channel $W$ is a BEC. This is due to the simple closed form expression of the process $Z_n$ given in (26). In the next section (Section III), we provide heuristic methods that lead to suitable numerical answers to Questions 1 and 2 for the BEC. As we will see in the next section, such heuristic derivations are in excellent compliance with numerical experiments. Using such derivations, we also give an answer to Question 3 for the BEC.

The heuristic results of Section III provide us then with a concrete path to analytically tackle the above questions. In Section IV we provide analytical answers to Questions 1 and 4 for the BEC as well as other BMS channels. Providing a complete answer to Questions 1 and 4 is beyond what we achieve in Section IV; nevertheless, we provide close and useful bounds. Finally, in Section V we conclude the paper.

III. HEURISTIC DERIVATION FOR THE BEC

In this section we provide a heuristic (and numerical) procedure that leads to a clear picture of how the process $Z_n$ evolves through time $n$ when the channel $W$ is a BEC. As we will see, this procedure guides us to a number of conclusions about the process $Z_n$, which we refer to as assumptions. By using these assumptions we can (numerically) compute the important parameters for the process $Z_n$ which will then enable us to predict scaling laws for the evolution of $Z_n$ as well as scaling laws for polar codes. Several plots are provided to show the excellent compliance of these scaling predictions with reality. The intuitive discussions as well as the numerical observations of this section will then help us in building a rigorous framework for the analysis of the evolution of $Z_n$. This is the subject of the next section (Section IV). Let us emphasize that none of heuristic assumptions of the current section (Section III) will be used in any of the proofs of the next section.

Throughout this section we assume that the channel $W$ is the BEC(z) where $z \in [0, 1]$. To avoid cumbersome notation, let us define $^{10}$

$$p_n(z, a, b) = \Pr(Z_n \in [a, b] \mid Z_0 = z),$$

(27)

where the condition $Z_0 = z$ means that $Z_0$ is the Bhattacharyya process of the BEC(z). We start by noticing that by (26) the function $p_n(z, a, b)$ satisfies the following recursion

$$p_{n+1}(z, a, b) = \frac{p_n(z^2, a, b) + p_n(1-(1-z)^2, a, b)}{2},$$

(28)

with

$$p_0(z, a, b) = \mathbb{1}_{[z \in [a, b]]}.$$  (29)

More generally, one can easily observe the following. Let $g : [0, 1] \rightarrow \mathbb{R}$ be an arbitrary bounded function. Define the functions $\{g_n\}_{n \in \mathbb{N}}$, $g_n : [0, 1] \rightarrow \mathbb{R}$, as

$$g_n(z) = \mathbb{E}[g(Z_n) \mid Z_0 = z].$$

(30)

The functions $\{g_n\}_{n \in \mathbb{N}}$ satisfy the following recursion for $n \in \mathbb{N}$

$$g_{n+1}(z) = \frac{g_n(z^2) + g_n(1-(1-z)^2)}{2}.$$  (31)

This observation motivates us to define the polar operator, denoted by $T$, as follows. Let $B$ be the space of all bounded and real valued functions $g$ over $[0, 1]$. The polar operator $T : B \rightarrow B$ maps a function $g \in B$ to another function in $B$ in the following way

$$T(g) = \frac{g(z^2) + g(1-(1-z)^2)}{2}.$$  (32)

It is now clear that

$$\mathbb{E}[g(Z_n) \mid Z_0 = z] = \overbrace{T \circ T \circ \cdots \circ T(g)}^{n \text{ times}} \equiv T^n(g).$$  (33)

In this new setting, our objective is to study the limiting behavior as well as the dynamics of the functions $T^n(g)$ when $g$ is a simple function as in (29). This task is intimately related to studying the eigenvalues of the polar operator $T$ and their corresponding eigenfunctions. Also, a check shows that both of the functions

$$v_0(z) = 1, v_1(z) = z,$$  (34)

are eigenfunctions associated to the eigenvalue $\lambda = 1$.

Consider now a function $g \in B$. For simplicity, let us also assume that $g$ is continuous at $z = 0$ and $z = 1$. By using the fact that $Z_n$ polarizes, it is easy to see that

$$\mathbb{E}[g(Z_n) \mid Z_0 = z] \xrightarrow{n \to \infty} (1-z)g(0) + zg(1).$$

Equivalently by (33) we have

$$T^n(g) \xrightarrow{n \to \infty} g(0) - z(g(1) - g(0)).$$  (35)

In other words, $T^n(g)$ converges to a linear combination of the two eigenfunctions $v_0(z) = 1$ and $v_1(z) = z$ that are associated to the eigenvalue $\lambda = 1$. However, our main interest is to find out how fast the convergence in (35) is taking place.
in terms of \( n \). In this regard, to keep things simple and in a more manageable setting, let us consider finite-dimensional approximations of \( T \). This is done by discretizing the unit interval into very small sub-intervals with the same length and by assuming that \( T \) operates on all the points of each sub-interval in the same way. More concretely, consider a (large) number \( L \in \mathbb{N} \) and let the numbers \( x_i, i \in \{0,1,\cdots,L−1\} \), be defined as \( x_i = \frac{i}{L} \). Hence, the unit interval \([0,1]\) can be thought of as the union of the small sub-intervals \([x_i,x_{i+1}]\). Now, for simplicity assume that \( g \) is a (piece-wise) continuous function on \([0,1]\). Intuitively, by assuming \( L \) to be large, we expect that the value of \( g \) is the same throughout each of the intervals \([x_i,x_{i+1}]\). Such an assumption seems also reasonable for the function \( T(g) \) given in \( (32) \). Thus, we can approximate the function \( g \) as an \( L \) dimensional vector

\[
g_L \approx [g(x_0),g(x_1),\cdots,g(x_{L−1})].
\]  

(36)

In this way, from \( (32) \) we expect that the function \( T(g) \) can be well approximated by a matrix multiplication

\[
T(g) \approx g_LT_L,
\]  

(37)

where the \( L \times L \) matrix \( T_L \) is defined as follows. Let \( T_L(i,j) \) be an element of \( T_L \) in the \( i \)-th row and the \( j \)-th column. Define \( T_L(1,1) = T_L(L,L) = 1 \) and for the other elements \( i,j \in \{0,1,\cdots,L−1\} \) we let

\[
T_L(i+1,j+1) = \begin{cases} \frac{1}{L^2}, & \text{if } i = [(L−1)(\frac{j}{L^2})^2], \\ \frac{1}{L^2}, & \text{if } i = [(L−1)(1−(1−\frac{j}{L^2})^2)], \\ 0, & \text{o.w.} \end{cases}
\]  

(38)

As an example, the matrix \( T_L \) for \( L = 10 \) has the following form

\[
T_{10} = \begin{pmatrix} 1 & \frac{1}{10} & \frac{1}{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{1}{10} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{10} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{10} \end{pmatrix}
\]  

All the columns of \( T_L \) sum up to 1. Hence, an application of the Perron–Frobenius theorem [15 Chapter 8] shows that the eigenvalues of \( T_L \) are all inside the interval \([-1,+1]\). Also, a check shows that the matrix \( T_L \) has an eigenvalue equal to \( \lambda_0 = 1 \) with two corresponding (left) eigenvectors

\[
v_{0,L} = (1,1,\cdots,1), \\
v_{1,L} = (x_0,x_1,\cdots,x_{L−1}),
\]

where \( x_i = \frac{i}{L−1} \). By using \( (36) \), it is easy to see that the vectors \( v_{0,L} \) and \( v_{1,L} \) are the corresponding \( L \)-dimensional approximations of the eigenfunctions \( v_0 \) and \( v_1 \) given in \( (34) \). We thus expect

\[
g_LT^n_L \xrightarrow{n \to \infty} c_0v_{0,L} + c_1v_{1,L},
\]  

(39)

where \( c_0 \) and \( c_1 \) are constants. Moreover, from \( (35) \) we have

\[
c_0 \xrightarrow{L \to \infty} g(0), \\
c_1 \xrightarrow{L \to \infty} g(1)−g(0).
\]  

In order to find out how fast the convergence in \( (39) \) is, we look at the second and third largest eigenvalues (in absolute value) of \( T_L \) as \( L \) grows large. We denote the second largest eigenvalue of \( T_L \) by \( \lambda_2(L) \), and the third largest eigenvalue is denoted by \( \lambda_3(L) \). Table \( \text{II} \) contains the value of these eigenvalues computed numerically for several (large) values of \( L \). It can thus be conjectured that

\[
\lim_{L \to \infty} \lambda_2(L) \approx 0.826, \\
\lim_{L \to \infty} \lambda_3(L) \approx 0.705.
\]  

(40)

(41)

This belief guides us to conclude that, for \( L \) growing large, if we start from any vector \( g_L \), then

\[
g_LT^n_L \approx c_0v_{0,L} + c_1v_{1,L} + c_2\lambda_2^nLv_2 + O(n\lambda_3^n).
\]  

The above approximate relation indicates that for large \( L \), the distance of \( g_LT^n_L \) from its value in the limit is roughly equal to \( c_2\lambda_2^nL \).

One particular instance of the function \( g \), is the one given in \( (29) \), i.e., \( g(z) = \mathbb{I}_{\{x \in [a,b]\}} \). If \( a,b \in (0,1) \) we know that \( T^n(g) = \Pr(Z_n \in [a,b]) \) converges to 0 everywhere (see \( (28) \)). If we consider the \( L \)-dimensional approximations of \( g \) and \( T \) for \( L \) large, then the final limit of \( g_LT^n_L \) would be arbitrarily close to 0 (depending on how large \( L \) is). Also, by \( (42) \) the distance to this final limit is around \( \lambda_2^nL \approx 2^{-n\log\frac{1}{\lambda_2}} \). In words, the speed of this convergence is large for \( L \). Now, let us go back the original polar operator \( T \) defined in \( (32) \). As we argued above, the operators \( T_L \), for \( L \) large, are good finite-dimensional approximations of \( T \). The (experimental) relation \( (42) \) brings us to the following assumption about \( T \).

**Assumption I (Scaling Assumption):** There exists \( \mu \in (0,\infty) \) such that, for any \( z,a,b \in (0,1) \) such that \( a < b \), the limit \( \lim_{n \to \infty} 2^n\Pr(Z_n \in [a,b]) \) exists in \((0,\infty)\). We denote this limit by \( q(z,a,b) \). In other words,

\[
\lim_{n \to \infty} 2^n\Pr(Z_n \in [a,b]) = q(z,a,b).
\]  

(43)

We call the value \( \mu \) the scaling exponent of polar codes for the BEC.

By \( (43) \) the value of \( \Pr(Z_n \in [a,b]) \) converges to 0 like \( 2^{-\frac{1}{\mu}} \). Hence, the speed of polarization for the process \( Z_n \) over the BEC is equal to \( \frac{1}{\mu} \).

Note here that by \( (40) \) we expect that

\[
2^{-\frac{1}{\mu}} = \lim_{L \to \infty} \lambda_2(L) \approx 0.826 = \frac{1}{\mu} \approx 0.275.
\]  

(44)

| \( L \) | 1000 | 2000 | 4000 | 8000 |
|-----|------|------|------|------|
| \( \lambda_2(L) \) | 0.8227 | 0.8240 | 0.8248 | 0.8253 |
| \( \lambda_3(L) \) | 0.6878 | 0.6898 | 0.7012 | 0.7046 |

**TABLE I**

VALUES OF \( \lambda_2(L) \) AND \( \lambda_3(L) \), WHICH CORRESPOND TO THE SECOND AND THIRD LARGEST EIGENVALUES OF \( T_L \) (IN ABSOLUTE VALUE), ARE COMPUTED NUMERICALLY FOR DIFFERENT VALUES OF \( L \).
Let us now describe a numerical method for computing \( \mu \) and \( q(a, b, z) \). In this regard, we follow the approach of (11). First, by (28) and the scaling law assumption we conclude that
\[
2^{-\frac{n}{2}} q(z, a, b) = q(z^2, a, b) + q(1 - z, a, b) / 2.
\]
(45)

Equation (45) can be solved numerically by recursion. In general, the equation can have many solutions. The idea here is to use the scaling assumption to properly initialize a recursion procedure to compute the desired solution of (45) that is compatible with (43) (i.e., a recursion that gives us the desired function \( q \) in (43)). Let us now describe the recursion. First of all, note that equation (45) is invariant under multiplicative scaling of \( q \). Also, from this equation one can naturally guess that \( q(z, a, b) \) can be factorized into
\[
q(z, a, b) = c(a, b) q(z),
\]
(46)
where \( q(z) \) is a solution of (45) with \( q(1) = 1 \). We iteratively compute \( \mu \) and \( q(z) \).

Initialize \( q_0(z) \) -say- with \( q_0(z) = 1_{(z \in (0, 1 / 2))} \) and compute recursively new estimates of \( q_{n+1}(z) \) by first computing
\[
q_{n+1}(z) = q_{n}(z^2) + q_{n}(1 - (1 - z)^2),
\]
and then by normalizing \( q_{n+1}(z) = \tilde{q}_{n+1}(z) / \tilde{q}_{n+1}(1 / 2) \), so that \( q_{n+1}(1 / 2) = 1 \). It is easy to see that \( \tilde{q}_n \) indeed converges to \( q(z) \) provided that the scaling assumption as well as (46) hold true. We have implemented the above functional recursion numerically by discretizing the \( z \) axis. Figure 3 shows the resulting numerical approximation of \( q_n(z) \) as obtained by iterating the above procedure until \( \|q_{n+1}(z) - q_n(z)\|_\infty \leq 10^{-10} \) \((\forall z \in [0, 1])\) and by using a discretization with \( 10^6 \) equispaced values of \( z \). From this recursion we also get a numerical
\[
c(a, b) = 0.92 \text{ match very well. Even for moderate values of } n \text{ (such as } n = 10 \text{) we observe that the curves have a fairly good agreement.}
\]

Consider the process \( Z_n \) with \( Z_0 = z = 1 / 2 \). It is easy to see that the set of possible values that \( Z_n \) takes in \([0, 1]\) is symmetric around \( z = 1 / 2 \). Now, according to the scaling law for \( x \in [0, 1] \), there is a constant \( c(1 / 2, x, 1 / 2) \approx c(x) \) such that
\[
Pr(Z_n \in [x, 1 / 2]) \approx c(x) 2^{-n},
\]
(47)
As we notice, by noticing the fact that \( Z_n \) is symmetric around the point \( z = 1 / 2 \), we get
\[
Pr(Z_n \in [0, x]) \leq 1 / 2 - c(x) 2^{-n}. \]
(48)

From the construction procedure of polar codes (and specially relation (25)), we know the following. Let \( z(1) \leq z(2) \cdots \leq z(N) \) be a re-ordering of the \( N \) possible outputs of \( Z_n \) in an ascending order. Then, by using (25) the error probability of a polar code with rate \( R \) is bounded from below by
\[
P_e \geq \max_{i \in [N: R]} E(W^{(i)}_N) = \max_{i \in [1, \cdots, N : R]} \frac{z(i)}{2} = \frac{z(N : R)}{2}. \]
(49)

So in order to achieve error probability \( P_e \), we should certainly have \( \frac{z(N : R)}{2} \leq P_e \) or \( z(N : R) \leq 2P_e \). As a result, we obtain
\[
R \leq \Pr(Z_n \in [0, 2P_e]),
\]
(50)

Note that if \( W \) is a BEC, then we have \( Z(W) = 2E(W) \). Also, for general BMS channels we have the relation (8).

Fig. 3. The functions \( 2^{-n} q_n(a, b, z) \) for various values of \( n \). Here we have fixed \( a = 1 - b = 0.9 \) and \( \frac{\mu}{2} = 0.2757 \). In all of the four plots the dashed curve corresponds to \( c(a, b)q(z) \) with \( c(a, b) = 0.92 \). Here, the function \( q(z) \) corresponds to the numerical solution of (45).

Fig. 4. The function \( q(z) \) for \( z \in [0, 1] \).

estimate of the scaling exponent \( \mu \). In particular we expect \( \tilde{q}_n(1 / 2) \to 2^{1 + \mu} \) as \( n \to \infty \). Using this method, we obtain the estimate \( 1 / \mu = 0.2757 \).

As mentioned above, the function \( q(a, b, z) \) differs from \( q(z) \) by a multiplicative constant \( c(a, b) \) that is to be found by other means. In Figure 5 we plot the functions \( 2^{-n} q_n(z, a, b) \) for \( a = 1 - b = 1 / 10 \) and different values of \( n \). We observe that, as \( n \) increases these plots and the curve \( c(a, b)q(z) \) with

Note that choosing \( q(1) = 1 \) is an arbitrary normalization choice.

This is an arbitrary choice for \( q_0(z) \). One can try other starting points, e.g. \( q_0(z) = 1_{(z \leq 1)} \) or \( q_0(z) = 4z(1 - z) \). All the initial points that we have tried have led to the same \( q(z) \). This is indeed compatible with the scaling assumption and (46).
and by using (48) we deduce that
\[ R \leq \Pr(Z_n \in [0, 2P_c]) \]
\[ \leq \frac{1}{2} - c'(2P_c)^2 \Theta \]
\[ = \frac{1}{2} - c(P_c)N^{-\frac{1}{3}}, \]
and finally
\[ N \geq \left( \frac{c(P_c)}{\frac{1}{2} - R} \right)^\mu. \] (50)

Now, from the above calculations we know that \( \frac{1}{\mu} \approx 0.2757 \).

As a result, for the channel \( W = \text{BEC}(\frac{1}{2}) \) we have
\[ N \geq \Theta \left( \frac{1}{(U(W) - R)^{3.627}} \right). \] (51)

For other empirical scaling laws of this type, we refer to [11]. In the next section, we provide methods that analytically validate the above observations. We also extend some of these observations and results to other BMS channels.

IV. ANALYTICAL APPROACH: FROM BOUNDS FOR THE BEC TO UNIVERSAL BOUNDS FOR BMS CHANNELS

In this section we provide a rigorous basis for the observations and derivations of the previous section. Proving the full picture of Section III is beyond what we achieve here, but we come up with close and useful bounds. As previously mentioned, we only use the heuristic arguments as well as the numerical observations of the previous section to give an intuitive picture for the ideas and proofs of this section. In other words, the proofs of this section do not rely on any of the assumptions of the previous section and can be read independently.

This section consists of three smaller parts. In the first part we provide lower and upper bounds on the speed of polarization for the BEC family. Similar types of bounds are obtained for general BMS channels in the second part. Finally, in the last part we use these bounds to derive trade-offs between the rate and the block-length for polar codes.

A. Speed of Polarization for the BEC Family

The (heuristic) arguments of the previous section led us to the conclusion that (see (45)) for the channel \( W = \text{BEC}(z) \) the quantity \( \Pr(Z_n \in [a, b]) \) vanishes in \( n \) like \( \Theta(2^{-\frac{1}{3}}) \) (here, \( z, a, b \in (0, 1) \)). In other words, the speed of polarization for the process \( Z_n \) is equal to \( \frac{1}{n} \). The value of \( \mu \) was also computed to be \( \mu \approx 3.627 \) (or \( \frac{1}{\mu} \approx 0.2757 \)).

Analytically speaking, proving the scaling assumption (43) seems to be a difficult task. It is not even clear whether the value \( \mu \) exists. The objective of this section is to provide (analytical) lower and upper bounds on the value \( \mu \).

More precisely, we look for numbers \( \mu \) and \( \overline{\mu} \) such that
\[ \Theta(2^{-\frac{1}{3}}) \leq \Pr(Z_n \in [a, b]) \leq \Theta(2^{-\frac{1}{3}}), \]
and in words, the speed of polarization of \( Z_n \) is bounded between the values \( \frac{1}{\overline{\mu}} \) and \( \frac{1}{\mu} \).

In this regard, we provide two approaches that exploit different techniques. The first approach is based on a more careful look at equation (32). From the arguments of the previous section, the value \( \mu \) is related to a significant (and non-trivial) eigenvalue of the polar operator \( T \). Here, we observe that simple bounds can be derived on the this eigenvalue of \( T \) by carefully analyzing the effect of \( T \) on some suitably chosen test functions. This approach provides us with a sequence of analytic bounds on \( \mu \). We conjecture (and observe empirically) that these bounds indeed converge to the value of \( \mu \) that is computed in Section III. The second approach considers all the possible compositions of the two operations \( z^2 \) and \( 2z - z^2 \) and analyzes the asymptotic behavior of these compositions. This approach provides us with a good lower bound on \( \mu \).

1) First Approach: Let us begin by providing an intuitive picture behind the first approach. This picture is only intended for a better explanation of the contents that appear later. Hence, these explanations can be skipped without losing the main track. Consider the polar operator defined in (32) and its eigenvalues which are the solutions of
\[ T(q) = \lambda q. \] (52)

A check shows that both \( q(z) = z \) and \( q(z) = 1 \) are eigenfunctions associated to the eigenvalue \( \lambda = 1 \). Perhaps more interestingly, let us look at the eigenvalues of \( T \) inside the interval \((0, 1)\). Intuitively, equation (45) together with the scaling law (43) can be reformulated as follows. The operator \( T \) has an eigenvalue \( \lambda = 2^{-\frac{1}{3}} \) and a corresponding eigenfunction \( q(z) \) such that if we take any step function \( f(z) = \mathbb{1}_{[z \in [a, b]]} \), then
\[ \lambda^{-n}T^n(f) \xrightarrow{n \to \infty} c(a, b)q(z). \] (53)

Therefore, for \( f(z) = \mathbb{1}_{[z \in [a, b]]} \), the value of \( T^n(f) \) vanishes in \( n \) like \( \Theta(\lambda^n) \) (or equivalently \( \Theta(2^{-\frac{n}{3}}) \)). In fact, if the scaling law is true, then we naturally expect that (53) holds for a much larger class of functions rather than the class of step functions. Heuristic arguments of the previous section also suggest that (53) holds at least for all (piece-wise) continuous functions \( f(z) \) with \( f(0) = f(1) = 0 \). Therefore, for any function \( f \) in this larger class of functions the value of \( T^n(f) \) decays like \( \Theta(\lambda^n) \) (or equivalently \( \Theta(2^{-\frac{n}{3}}) \)). So to compute (or to provide bounds on) the value of \( \mu \), one can look for suitable continuous functions \( f \) such that the speed of decay of \( T^n(f) \) is “easy” to compute (or provide bounds on). As we will see, functions in the form of \( f(z) = z^\alpha(1 - z)^\beta \) are among such suitable functions.

Motivated by this picture, let us formalize the first approach to find bounds on the speed of polarization of \( Z_n \) (or the value \( \mu \)) through the following two steps: (1) choose a suitable “test function” \( f(z) \) for which we can provide good bounds on how fast \( T^n(f) \) approaches \( 0 \) (in \( n \)), and (2) turn these bounds into bounds on the speed for polarization of \( Z_n \) (or \( \mu \)). With this in mind, for a generic test function \( f(z) : [0, 1] \to [0, 1] \), let us define the sequence of functions \( \{f_n(z)\}_{n \in \mathbb{N}} \) as
\[ f_n(z) = \mathbb{E}[f(Z_n) \mid Z_0 = z] = T^n(f). \] (54)

Here, note that for \( z \in [0, 1] \) the value of \( f_n(z) \) is a deterministic value that is dependent on the choice of \( f \) and
the process $Z_n$ with the starting value $Z_0 = z$. Let us now recall once more the recursive relation of the functions $f_n$:

$$f_0(z) = f(z),$$

$$f_n(z) = \frac{f_{n-1}(z^2) + f_{n-1}(1 - (1-z)^2)}{2}.$$  

(55)

In order to find lower and upper bounds on the speed of decay of the sequence $f_n$, we define sequences of numbers $\{a_m\}_{m\in\mathbb{N}}$ and $\{b_m\}_{m\in\mathbb{N}}$ as

$$a_m = \inf_{z\in(0,1)} \frac{f_{m+1}(z)}{f_m(z)},$$

(56)

$$b_m = \sup_{z\in(0,1)} \frac{f_{m+1}(z)}{f_m(z)}.$$  

(57)

**Lemma 3:** Fix $m \in \mathbb{N}$. For all $n \geq m$ and $z \in (0,1)$, we have

$$\left( a_m \right)^{n-m} f_m(z) \leq f_n(z) \leq \left( b_m \right)^{n-m} f_m(z).$$  

(58)

Furthermore, the sequence $a_m$ is an increasing sequence and the sequence $b_m$ is a decreasing sequence.

**Proof:** Here, we only prove the left-hand side of (58) and note that the right-hand side follows similarly. The proof goes by induction on $n - m$. For $n = m = 0$ the result is trivial. Assume that the relation (58) holds for a $n - m \leq k$, i.e., for $z \in (0,1)$ we have

$$\left( a_m \right)^k f_m(z) \leq f_{m+k}(z).$$  

(59)

We show that (58) is then true for $k + 1$ and $z \in (0,1)$. We have

$$f_{m+k+1}(z) = \frac{f_{m+k}(z^2) + f_{m+k}(1 - (1-z)^2)}{2}$$

$$\geq \left( a_m \right)^k f_m(z) + \left( a_m \right)^k f_m(1 - (1-z)^2)$$

$$= \left( a_m \right)^k f_{m+1}(z)$$

$$= \left( a_m \right)^k \frac{f_{m+1}(z)}{f_m(z)} f_m(z)$$

$$\geq \left( a_m \right)^k \left[ \inf_{z\in(0,1)} \frac{f_{m+1}(z)}{f_m(z)} \right] f_m(z)$$

$$= \left( a_m \right)^{k+1} f_m(z).$$

Here, (a) follows from (55) and (b) follows from (59), and hence the lemma is proved via induction.

Finally, the sequence $a_m$ increases by $m$ because if we plug in $k = 1$ to the above set of inequalities, and stop after the third line, then we obtain that $f_{m+2}(z) \geq a_m f_{m+1}(z)$ for $z \in (0,1)$. From this and the definition of $a_m$ in (56), it is then easy to see that $a_{m+1} \geq a_m$.

Let us now begin searching for suitable test functions, i.e., candidates for $f(z)$ that provide us with good lower and upper bounds $a_m$ and $b_m$. First of all, it is easy to see that a test function $f(z) = \mathbb{1}_{\{z\in[a,b]\}}$ results in trivial values of $a_m$ and $b_m$ (namely $b_m = \infty$ and $a_m$ is not well-defined), and hence such test functions are not suitable for this bounding technique. Second, we expect that having a polynomial test function might be slightly preferable. This is due to the fact that if $f$ is a polynomial, then $T^a(f)$ is also a polynomial and computing $a_m$ and $b_m$ is equivalent to finding roots of polynomials which is a manageable task. Of course the simplest polynomial that takes the value $0$ on $z = 0, 1$ is $f(z) = z(1-z)$. Hence, let us take our test function as $f(z) = z(1-z)$ and consider the corresponding sequence of functions $\{f_n(z)\}_{n\in\mathbb{N}}$ with $f_0(z) = f(z)$ and $f_n(z) = E[Z_n(1-Z_n)] = T^n(f_0)$.

(60)

A moment of thought shows that with $f_0 = z(1-z)$ the function $2^n f_n$ is a polynomial of degree $2^{n+1}$ with integer coefficients. Let us first focus on computing the value of $a_m$ for $m \in \mathbb{N}$.

**Remark 4:** One can compute the value of $a_m$ by finding the extreme points of the function $\text{sup}_{j\in\mathbb{N}} f_j$ (i.e., finding the roots of the polynomial $g_m = f_{m+1} f_m - f_{m+1} f_m'$), and then minimizing the function $\frac{f_j}{g_m}$ on these extreme points as well as boundary points. Assuming $f_0 = z(1-z)$, for small values e.g., $m = 0, 1$, pen and paper suffice to find the extreme points. For higher values of $m$, we can automatize the process: all these polynomials have rational coefficients and therefore it is possible to determine the number of real roots exactly and to determine their value to any desired precision. This task can be accomplished precisely by computing so-called Sturm chains (see Sturm’s Theorem [18]). Computing Sturm chains is equivalent to running Euclid’s algorithm starting with the second and third derivative of the original polynomial. Hence, we can analytically find the value of $a_m$ to any desired precision. Table II contains the numerical value of $a_m$ up to precision $10^{-4}$ for $m \leq 10$. As the table shows, the values $a_m$ are increasing (see Lemma 3), and we conjecture that they converge to $2^{-0.2757} = 0.8260$, the corresponding value for the channel BEC.

We now focus on computing the value of $b_m$. On the negative side, for the specific test function $f(z) = z(1-z)$ we obtain $b_m = 1$ for $m \in \mathbb{N}$ and therefore the upper bounds implied by (57) are trivial. In fact, it is not hard to show that if we plug in any polynomial as the test function then we get $b_m = 1$ for any $m$. On the positive side, we can consider other test functions that result in non-trivial values for $b_m$. The problem with non-polynomial functions is that methods such as the Sturm-chain method no longer apply. Hence, finding the precise value of $b_m$ up to any desired precision can in general be a difficult task and we might lose the analytical tractability of $b_m$. As an example, choose

$$f_0(z) = z^\alpha (1-z)^\beta,$$

(61)

\footnote{Note that in spite of the fact that the supremum and the infimum are defined for $z \in (0,1)$, we should check the value of $\frac{f_{m+1}}{f_m}$ around the boundary points $z = 0, 1$.}

\footnote{This follows from repeated applications of L’Hôpital’s rule.}

| $m$ | 0 | 2 | 4 | 6 | 10 |
|-----|---|---|---|---|----|
| $a_m$ | 0.75 | 0.7897 | 0.8074 | 0.8180 | 0.8239 |
| $\log a_m$ | -0.4150 | -0.3406 | -0.3086 | -0.2880 | -0.2794 |
for some choice of $\alpha, \beta \in (0, 1)$. Then, from (57) we have
\[
 b_0 = \sup_{z \in (0,1]} f_1(z) = \sup_{z \in [0,1]} \frac{z^\alpha (1+z)^\beta + (2-z)^\alpha (1-z)^\beta}{2}.
\]
(62)

We can compute $b_0$ to any desired precision either by finding the extreme points of the expression in (62), or by simple numerical methods.

**Remark 5:** Let us explain what we mean by a simple numerical method. The idea is to take a fine grid for the unit interval and maximise the right-hand side of (62) on this grid. Let $g(z) = z^\alpha (1+z)^\beta + (2-z)^\alpha (1-z)^\beta$. We now describe briefly a numerical procedure to find precisely the maximum value that $g$ attains over $[0, 1]$: (i) Fix a number $\delta > 0$. The function $g$ has a finite derivative on the interval $(\delta, 1 - \delta)$. Thus, the maximum of $g$ over the interval $(\delta, 1 - \delta)$ can be found to any desired precision by making the grid sufficiently fine. (ii) The maximum value of $g$ over the region $[0, \delta] \cup [1 - \delta, 1]$ can be upper-bounded by simple Taylor-type methods. This upper bound becomes tighter when $\delta$ is smaller. It is then straightforward to conclude that by this procedure we can compute, to any desired precision, the maximum value that $g$ attains over the unit interval (provided that we choose a sufficiently small $\delta$ and grid size).

By letting $\alpha = \beta = \frac{2}{3}$, we obtain $b_0 = 0.8312$ which is already a good bound for $\lambda$ (recall from the calculations done in Section III that $\lambda = 2^{-0.2757} = 0.8260$). This suggests that the test function $f_0(z) = (z(1-z))^{\frac{2}{3}}$ is a suitable candidate for obtaining good upper bounds. For this specific test function, the value of $b_0$ for various values of $m$ has been numerically computed in Table III. As we observe from Table III even for moderate values of $m$ the (numerically computed) bound $b_m$ is very close to the “true” value of $\lambda$.

Finally, let us relate the bounds $a_m$ and $b_m$ to bounds on the value of $\Pr(Z_n \in [a, b])$. This is the subject of the following lemma which is proven in Appendix A.

**Lemma 6:** Let $a, b \in (0, 1)$ be such that $\sqrt{a} \leq 1 - \sqrt{1-b}$. Then, there exists a constant $c_1 > 0$ such that for any $z \in [0, 1]$
\[
\frac{1}{n} \log \mathbb{E}[Z_n(1-Z_n)] - \frac{c_1 \log n}{n} \leq \frac{1}{n} \log (2^{-n} + \Pr(Z_n \in [a,b]))
\]
(63)
where $Z_n$ is defined in (26) with $Z_0 = z$. Also, for any continuous function $f : [0, 1] \to [0, 1]$ such that $f(z) > 0$ for $z \in (0, 1)$, we have for $a, b \in (0, 1)$ that
\[
\frac{1}{n} \log \Pr(Z_n \in [a, b]) \leq \frac{1}{n} \log \mathbb{E}[f(Z_n)] + \frac{c_3}{n}
\]
(64)
where $c_3$ is a positive constant that depends on $a, b$, and $f$. Examples of such function $f$ can be $f(z) = z(1-z)$ or $f(z) = (z(1-z))^{\frac{2}{3}}$.

We can now easily conclude the following.

**Corollary 7:** Fix $m \in \mathbb{N}$. For $a, b \in (0, 1)$ such that $\sqrt{a} \leq 1 - \sqrt{1-b}$ and $n \geq m$ we have
\[
\log a_m + O\left(\frac{\log n}{n}\right) \leq \frac{1}{n} \log (2^{-n} + \Pr(Z_n \in [a,b])) 
\]
\[
\leq \log b_m + O\left(\frac{1}{n}\right),
\]
(65)
where $a_m$ is defined in (66) with the test function $f(z) = z(1-z)$ (see Table II), and $b_m$ is defined in (66) with the test function $f(z) = (z(1-z))^{\frac{2}{3}}$ (see Table III).

**Remark 8:** Two comments are in order: (i) The additional term $2^{-n}$ in (63) and (65) is to avoid trivial conflicts when $\Pr(Z_n \in [a,b]) = 0$. However, these cases are very rare as for every $z \in (0, 1)$ and $a, b \in (0, 1)$ s.t. $\sqrt{a} \leq 1 - \sqrt{1-b}$, it is not hard to prove that there exists an integer $n_0 \in \mathbb{N}$ such that we have $\Pr(Z_n \in [a,b]) > 0$ for $n \geq n_0$. Note that if $\Pr(Z_n \in [a,b]) > 0$, then we certainly have $\Pr(Z_n \in [a,b]) > 2^{-n}$. (ii) We expect that the result of Lemma 6 holds for any choice of $a$ and $b$ such that $a < b$. That is, the condition $\sqrt{a} \leq 1 - \sqrt{1-b}$ is only a technical restriction.

2) Second Approach: We will now explain another approach for finding the value $\mu$ for the BEC. Let us point out the fact that the content of this section (Section IV-A2) is not necessary for the forthcoming parts of the paper and hence can be skipped without losing the main track.

Throughout this section we will prove the following theorem.

**Theorem 9:** We have
\[
\liminf_{n \to \infty} \frac{1}{n} \log \left( \int_0^1 \Pr(Z_n \in [a,b])dz \right) \geq \frac{1}{2 \ln 2} - 1 \approx -0.2787.
\]
(66)
Let us now explain, at the intuitive level, the main consequence of Theorem 9. By using the scaling law assumption, and specifically (42) and (43), we have that
\[
\int_0^1 \Pr(Z_n \in [a,b])dz \approx \int_0^1 z^{2-\frac{2}{3}} q(z, a, b)dz + o(2^{-\frac{4}{3}}).
\]
This relation together with (66) implies that $\mu \geq \frac{1}{2 \ln 2} - 1 \approx -0.2787$. For the sake of brevity, we do not address here further (analytic) conclusions of Theorem 9 and we refer the reader to [12].

To proceed with the proof of Theorem 9 let us recall from Section II-D the definition of $Z_n$ (for the BEC) in terms of the sequence $\{B_n\}_{n \in \mathbb{N}}$. We start by $Z_0 = z$ and
\[
Z_{n+1} = \begin{cases} 
Z_{n-1}^2 & \text{if } B_n = 1, \\
2Z_{n-1} - Z_{n-1}^2 & \text{if } B_n = 0.
\end{cases}
\]
(67)

Hence, by considering the two maps $t_0, t_1 : [0, 1] \to [0, 1]$ defined as
\[
t_0(z) = 2z - z^2, \quad t_1(z) = z^2
\]
(68)
the value of $Z_n$ is obtained by applying $t_{B_n}$ on the value of $Z_{n-1}$, i.e.,
\[
Z_n = t_{B_n}(Z_{n-1}).
\]
(69)
The same rule applies for obtaining the value of $Z_{n-1}$ form $Z_{n-2}$ and so on. Thinking this through recursively, the value of $Z_n$ is obtained from the starting point of the process, $Z_0 = z$, via the following (random) maps [16].

---

[16] The necessary notation is reviewed in Section II-D.
**Definition 10:** For each \( n \in \mathbb{N} \) and a realization \( (b_1, \cdots, b_n) \in \omega_n \in \Omega_n \) define the map \( \phi_{\omega_n} \) by
\[
\phi_{\omega_n} = t_{b_n} \circ t_{b_{n-1}} \circ \cdots \circ t_{b_1}.
\] (70)

Also, let \( \Phi_n \) be the set of all such \( n \)-step maps.

As a result, an equivalent description of the process \( Z_n \) is as follows. At time \( n \) the value of \( Z_n \) is obtained by picking uniformly at random one of the functions \( \phi_{\omega_n} \in \Phi_n \) and assigning the value \( \phi_{\omega_n}(z) \) to \( Z_n \). Consequently we have,
\[
\Pr(Z_n \in [a, b]) = \sum_{\omega_n \in \Omega_n} \frac{1}{2^n} \mathbb{1}_{\{\phi_{\omega_n}(z) \in [a, b]\}}.
\] (71)

By using (71), it is apparent that in order to analyze the behavior of the quantity \( \Pr(Z_n \in [a, b]) \) as \( n \) grows large, it is necessary to characterize the asymptotic behavior of the random maps \( \phi_{\omega_n} \). Continuing the theme of Definition 10, we can assign to each realization of the infinite sequence \( \{b_k\} \in \Omega \), a sequence of maps \( \phi_{\omega_n}(z), \phi_{\omega_n}(z), \cdots \) where \( \omega_i \in \{b_1, \cdots, b_i\} \). We call the sequence \( \{\phi_{\omega_n}\}_{n \in \mathbb{N}} \) the corresponding sequence of maps for the realization \( \{b_k\}_{k \in \mathbb{N}} \).

We also use the realization \( \{b_k\}_{k \in \mathbb{N}} \) and its corresponding \( \{\phi_{\omega_n}\}_{n \in \mathbb{N}} \) interchangeably. Let us now focus on the asymptotic characteristics of the functions \( \phi_{\omega_n} \). Firstly, since \( \{\phi_{\omega_n}(z)\}_{\omega_n \in \Omega} \) has the same law as \( Z_n \) starting at \( Z_0 = z \), we conclude that for \( z \in [0, 1] \), with probability one, the quantity \( \lim_{n \to \infty} \phi_{\omega_n}(z) \) takes on a value in the set \( \{0, 1\} \). In Figure 6 the functions \( \phi_{\omega_n} \) are plotted for a random realization. As it is apparent from the figure, the functions \( \phi_{\omega_n} \) seem to converge point-wise to a jump function (i.e., a sharp rise from 0 to 1). An intuitive justification of this fact is as follows. Consider a random function \( \phi_{\omega_n} \). Due to polarization, as \( n \) grows large, almost all the values that this function takes are very close to 0 or 1. This function is also increasing and continuous (more precisely, it is a polynomial). A little thought reveals that the only choice to imagine for \( \phi_{\omega_n} \) is a very sharp rise from being almost 0 to almost 1. The formal and complete statement is given as follows.

**Lemma 11** (Almost every realization has a threshold point): For almost every realization of \( \omega = \{b_k\}_{k \in \mathbb{N}} \in \Omega \), there exists a point \( z^*_{\omega} \in [0, 1] \) such that
\[
\lim_{n \to \infty} \phi_{\omega_n}(z) = \begin{cases} 
0 & z \in [0, z^*_{\omega}) \\
1 & z \in (z^*_{\omega}, 1]
\end{cases}
\]
Furthermore, \( z^*_{\omega} \) has uniform distribution on \([0, 1]\). We call the point \( z^*_{\omega} \) the threshold point of the realization \( \{b_k\}_{k \in \mathbb{N}} \) or the threshold point of its corresponding sequence of maps \( \{\phi_{\omega_n}\}_{n \in \mathbb{N}} \).

Looking more closely at (71), by the above lemma we conclude that as \( n \) grows large, the maps \( \phi_{\omega_n} \) that activate the identity function \( \mathbb{1}_{\{\cdot\}} \) must have their threshold point sufficiently close to \( z \). Let us now give an intuitive discussion about the idea behind the proof of Theorem 9. By using (71) we can write
\[
\Pr(Z_n \in [a, b]) = \sum_{\omega_n \in \Omega_n} \frac{1}{2^n} \mathbb{1}_{\{\phi_{\omega_n}(z) \in [a, b]\}} = \sum_{\omega_n \in \Omega_n} \frac{1}{2^n} \mathbb{1}_{\{z \in [\phi^{-1}_n(a), \phi^{-1}_n(b)]\}}.
\] (72)

Hence, by Lemma 11 for a large choice of \( n \) the intervals \( [\phi^{-1}_n(a), \phi^{-1}_n(b)] \) have a very short length and are distributed almost uniformly along \([0, 1]\). Now, if we assume that the length of the intervals \( [\phi^{-1}_n(a), \phi^{-1}_n(b)] \) is very close to their average, then we can replace the average in (72) by the average length of \( [\phi^{-1}_n(a), \phi^{-1}_n(b)] \). That is,
\[
\Pr(Z_n \in [a, b]) \approx \mathbb{E}[\phi^{-1}_n(b) - \phi^{-1}_n(a)].
\]

So intuitively, all that remains is to compute the average length of the random intervals \( [\phi^{-1}_n(a), \phi^{-1}_n(b)] \).

In fact we are not able to make all these heuristics precise for the point-wise values \( \frac{1}{n} \log \Pr(Z_n \in [a, b]) \). Nonetheless, the picture is naturally precise for the average of \( \Pr(Z_n \in [a, b]) \) over \( z \in [0, 1] \), i.e.,
\[
\frac{1}{n} \log \left\{ \int_0^1 \Pr(Z_n \in [a, b])dz \right\}.
\] (73)

To see this, we proceed as follows. By (72) we have
\[
\int_0^1 \Pr(Z_n \in [a, b])dz = \int_0^1 \left\{ \sum_{\omega_n \in \Omega_n} \frac{1}{2^n} \mathbb{1}_{\{z \in [\phi^{-1}_n(a), \phi^{-1}_n(b)]\}}dz \right\} = \sum_{\omega_n \in \Omega_n} \frac{1}{2^n} \int_0^1 \mathbb{1}_{\{z \in [\phi^{-1}_n(a), \phi^{-1}_n(b)]\}}dz = \mathbb{E}[\phi^{-1}_n(b) - \phi^{-1}_n(a)],
\]
and by applying \( \frac{1}{n} \log(\cdot) \) to both sides we have
\[
\frac{1}{n} \log \left\{ \int_0^1 \Pr(Z_n \in [a, b])dz \right\} = \frac{1}{n} \log \mathbb{E}[\phi^{-1}_n(b) - \phi^{-1}_n(a)] \\
\geq \frac{1}{n} \mathbb{E}[\log(\phi^{-1}_n(b) - \phi^{-1}_n(a))],
\] (74)
where in the last step we have used Jensen’s inequality. The value of \( \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\log(\phi^{-1}_n(b) - \phi^{-1}_n(a))] \) can be computed precisely.
Lemma 12: We have
\[ \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\log(\phi_{m}^{-1}(b) - \phi_{m}^{-1}(a)))] = \frac{1}{2 \ln 2} - 1 \approx -0.2787. \]
As a result, we have
\[ \liminf_{n \to \infty} \frac{1}{n} \log \{ \int_{0}^{1} \Pr(Z_n \in [a, b])dz \} \geq \frac{1}{2 \ln 2} - 1. \]

The result of Theorem 9 provides a lower bound on \( \mu \) that is very close to the value we obtained in Section III but is not exactly equal. This is because we have used Jensen’s inequality in (74).

B. Speed of Polarization for General BMS Channels

In the previous part, we derived bounds on the speed of polarization for the process \( Z_n \) associated to the BEC. To this end, we used the recursion (26) for \( Z_n \) and the fact that the speed of polarization can be “measured” by computing the rate of decay of a sequence \( \{\mathbb{E}[f(Z_n)]\}_{n \in \mathbb{N}} \), where \( f \) is a suitable “test” function such as \( f(z) = z(1-z) \).

In this part, we use a similar approach to bound the speed of polarization for any BMS channel. For a BMS channel \( W \), there is no simple and closed-form (scalar) recursion for the process \( Z_n \) as for the BEC. However, by using (13) and (14), we can provide bounds on how \( Z_n \) evolves:
\[ Z_{n+1} = \begin{cases} Z_n^2 / \sqrt{2 - Z_n^2}, & \text{if } B_n = 1, \\ 2Z_n - Z_n^2, & \text{if } B_n = 0. \end{cases} \tag{75} \]

As a warm-up, we notice that similar techniques as used in Section IV-A1 can be used to provide general lower and upper bounds. For instance, to find upper bounds we can proceed as follows. For any continuous function \( g : [0, 1] \to \mathbb{R} \) such that \( g(0) = g(1) = 0 \) and \( g(z) > 0 \) for \( z \in (0, 1) \), let
\[ L_g = \sup_{z \in (0, 1), g \in \mathcal{J}} \frac{g(z^2) + g(y)}{2g(z)}. \tag{76} \]

Similar to the discussion in Section IV-A1 (in particular the proof of Lemma 3), it is easy to see from (75) and (76) that for \( n \in \mathbb{N} \)
\[ \mathbb{E}[g(Z_{n+1}) | Z_n] \leq g(Z_n) L_g, \]
and consequently,
\[ \mathbb{E}[g(Z_n)] \leq g(0) L_n^g. \tag{77} \]

As a result, for the process \( Z_n = Z(W_n) \) we have
\[ \mathbb{E}[g(Z_n)] \leq cL_n^g, \tag{78} \]
where \( c = \sup_{z \in [0, 1]} g(z) \) is a constant. Also, by using the Markov inequality we have for \( a, b \in (0, 1) \),
\[ \frac{1}{n} \log \Pr(Z_n \in [a, b]) \leq \log L_g + O\left(\frac{1}{n}\right). \tag{79} \]

It thus remains to find good candidates for the function \( g \) (with the properties mentioned above) such that the value \( L_g \) defined in (78) is minimized. For instance, we can let the function \( g \) take the following closed form:
\[ g(z) = (az^2 + bz + c)(z(1-z))^{d} \]
where \( a, b, c, d \in (0, 1) \) and optimize the value of \( L_g \) over the choice of \( a, b, c, d \). For example, by choosing \( a = b = 0 \), \( c = 1 \), and \( d = \frac{2}{3} \), we have \( g(z) = (z(1-z))^{\frac{2}{3}} \) and we obtain \( \log L_g = -0.169 \). That is
\[ \mathbb{E}[g(Z_n)] \leq 2^{-0.169} \tag{80} \]

Also, by choosing \( a = \frac{2}{3}, b = \frac{1}{4}, c = \frac{4}{27}, \) and \( d = \frac{1}{4} \), we have \( g(z) = \frac{1}{20}(8z^2 + 5z + 19)(z(1-z))^{\frac{2}{3}} \). The value of \( L_g \) can be computed to a desirable precision using simple numerical methods (see Remark 5). We thus obtain \( \log L_g = -0.202 \) and as a result
\[ \mathbb{E}[g(Z_n)] \leq 2^{-0.202} \tag{81} \]

As a final remark, we note that for \( z \in [0, 1] \) we have \( g(z) \geq \frac{3}{4}(z(1-z))^{\frac{2}{3}} \). Therefore, we can conclude that for any BMS channel \( W \) we have
\[ \mathbb{E}[g(Z_n)] \leq 2^{0.202} \tag{82} \]

The relations of type (79) and (82) are upper bounds on the speed of polarization that hold universally over all BMS channels. Let us now compute universal lower bounds. In the rest of this section, it is more convenient for us to consider another stochastic process related to \( W_n \), which is the process \( H_n = H(W_n) \). The main reason to consider \( H_n \) rather than \( Z_n \) is that the process \( H_n \) is a martingale and this martingale property will help us to use the functions \( \{f_n\}_{n \in \mathbb{N}} \) defined in (85) (with the starting function \( f(z) = z(1-z) \)) to provide universal lower bounds on the quantity \( \mathbb{E}[H_n(1 - H_n)] \). We begin by introducing one further technical condition given as follows.

Definition 13: We call an integer \( m \in \mathbb{N} \) suitable if the function \( f_m(z) \), defined in (85) (with the starting function \( f(z) = z(1-z) \)), is concave on \( [0, 1] \).

Remark 14: For small values of \( m \), i.e., \( m \leq 2 \), it is easy to verify by hand that the function \( f_m \) is concave. As discussed previously, for larger values of \( m \) we can use Sturm’s theorem [18] and a computer algebra system to verify this. Note that the polynomials \( 2^m f_m \) have integer coefficients. Hence, all the required computations can be done exactly. We have checked up to \( m = 10 \) that \( f_m \) is concave and we conjecture that in fact this is true for all \( m \in \mathbb{N} \). We now show that for any BMS channel \( W \), the value of \( a_m \), defined in (85), is a lower bound on the speed of polarization of \( H_n \) provided that \( m \) is a suitable integer.

Lemma 15: Let \( m \in \mathbb{N} \) be a suitable integer and \( W \) a BMS channel with \( I(W) \in (0, 1) \). We have for \( n \geq m \)
\[ \mathbb{E}[H_n(1 - H_n)] \geq (a_m)^{n-m} f_m(H(W)), \tag{83} \]
where \( a_m \) is given in (85).

17For the BEC the processes \( H_n \) and \( Z_n \) are identical.
Proof: We use induction on \( n - m \); for \( n - m = 0 \) there is nothing to prove. Assume that the result of the lemma is correct for \( n - m = k \). Hence, for any BMS channel \( W \) with \( H_n = H(W_n) \) we have

\[
\mathbb{E}[H_{m+k}(1 - H_{m+k})] \geq (a_m)^k f_m(H(W)).
\]  

(84)

We now prove the lemma for \( m - n = k + 1 \). For the BMS channel \( W \), let us recall from Section [18] that the transform \( W \to (W^0, W^1) \) yields two channels \( W^0 \) and \( W^1 \) such that (12) holds. Define the process \( \{W^n\}_{n \in \mathbb{N}} \) as the channel process that starts with \( W^0 \) and evolves as in (17). We define \( \{W^n\}_{n \in \mathbb{N}} \) similarly. Furthermore, define the two processes \( H_n^0 = H((W^n)_0) \) and \( H_n^1 = H((W^n)_1) \). We have,

\[
\mathbb{E}[H_{m+k+1}(1 - H_{m+k+1})] \\
\geq \frac{\mathbb{E}[H_{m+k}^0(1 - H_{m+k}^0) + H_{m+k}^1(1 - H_{m+k}^1)]}{2} \\
\geq (a_m)^k \frac{f_m(H(W^0))}{2} + f_m(H(W^1)) \\
\geq (a_m)^k \frac{f_m(H(W))}{2} \\
\geq (a_m)^k f_m(H(W)) \\
= (a_m)^k \frac{\max_{h \in (0,1)} f_m(h)}{\min_{h \in (0,1)} f_m(h)} f_m(H(W)) \\
\geq (a_m)^k \min_{h \in (0,1)} f_m(h) f_m(H(W)) \\
\geq (a_m)^k f_m(H(W)).
\]

(85)

In the above chain of inequalities, relation (a) follows from the fact that \( W_m \) has \( 2^m \) possible outputs among which half of them are branched out from \( W^0 \) and the other half are branched out from \( W^1 \). Relation (b) follows from the induction hypothesis given in (84). Relation (c) follows from (23), (24) and the fact that the function \( f_m \) is concave. More precisely, because \( f_m \) is concave on \([0,1]\), we have the following inequality for any sequence of numbers \( 0 \leq x' \leq x \leq y \leq y' \leq 1 \) that satisfy \( \frac{x+y}{2} = \frac{x'+y'}{2} \):

\[
\frac{f_m(x') + f_m(y')}{2} \leq \frac{f_m(x) + f_m(y)}{2}.
\]  

(85)

In particular, we set \( x' = H(W^1), x = H(W^1), y = H(W^0), y' = 1 - (1 - H(W^0))^2 \) and we know from (23) and (24) that \( 0 \leq x' \leq x \leq y \leq y' \leq 1 \). Hence, by (85) we obtain (c). Relation (d) follows from the recursive definition of \( f_m \) given in (55). Finally, relation (e) follows from the definition of \( a_m \) given in (56).

Up to now, we have provided bounds on the speed of polarization for the BEC as well as general BMS channels. In the final part of this section, we rigorously relate the results obtained in previous parts to finite-length performance of polar codes. In other words, answering Question [4] stated in Section [11] is the main focus for the remaining part of this section.

C. Universal Bounds on the Scaling Behavior of Polar Codes

1) Universal Lower Bounds: Consider a BMS channel \( W \) and let us assume that a polar code is required with block-error probability at most a given value \( P_e < 1 \). One way to accomplish this is to ensure that the right side of (25) is less than \( P_e \). However, this is only a sufficient condition that might not be necessary. Hence, we call the right side of (25) the strong reliability condition. Numerical and analytical investigations (see [11] and [19]) suggest that once the sum of individual errors in the right side of (25) is less than 1, then it provides a fairly good estimate of \( P_e \). In fact, the smaller the sum is the closer it is to \( P_e \). Hence, the sum of individual errors can be considered as a fairly accurate proxy for \( P_e \). Based on this measure of the block-error probability, we provide bounds on how the rate \( R \) scales in terms of the block-length \( N \).

Theorem 16: For any BMS channel \( W \) with capacity \( I(W) \in (0,1) \), there exist constants \( P_e, \alpha > 0 \), that depend only on \( I(W) \), such that

\[
\sum_{n \in \mathbb{N}, R} E(W^{(i)}_n) \leq P_e,
\]  

(86)

implies

\[
R < I(W) - \frac{\alpha}{N^2}.
\]  

(87)

Here, \( \mu \) is a universal parameter equal to \( \mu = 3.579 \).

A few comments are in order:

(i) The value of \( \mu \) stated in Theorem [16] (i.e. \( \mu = 3.579 \)) can be slightly improved by the following procedure. As we will see shortly, we can obtain an increasing sequence of candidates, call this sequence \( \{\mu_m\}_{m \in \mathbb{N}} \), for the universal parameter \( \mu \) in (87). For each \( m \), in order to show the validity of \( \mu_m \), we need to verify the concavity of a certain polynomial on \([0,1] \) (the polynomial is defined in (55) with \( f(z) = z(1-z) \)). We explained in Remark [14] how we can accomplish this using the Sturm chain method. The value of \( \mu \) stated in Theorem [16] is the one corresponding to \( m = 10 \), an arbitrary choice. If we increase \( m \), we get a new candidate for \( \mu \) to plug into (87), i.e., \( \mu_{16} = 3.614 \). We conjecture that the sequence \( \mu_m \) converges to \( \mu_{\infty} = 3.627 \), the parameter for the BEC. If such a conjecture holds, then the channel BEC polarizes the fastest among the BMS channels (see Question [2]).

(ii) Let \( P_e, \alpha, \mu \) be as in Theorem [16] If we require the block-error probability to be less than \( P_e \) (in the sense that the condition (86) is fulfilled), then the block-length \( N \) should be at least

\[
N > \left( \frac{\alpha}{I(W) - R} \right)^{\mu}.
\]  

(88)

(iii) From (11) we know that the value of \( \mu \) for the random linear ensemble is \( \mu = 2 \), which is the optimal value since the variations of the channel itself require \( \mu \geq 2 \). Thus, given a rate \( R \), reliable transmission by polar codes requires a larger block-length than the optimal value.

Proof of Theorem [17] To fit the bounds of Section [IV-A1] into the framework of Theorem [16] let us first introduce the sequence \( \{\mu_m\}_{m \in \mathbb{N}} \) as

\[
\mu_m = -\frac{1}{\log a_m}.
\]  

(89)
where $a_m$ is defined in (85) with starting function $f(z) = z(1-z)$. From Lemma [15] we know that for a suitable $m$, the speed with which the quantity $E[H_n(1-H_n)]$ decays is lower bounded by $a_m = 2 \frac{\gamma^2}{4 \nu_m}$. More precisely, for $n \geq m$ we have $E[H_n(1-H_n)] \geq 2 \frac{\gamma^2}{4 \nu_m} f_m(H(W))$. To relate the strong reliability condition in (86) to the rate bound in (87), we need the following lemma.

**Lemma 17:** Consider a BMS channel $W$ and assume that there exist positive real numbers $\gamma, \theta$, and $m \in \mathbb{N}$ such that $E[H_n(1-H_n)] \geq \gamma^2 2^{-\theta n}$ for $n \geq m$. Let $\alpha, \beta \geq 0$ be such that $2\alpha + \beta = \gamma$, we have for $n \geq m$

$$\Pr(H_n \leq \alpha 2^{-n\theta}) \leq I(W) - \beta 2^{-n\theta}. \quad (90)$$

The proof of this lemma is provided in the appendices. Let us now use the result of Lemma [17] to conclude the proof of Theorem [16]. By Lemma [15] we have for $n \geq m$

$$E[H_n(1-H_n)] \geq 2 \frac{\gamma^2}{4 \nu_m} f_m(H(W)). \quad (92)$$

Thus, if we now let $\gamma = 2 \frac{\gamma^2}{4 \nu_m} f_m(H(W)), \theta = \frac{1}{\mu_m}$, and $2\alpha = \beta = \frac{\gamma}{2}$, then by using Lemma [17] we obtain

$$\Pr(H_n \leq \frac{\gamma^2}{4 2^{-\nu_m}}) \leq I(W) - \frac{\gamma}{2} 2^{-\nu_m}. \quad (91)$$

Assume that we desire to achieve a rate $R$ equal to

$$R = I(W) - \frac{\gamma^2}{4 2^{-\nu_m}}. \quad (93)$$

Let $\mathcal{I}_{N,R}$ be the set of indices chosen for such a rate $R$, i.e., $\mathcal{I}_{N,R}$ includes the $2^n R$ indices of the sub-channels with the least value of error probability. Define the set $A$ as

$$A = \{ i \in \mathcal{I}_{N,R} : H(W(i)_N) \geq \frac{\gamma}{4} 2^{-\nu_m} \}. \quad (94)$$

In this regard, note that (91) and (92) imply that

$$|A| \geq \frac{\gamma^2}{4} 2^n (1 - \frac{1}{\mu_m}). \quad (95)$$

As a result, by using (5) and (6) we obtain for $n \geq m$

$$\sum_{i \in \mathcal{I}_{N,R}} E(W(i)_N^2) \geq \sum_{i \in A} E(W(i)_N) \geq \sum_{i \in A} h_2^2(H(W(i)_N)) \geq \sum_{i \in A} h_2^2(H(W(i)_N)^2) \geq \sum_{i \in A} h_2^2(\frac{\gamma}{4} 2^{-\nu_m}) \geq |A| \frac{\gamma^2}{4} 2^{-\nu_m} \geq \frac{\gamma^2}{16} \frac{2^n (1 - \frac{1}{\mu_m}) h_2^2(\frac{\gamma}{4} 2^{-\nu_m})}{\mu_m^2 + \log \frac{2}{\gamma}} \geq \frac{\gamma^2}{16} \frac{2^n (1 - \frac{1}{\mu_m}) h_2^2(\frac{\gamma}{4} 2^{-\nu_m})}{\mu_m^2 + \log \frac{2}{\gamma}}, \quad (98)$$

where the last step follows from the fact that for $x \in [0, \frac{1}{2}]$, we have $h_2^2(x) \geq 2^{x(1-x)} x^{x} e^{-2x}$. Thus, having a block-length $N = 2^n$, in order to have error probability (measured by (25)) less than $\frac{\gamma^2}{16} \frac{2^n (1 - \frac{1}{\mu_m}) h_2^2(\frac{\gamma}{4} 2^{-\nu_m})}{\mu_m^2 + \log \frac{2}{\gamma}}$, the rate can be at most $I(W) - \frac{\gamma}{2} 2^{-\nu_m}$. Finally, if we let $m = 10$ (by the discussion in Remark [14] we know that $m = 10$ is suitable), then $\mu_{10} = \frac{1}{-\log(\alpha_{10})} = 3.579$ and choosing

$$P_e = \inf_{n \in \mathbb{N}} \left\{ \sum_{i \in \mathcal{I}_{N,R}} E(W(i)_N) \right\}, \quad (100)$$

where $R$ is given in (92), then it is easy to see from (99) that $P_e > 0$ (since $\frac{1}{\mu_{10}} < \frac{1}{2}$). In other words, from the definition of $P_e$ in (100), we see that $P_e$ is the infimum of a sequence of numbers. Each member of this sequence is lower bounded in (99). However, it is easy to see that this lower bound (and hence the sequence) diverges in $n$ (note that $\frac{1}{\mu_{10}} < \frac{1}{2}$). As a result, the value of $P_e$, which is defined as the infimum of this sequence, is strictly positive, i.e., $P_e > 0$. Furthermore, from (100), it is easy to see that to have the value of the sum $\sum_{i \in \mathcal{I}_{N,R}} E(W(i)_N)$ to be less than $P_e$, the rate should be less than $R$ given in (92).

**2) Universal Upper Bounds:** In this part, we provide upper bounds on the block-length $N$ for polar codes, in terms of the rate $R$, that is required to obtain an error probability less than a given value $P_e$ (see Question [4] in Section II). Again, the key component here is the upper-bounds on the speed of polarization, e.g. the bounds derived in Table III for the BEC and the universal bound (82).

**Theorem 18:** Let $Z_n = Z(W_n)$ be the Bhattacharyya process associated to a BMS channel $W$. Assume that for $n \in \mathbb{N}$ we have

$$E[(Z_n(1 - Z_n))^\alpha] \leq \beta^2 2^{-\rho m}, \quad (101)$$

where $\alpha, \beta, \rho$ are positive constants and $\alpha < 1$. Then, the block-length $N$ required to achieve an error probability $P_e > 0$ at a given rate $R < I(W)$ is bounded from above by

$$\log N \leq (1 + \frac{1}{\rho}) \log \frac{1}{d} + c(\log(\frac{4}{d}))^2, \quad (102)$$

where $d = I(W) - R$ and $c$ is a universal positive constant that depends on $\alpha, \beta, \rho, P_e$.

Before proceeding with the proof of Theorem [18], let us note a few comments:

(i) In the previous sections we have computed several candidates for the value $\rho$ required in Theorem [18]. As an example, using the universal candidate for $\rho$ given in (82) (i.e., $\rho = 0.202$), we obtain the following corollary.

**Corollary 19:** For any BMS channel $W$, the block-length $N$ required to achieve a rate $R < I(W)$ scales at most as

$$N \leq \Theta\left( \frac{1}{(I(W) - R)^\alpha} \right). \quad (103)$$

One important consequence of this corollary is that polar codes require a block-length that scales polynomially in terms of the reciprocal of gap to capacity.

(ii) As we will see in the proof of Theorem [18] the result of this theorem is also valid if we replace $P_e$ with the sum of Bhattacharyya values of the channels that correspond to the good indices (this sum is indeed an upper bound for $P_e$).

18The fact that polar codes need a polynomial block-length in terms of the reciprocal of the gap to capacity is also proven in the recent independently-derived result of [81]...
Proof of Theorem 18: Throughout the proof we will be using two key lemmas (Lemma 21 and Lemma 22) that are stated in the appendices. Let
\[ d = I(W) - R. \]  
(104)

We define \( n_0 \in \mathbb{N} \) to be
\[ n_0 = \left[ \frac{1}{\rho} \log \frac{3(1 + c_1)(1 + 2c_2c_3\beta)}{d} \right]. \]  
(105)

where \( \beta \) is given in (101) and the constants \( c_1, c_2 \) and \( c_3 \) are given in Lemmas 21, 22 and 23 respectively. As a result of Lemma 21 and (105), we have for \( n \geq n_0 \)
\[ \Pr(Z_n \leq \frac{1}{2}) \geq I(W) - c_12^{-n\rho} \]
\[ \geq I(W) - \frac{d}{3} \]
\[ = R + \frac{2}{3}d, \]  
(106)

where step (a) is a consequence of (105) that for \( n \geq n_0 \) we have \( c_12^{-n\rho} \leq \frac{d}{3} \). We now define the set \( \mathcal{A} \) as follows. Let \( N_0 = 2^{n_0} \) and
\[ \mathcal{A} = \{i \in \{0, \ldots, N_0 - 1\} : Z(W^{(i)}_{N_0}) \leq \frac{1}{2}\}. \]  
(107)

In other words \( \mathcal{A} \) is the set of indices at level \( n_0 \) of the corresponding infinite binary tree of \( W \) (see Section I-C whose Bhattacharyya parameter is not so large. Also, from (106) the set \( \mathcal{A} \) contains more than a fraction \( R \) of all the sub-channels at level \( n_0 \). The idea is then to go further down through the infinite binary tree at a level \( n_0 + n_1 \) (the value of \( n_1 \) will be specified shortly). We then observe that the sub-channels at level \( n_0 + n_1 \) that are branched out from the set \( \mathcal{A} \) are polarized to a great extent in the sense that sum of their Bhattacharyya parameters is below \( P_e \) (see Figure 7 for a schematic illustration of the idea).

We proceed by finding a suitable candidate for \( n_1 \). Our objective is to choose \( n_1 \) large enough s.t. there is a set of indices at level \( n_0 + n_1 \) with the following properties: (i) sum of the Bhattacharyya parameters of the sub-channels in this set is less than \( P_e \) and (ii) the cardinality of this set is at least \( 2^{n_0 + n_1} \). In what follows, we will first use the hypothesis of Lemma 22 to give a candidate for \( n_1 \) and then we make it clear that such a candidate is suitable for our needs. Let \( \{B_m\}_{m \in \mathbb{N}} \) be a sequence of iid Bernoulli(\( \frac{1}{2} \)) random variables. We let \( n_1 \) be the smallest positive integer such that the following holds
\[ \Pr(2^{-2^{n_1+1}}) \leq \frac{P_e}{2^{n_0 + n_1}} \geq 1 - \frac{d}{6}. \]  
(108)

It is easy to see that (108) is equivalent to
\[ \Pr(\sum_{i=1}^{n_1} B_i \geq \log(\frac{1}{P_e} + n_0 + n_1)) \geq 1 - \frac{d}{6}. \]  
(109)

Now, note that we can write
\[ \log(\frac{1}{P_e} + n_0 + n_1) \]
\[ = \log(1 + \log(\frac{1}{P_e} + 1 + n_0 + n_1 - 2)) \]
\[ \leq \log(1 + \log(\frac{1}{P_e})) + \log(1 + n_0 + n_1 - 2) \]
\[ \leq \log(\frac{2}{P_e}) + \log(n_0 + n_1), \]  
(110)

where (a) follows from the fact that the function \( f(x) = \log(1 + x) \) is a concave function with \( f(0) = 0 \), and for any such function the following is true: \( f(x + y) \leq f(x) + f(y), \forall x, y \geq 0 \). As a result of (109) and (110), in order for (108) to hold the following is sufficient:
\[ \Pr(\sum_{i=1}^{n_1} B_i \geq \log(\frac{2}{P_e}) + \log(n_0 + n_1)) \geq 1 - \frac{d}{6}. \]  
(111)

Also, as the random variables \( B_i \) are Bernoulli(\( \frac{1}{2} \)) and iid, the relation (111) is equivalent to
\[ \frac{\log(\log(\frac{2}{P_e}) + \log(n_0 + n_1))}{2^{n_1}} \leq \frac{d}{6}. \]  
(112)

A sufficient condition for (112) to hold is as follows:
\[ \frac{n_1 + \log(\frac{2}{P_e}) + \log(n_0 + n_1)}{2^{n_1}} \leq \frac{d}{6}. \]  
(113)

and after applying the function \( \log(\cdot) \) to both sides and some further simplifications we reach to
\[ n_1 + (1 + \log(\log(\frac{2}{P_e}) + \log(n_0 + n_1))) \log n_1 \geq \log \frac{6}{d}. \]  
(114)

It can be shown through some simple steps that there is a constant \( c_6 > 0 \) (that also depends on \( P_e \)) s.t. if we choose
\[ n_1 = \left[ \log \frac{6}{d} + c_6(\log(\log(\frac{6}{d})))^2 \right]. \]  
(115)

then the inequality (113) holds. Now, let \( \tilde{N} = 2^{n_0 + n_1} \) and consider the set \( \mathcal{A}_1 \) defined as
\[ \mathcal{A}_1 = \{i \in \{0, \ldots, \tilde{N} - 1\} : Z(W^{(i)}_{\tilde{N}}) \leq \frac{P_e}{\tilde{N}}\}. \]  
(115)
We now show that
\[
\frac{|A_1|}{N} \geq R. \tag{116}
\]
This relation together with (115) shows that block error probability of the polar code of block-length \(\hat{N}\) and rate \(R\) is at most \(P_e\).

In order to show (116), we consider the sub-channels in \(A_1\) that are branched out from the ones in the set \(A\) (defined in (107)). Let \(i \in A\) and consider the sub-channel \(W^{(i)}_{N_0}\). At level \(n_0 + n_1\) there are in total \(2^{n_1}\) sub-channels that branch out from the sub-channel \(W^{(i)}_{N_0}\) (which is itself at level \(n_0\)). By using (75) it is easy to see that the process \(Z_n\) fulfills the condition (161) of Lemma 22. From Lemma 22 relation (108), and the fact that for any two events \(A\) and \(B\) we have \(\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1\), we obtain the following: At level \(n_0 + n_1\), there are in total \(2^{n_1}\) sub-channels that are branched out from \(W^{(i)}_{N_0}\), and among these sub-channels, a fraction at least
\[
1 - \frac{d}{6} - c_2 Z(W^{(i)}_{N_0}) (1 + \log \frac{1}{Z(W^{(i)}_{N_0})})
\]
have Bhattacharyya value less than \(\frac{\mu}{N}\). Therefore, the number of channels at level \(n_0 + n_1\) that are branched out from \(W^{(i)}_{N_0}\) and have Bhattacharyya value less than \(\frac{\mu}{N}\) is at least
\[
2^{n_1} \left(1 - \frac{d}{6} - c_2 Z(W^{(i)}_{N_0}) (1 + \log \frac{1}{Z(W^{(i)}_{N_0})})\right).
\]
Hence, the total number of sub-channels at level \(n_0 + n_1\) that are branched out from a sub-channel in \(A\) and have Bhattacharyya value less than \(\frac{\mu}{N}\) is at least
\[
2^{n_1} \sum_{i \in A} \left(1 - \frac{d}{6} - c_2 Z(W^{(i)}_{N_0}) (1 + \log \frac{1}{Z(W^{(i)}_{N_0})})\right). \tag{117}
\]
We can further write
\[
\frac{|A|}{N} - c_2 2^{n_1} \sum_{i \in A} Z(W^{(i)}_{N_0}) (1 + \log \frac{1}{Z(W^{(i)}_{N_0})}).
\]
Now, by using (106) and (107) we have \(|A| \geq 2^{n_0} (R + \frac{2}{3}d)\), and hence (117) can be lower bounded by
\[
2^{n_0+n_1} \left((R + \frac{2}{3}d) (1 - \frac{d}{6}) - c_2 2^{-n_0} \sum_{i \in A} Z(W^{(i)}_{N_0}) (1 + \log \frac{1}{Z(W^{(i)}_{N_0})})\right).\tag{118}
\]
We further have
\[
c_2 2^{-n_0} \sum_{i \in A} Z(W^{(i)}_{N_0}) (1 + \log \frac{1}{Z(W^{(i)}_{N_0})}) \\
\leq 2 c_2 2^{-n_0} \sum_{i \in A} Z(W^{(i)}_{N_0}) \log \frac{1}{Z(W^{(i)}_{N_0})} \\
\leq 2 c_2 c_3 2^{-n_0} \sum_{i \in A} (Z(W^{(i)}_{N_0})(1 - Z(W^{(i)}_{N_0})))^a \\
\leq 2 c_2 c_3 (Z_{n_0} - Z_{n_0})^a. \tag{110}
\]
\[
\leq 2 c_2 c_3 \beta^2 2^{-n_0} \tag{109}
\]
\[
\leq \frac{d}{3},
\]
where (a) follows from the fact that for \(x \leq \frac{1}{2}\) we have \(1 + \log \frac{1}{x} \leq 2 \log \frac{1}{x}\). Therefore, the expression (118) (and hence (117)) is lower-bounded by
\[
2^{n_0+n_1} \left((R + \frac{2}{3}d) (1 - \frac{d}{6}) - \frac{d}{3}\right) \geq 2^{n_0+n_1} R = \hat{N} R.
\]
Hence, the relation (116) is proved and a block-length of size \(\hat{N}\) is sufficient to achieve a rate \(R\) and error at most \(P_e\). It is now easy to see that to log \(\hat{N} = n_0 + n_1\) has the form of (102).

V. Conclusion

Let us briefly summarize our main results and discuss some interesting avenues for future research.

We have considered the tradeoff between the rate and the block-length for a fixed error probability when we use polar codes and the successive cancellation (SC) decoder. For a BMS channel \(W\), consider the setting where we require the error probability (measured by the sum of the Bhattacharyya parameters) to be a fixed value \(P_e > 0\). We have shown that in this setting the block-length \(\hat{N}\) scales in terms of the rate \(R < I(W)\) as \(N \geq \frac{\alpha}{I(W) - R \hat{N}}\), where \(\alpha\) is a positive constant that depends on \(P_e\) and \(I(W)\), and \(\hat{N} = 3\). In other words, the required block-length \(N\) is at least \(\Theta\left(\frac{1}{I(W) - R \hat{N}}\right)\). A comparison with (10) indicates that polar codes require a larger block-length compared to the best possible codes (for which \(\mu = 2\)). This provides an analytical explanation for the rather long blocklengths which are required in numerical experiments involving polar codes.

In the same setting, we have also derived an upper bound on the required blocklength by showing that \(N \leq \frac{\beta}{I(W) - R \hat{N}}\), where \(\beta\) is a constant that depends on \(P_e\) and \(I(W)\), and \(\hat{N} = 6\). In other words, the required block-length is at most \(\Theta\left(\frac{1}{I(W) - R \hat{N}}\right)\).

We conjecture that the value of \(\mu\) can be increased up to \(\mu = 3.627\) (the corresponding parameter for the BEC). In the same vain, the value of \(\beta\) can be decreased below \(\beta = 6\) by searching for better candidates for the function \(g(\cdot)\) with a smaller \(L_g\) (see (76)). Indeed, in a follow up work [23], such functions are constructed by carefully evolving a suitable sequence of candidates \(g_m(\cdot)\) through the various polarization levels \(m\). In this way, a new scaling bound with \(\beta = 5.77\) is obtained.

In view of our results, perhaps the most important open question, both from the theoretical as well as the practical side, is to improve the finite-length performance of these codes. We can approach this problem from two perspectives: (i) by devising better decoding algorithms and (ii) by changing the construction of polar codes (e.g., by concatenating them with other codes, use other polarizing kernels, etc.). In any attempt to improve the finite-length performance, one main objective should be to improve the scaling exponent (or the speed of polarization).

In [23], the authors combine both of these perspectives and provide experimenetal evidence that the short-length performance of polar codes can be improved considerably. More precisely, a successive-cancellation list decoder (SCL) is proposed in [23] to boost the performance of the SC decoder to that of the MAP decoder. However, even under MAP decoding the performance of polar codes is still not competitive. Hence,
by a simple concatenation with a very high-rate code, the MAP performance is improved to a great extent. The main issue of the successive cancellation list decoder is its memory consumption which scales linearly with the list-size. There are by now various other techniques to improve the finite-length performance of polar codes. For a partial list see [24]-[29]. It is also an interesting open question to find out how the scaling exponent of the coding method of [23] changes with the list-size parameter. For a fixed finite list-size, it is proven in [30] that the scaling exponent does not change compared to original polar codes when we use the MAP decoder. We believe that the methods developed in this paper can be useful in this regard.

Another approach is to consider polar codes with general $\ell \times \ell$ kernels with the hope that polar codes with larger kernels might have a better finite-length behavior. The related discussions in [21] Chapter 1 support the fact that when $\ell$ grows large, for almost any kernel, the scaling exponent ($\mu$) of the associated polar code tends to $\frac{1}{2}$. Recall from [1] that the optimal value of $\mu$ over all the codes is $\frac{1}{2}$, and for polar codes (with $\ell = 2$) the scaling exponent is at most $\mu = \frac{1}{2} = 0.27$. We keep in mind that, in general, the decoding complexity of (extended) polar codes is $O(2^N \log N)$, where $N$ is the block-length. An interesting question here is to find suitable $\ell \times \ell$ kernels with a better scaling exponent than the $\ell = 2$ case, as well as a reasonable complexity.

Finally, let us note that all these scaling results are in principle extendable to further applications of polarization theory and polar codes in various other scenarios (see e.g. [22]).

ACKNOWLEDGMENT

The authors wish to thank Erdal Arıkan, Alexander Barg, Mani Bastani-Parizi, Ali Goli, Marco Mondelli, and Emre Telatar for their valuable comments on this topic.

REFERENCES

[1] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Transactions on Information Theory, vol. 55 (7), pp. 3051–3073, 2009.
[2] R.G. Gallager, “A simple derivation of the coding theorem and some applications”, IEEE Transactions on Information Theory, vol. 11, no. 1, pp. 3-18, 1965.
[3] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in proc. ISIT, Seoul, South Korea, pp.1493-1495, 2009.
[4] S. H. Hassani, R. Mori, T. Tanaka and R. Urbanke, “Rate dependent analysis of the asymptotic behavior of channel polarization”, IEEE Transactions on Information Theory, vol. 59 (4) pp. 2267–2276, 2013.
[5] 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, pp. 211-252, 1961.
[6] V. Strassen, “Asymptotische abschätzungn in Shannons informations-theorie”, in Trans. 3d Prague Conf. Inf. Theory, Prague, pp. 689-723, 1962.
[7] Y. Polyanskiy, H. V. Poor, and S. Verdu, “A channel coding rate in the finite block-length regime”, IEEE Transactions on Information Theory, 56 (5), pp. 2307-2359, 2010.
[8] S. B. Korada, “Polar codes for channel and source coding,” Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2009.
[9] R. Mori and T. Tanaka, “Performance and construction of polar codes on symmetric binary-input memoryless channels”, in proc. ISIT, Seoul, South Korea, pp.1496-1500, 2009.
[10] S. H. Hassani, S. B. Korada and R. Urbanke, “The compound capacity of polar codes”, in proc. 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp.16-21, 2009.
[11] S. B. Korada, A. Montanari, E. Telatar and R. Urbanke, “An empirical scaling law for polar codes”, in proc. ISIT, Texas, USA, pp.884-888, 2010.
[12] S. H. Hassani, K. Alishahi and R. Urbanke, “On the scaling of polar codes: II. The behavior of un-polarized channels”, in proc. ISIT, Texas, USA, pp.879-883, 2010.
[13] A. Goli, S. H. Hassani, and R. Urbanke, “Universal bounds on the scaling behavior of polar codes”, in proc. ISIT. Boston, USA, pp.1957-1961, 2012.
[14] T. Richardson and R. Urbanke, Modern Coding Theory. Cambridge University Press, 2008.
[15] Chris Godsil and Gordon Royle, Algebraic Graph Theory. Springer, 2001.
[16] I. Tal and A. Vardy, “How to construct polar codes,” IEEE Transactions on Information Theory, 59 (10), pp. 6562–6582, 2013.
[17] R. Pedarsani, H. Hassani, I. Tal and E. Telatar, “On the construction of polar codes,” in proc. ISIT, St. Petersburg, Russia, pp. 11–15, 2011.
[18] Q. I. Rahman, G. Schmeiser, “Analytic theory of polynomials,” London Mathematical Society Monographs, new Series 26. Oxford: Oxford University Press, 2002.
[19] M. Bastani Parizi and E. Telatar, “On Correlation Between Polarized BECs,” [online] Available: arXiv:1301.5536 [cs.IT].
[20] V. Gurusswami and P. Xia, “Polar codes: Speed of polarization and polynomial gap to capacity,” in Electronic Colloquium on Computational Complexity, Report No. 50, 2013.
[21] S. H. Hassani, “Polarization and spatial coupling: Two techniques to boost performance,” Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2013.
[22] D. Goldin and D. Burshtein, “Improved Bounds on the Finite Length Scaling of Polar Codes,” [online] Available: arXiv:1307.5510 [cs.IT].
[23] I. Tal and A. Vardy, “List decoding of polar codes,” [online] Available: arXiv:1206.0050 [cs.IT].
[24] M. Seidl and J. B. Huber, “Improving successive cancellation decoding of polar codes by usage of inner block codes”, in proc. of International Symposium on Turbo Codes and Iterative Information Processing, France, 2010.
[25] B. Li, H. Shen, and D. Tse, “An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check”, in IEEE Comm. Letters, vol. 16, pp. 2044–2047, 2012.
[26] H. Mahdavifar, M. El-Khamy, J. Lee, and J. Kang, “Performance limits and practical decoding of interleaved Reed-Solomon polar concatenated codes”, [online] Available: arXiv:1308.1144 [cs.IT].
[27] P. Trifonov and V. Miloslavskaya, “Polar codes with dynamic frozen symbols and their decoding by directed search”, [online] Available: arXiv:1307.2352 [cs.IT].
[28] K. Niu, K. Chen, and J. Lin, “Improved successive cancellation decoding of polar codes,” in IEEE Transactions on Communications, 61(8)(16), pp. 3100–3107, 2013.
[29] M. Mondelli, S. H. Hassani, and R. Urbanke, “From polar to Reed-Muller codes: A technique to improve the finite-length performance”, [online] Available: arXiv:1304.5220 [cs.IT].
[30] M. Mondelli, S. H. Hassani, and R. Urbanke, “Scaling exponent of list decoders with applications to polar codes”, [online] Available: arXiv:1301.3127 [cs.IT].

APPENDIX A

PROOFS

1) Proof of Lemma 6. The proof of (64) is an easy application of the Markov inequality: We have

\[ \Pr(Z_n \in [a, b]) \leq \Pr(f(Z_n) \geq \min_{z \in [a,b]} f(z)) \leq \frac{\mathbb{E}[f(Z_n)]}{\min_{z \in [a,b]} f(z)} \]

and (64) follows by applying \( \frac{1}{\ell} \log(\cdot) \) to (119).

To prove (63), we define sequences \( \{x_n\}_{n \geq 1} \) and \( \{y_n\}_{n \geq 1} \) as

\[ x_n = 2^{-n}, \]

(120)
We show that in each of these cases the statement of the lemma holds. The proof consists of two parts:

Continuing this way, we can show that for

\[ n \in \mathbb{N} \]

As a result, there exists an index \( j \in \{1, \ldots, n\} \) such that at least one of the following cases occurs:

\[ \Pr(Z_n \in [y_j, y_{j+1}) \mid Z_0 = z) = \Pr(Z_n \in [x_{j+1}, x_j) \mid Z_0 = 1 - z), \]

Hence, without loss of generality we can assume that the choice of \( \{a, b\} \)

The proof consists of two parts:

We first assume that \( a = 1 - b = \frac{1}{4} \) and prove (123) for this choice of \( a, b \).

**Lemma 20:** For any \( j \in \{1, \ldots, n\} \) we have

\[ 2^{-2j}\Pr(Z_n \in [x_{j+1}, x_j]) \leq (n+1)\Pr(Z_n \in [\frac{1}{4}, \frac{3}{4}]) + 2^{-n}, \]

The proof of this lemma will appear shortly. But before that, we note that by using the result of this lemma and (122) we obtain

\[ \Pr(Z_n \in [1 - Z_n]) \leq 2n(n+1)[\Pr(Z_n \in [\frac{1}{4}, \frac{3}{4}]) + (2n+1)2^{-n}, \]

and as a result, by taking \( \frac{1}{n} \log(\cdot) \) from both sides, (63) is proved for \( a = 1 - b = \frac{1}{4} \).

Now, for other choices of \( a, b \in (0, 1) \) s.t. \( \sqrt{a} \leq 1 - \sqrt{1 - b} \) we can proceed as follows. Let us first recall the definition of the maps \( t_0, t_1 \) from (68) as well as the maps \( \phi_{a,b} \) from Definition 10. Also, let \( p_n(z, a, b) \) be defined as in (23). We have

\[ p_{n+1}(z, a, b) = \sum_{\omega_{n+1} \in \Omega_{n+1}} \frac{1}{2^{m+1}} \mathbb{I}_{\{z \in \phi_{a,b}(y_n, \omega_{n+1}, [a, b])\}} \]

\[ = \sum_{\omega_{n+1} \in \Omega_{n+1}} \frac{1}{2^{m+1}} \left[ \mathbb{I}_{\{z \in \phi_{a,b}(y_n, t_0^{-1}(a), t_0^{-1}(b))\}} + \mathbb{I}_{\{z \in \phi_{a,b}(y_n, t_1^{-1}(a), t_1^{-1}(b))\}} \right] \]

\[ = \frac{1}{2} \left[ p_n(z, t_0^{-1}(a), t_0^{-1}(b)) + p_n(z, t_1^{-1}(a), t_1^{-1}(b)) \right]. \]

It is easy to see that if \( \sqrt{a} \leq 1 - \sqrt{1 - b} \), then

\[ [t_0^{-1}(a), t_1^{-1}(b)] \subseteq [t_0^{-1}(a), t_0^{-1}(b)] \cup [t_1^{-1}(a), t_1^{-1}(b)], \]

and hence,

\[ 2p_{n+1}(z, a, b) \geq p_n(z, t_0^{-1}(a), t_1^{-1}(b)). \]

Continuing this way, we can show that for \( m \in \mathbb{N} \)

\[ 2^m p_{n+m}(z, a, b) \]

As \( m \) grows large, we have

\[ t_0^{-1} \circ \cdots \circ t_0^{-1}(a) \to 0, \]

\[ t_1^{-1} \circ \cdots \circ t_1^{-1}(b) \to 1. \]

Therefore, by (125) there exists a positive integer \( m_0 \in \mathbb{N} \) that only depends on \( a, b \) and for all \( z \in [0, 1] \)

\[ 2^{m_0} p_{n+m_0}(z, a, b) \geq p_n(z, \frac{1}{4}, \frac{3}{4}). \]

The proof of (63) now follows from this relation together with (124) and the result of Lemma 5. It remains to prove Lemma 20.

**Proof of Lemma 20:** Consider the relation (124) for \( 1 \leq j \leq n \). If \( j = 1 \), then there is nothing to prove. Hence, in the following we assume that \( 2 \leq j \leq n \). We prove that for any fixed \( j \), such that \( 2 \leq j \leq n \), the claim of (124) holds true. So let us fix the index \( j \) and prove (124) for any value of \( n \in \mathbb{N} \). The proof consists of two steps.

**Step 1:** We first show that \( \forall m \in \mathbb{N} \),

\[ \Pr(Z_n \in [x_{2j+1}, x_j]) \leq m \Pr(Z_n \in [x_j, \frac{3}{4}]) + \frac{1}{2^m}. \]

To prove (126), fix \( m \in \mathbb{N} \) and define the sets \( A \) and \( B \) as

\[ A = \{(b_1, \ldots, b_m) \in \Omega_m : t_{b_m} \circ \cdots \circ t_{b_1}(z) \in [x_{2j+1}, x_j]\}, \]

\[ B = \{(b_1, \ldots, b_m) \in \Omega_m : t_{b_m} \circ \cdots \circ t_{b_1}(z) \in [x_j, \frac{3}{4}]\}. \]

In words, \( A \) is the set of all the paths that start from \( z = Z_0 \) and end up in \([x_{2j+1}, x_j]\) and \( B \) is the set of paths that start from \( z \) and end up in \([x_j, \frac{3}{4}]\). Consider the sets \( A_k, k \in \{1, \ldots, m\} \), defined as

\[ A_k = \{(b_1, \ldots, b_m) \in A : b_k = 1; b_i = 0 \ \forall i > k\}. \]

It is easy to see \( A_k \)'s are disjoint and

\[ |A - \cup_k A_k| \leq 1. \]

Our aim is now to show that for \( k \in \{1, \ldots, m\} \),

\[ |A_k| \leq |B|. \]

Before proving (129), let us show how the relation (126) follows from (128) and (129). We have

\[ \Pr(Z_n \in [x_{2j+1}, x_j]) = \frac{|A|}{2^m} \leq \sum_{i=1}^{2^m} |A_i| + 1 \]

\[ \leq m |B| + \frac{1}{2^m} \leq m \Pr(Z_n \in [x_j, \frac{3}{4}]) + \frac{1}{2^m}. \]

It thus remains to prove (129) and Step 1 is over. We show that there exists a one-to-one correspondence between \( A_k \) and a subset of \( B \). In other words, we claim that we can map each
member of $A_k$ to a distinct member of $B$. In this way, the relation (129) is immediate. Consider $(b_1,\ldots,b_m) \in A_k$. We now construct a distinct member $(b'_1,\ldots,b'_m) \in B$ corresponding to $(b_1,\ldots,b_m)$. We first set $b'_i = b_i$ for $i < k$ and hence the uniqueness condition is fulfilled (i.e., the choice of $b'_i = b_i$ for $i < k$ guarantees that the mapping from $A_k$ to $B$ is an invertible mapping). Consider the number $y$ defined as

$$
y = \begin{cases} 
    z & \text{if } k = 1, \\
    t_{b_{k-1}} \cdots t_{b_1}(z) & \text{if } k > 1.
\end{cases}
$$

(130)

Note that since $(b_1,\ldots,b_m) \in A_k$ we have

$$
t_{b_m} \cdots t_{b_1}(y) \in [x_{2j+1}, x_j].
$$

(131)

Now, note that as $(b_1,\ldots,b_m) \in A_k$, we have $b_k = 1$ and $b_i = 0$ for $i > k$. Thus, in this setting (131) becomes

$$
t_0 \cdots t_0(y^2) \in [x_{2j+1}, x_j].
$$

(132)

Hence,

$$x_{2j+1} \leq 1 - (1 - y^2)^{m-k} \leq x_j.
$$

(133)

From the left side of (133) and by using Bernoulli’s inequality

$$1 - (1 - x)^\beta \leq \beta x, \text{ where } \beta \geq 1 \text{ and } x \in [0,1],$$

we obtain

$$x_{2j+1} \leq 2^{m-k}y^2 \Rightarrow 2^{-j} \frac{m-k}{m} \leq y.
$$

(134)

From the right side of (133) we have

$$\ln(1 - x_j) \leq 2^{-k} \ln(1 - y^2),
$$

and by using the inequality

$$-x - \frac{x^2}{2} \leq \ln(1 - x) \leq -x, \text{ where } x \in (0,1),$$

we obtain

$$y \leq 2^{-j} \frac{m-k}{m-m-k}.
$$

(135)

Let us recall that we let $b'_i = b_i$ for $i < k$ (and this makes the mapping from $A_k$ to $B$ an invertible mapping). We now construct the remaining values $b'_k,\ldots,b'_m$ by the following algorithm: Consider the number $y$ given in (130). In the following, we will also construct a sequence $y = y_{k-1}, y_k, y_{k+1},\ldots,y_m$ such that for $i \geq k$ we have $y_i = t_{b'_i}(y_{i-1})$. Begin with the initial value $y_{k-1} = y$ and for $i \geq k$ recursively construct $b'_i$ and $y_i$ from $y_{i-1}$ by the following rule: if $t_0(y_{i-1}) \leq \frac{3}{4}$, then $b'_i = 0$ and $y_i = t_0(y_{i-1})$, otherwise $b'_i = 1$ and $y_i = t_1(y_{i-1})$. We show that by this construction the value of $y_m$ would always fall in the interval $[x_j, \frac{3}{4}]$. This in regard, an important observation is that for $i$ s.t. $k - 1 \leq i \leq m$, once the value of $y_i$ lies in the interval $[x_j, \frac{3}{4}]$, then for all $i \leq t \leq m$ we have $y_t \in [x_j, \frac{3}{4}]$ (this is clear from construction rule of $y_i$). Hence, we only need to show that by the above algorithm, the exists an index $i$, s.t. $k - 1 \leq i \leq m$, and the value of $y_i$ lies inside the interval $[x_j, \frac{3}{4}]$. On the one hand, observe that due to (135) and the fact that $j \geq 2$, we have $y \leq 2^{-\frac{3}{2}} < \frac{3}{4}$. Thus, the value of $y_i$ is definitely less than $\frac{3}{4}$ for $i \geq k$. If the value of $y_{k-1}$ is also greater than $x_j$ then we have nothing to prove. Else, it might be the case that $y < x_j$. We now prove that in this case the algorithm moves in a way that the value of $y_m$ falls eventually in the desired region $[x_j, \frac{3}{4}]$. To show this, a moment of thought reveals that this is equivalent to showing that we always have

$$2^{m-k+1} y \leq 1 - (1 - y^2)^{2^{m-k}} \geq x_j.
$$

(136)

In order to have (136) it is equivalent that

$$2^{-j} \frac{m-k}{m} \leq \ln(1 - x_j),$$

and after some further simplification using the inequality $-x - \frac{x^2}{2} \leq \ln(1 - x) \leq -x$, we deduce that a sufficient condition to have (136) is

$$x_j \leq 2^{m-k}y \Rightarrow 2^{-j} \frac{m-k}{m} \leq y.
$$

(137)

But this sufficient condition is certainly met by considering the inequality (134) and noting the fact that $-j \frac{m-k}{m} \geq -j + k - m$ (recall that $k \leq m$). Hence, the claim in (129) is proved and as a result, the claim in (126) is true.

**Step 2:** Firstly note that in order for $Z_n$ to be in the interval $[x_{j+1}, x_j]$, the value of $Z_{n-j}$ should certainly lie somewhere in the interval $[x_{2j+1}, x_{2j+2}]$. As a result, we can write

$$\Pr(Z_n \in [x_{j+1}, x_j]) = \Pr(Z_n \in [x_{j+1}, x_j] \mid Z_{n-j} \in [x_{2j+1}, x_j]) \times \Pr(Z_{n-j} \in [x_{2j+1}, x_j])$$

$$+ \Pr(Z_n \in [x_{j+1}, x_j] \mid Z_{n-j} \in (x_j, x_{2j+2}^2]) \times \Pr(Z_{n-j} \in (x_j, x_{2j+2}^2]),
$$

(138)

and by letting $m = n - j$ in relation (126), we can easily obtain

$$\Pr(Z_{n-j} \in [x_{2j+1}, x_j]) \leq n \Pr(Z_{n-j} \in [x_j, \frac{3}{4}]) + \frac{1}{2n-j}.
$$

(139)

Thus, by combining (138) and (139), we obtain

$$\Pr(Z_n \in [x_{j+1}, x_j]) \leq n \Pr(Z_{n-j} \in [x_j, \frac{3}{4}]) + \Pr(Z_{n-j} \in [x_j, x_{2j+2}^2]) + \frac{1}{2n-j}.
$$

(140)

Finally, in order to conclude the proof of (124) (for $j \in \{2,\ldots,n\}$), we prove the following relations:

$$2^{-j} \Pr(Z_{n-j} \in [x_j, \frac{3}{4}]) \leq \Pr(Z_n \in [\frac{1}{4}, \frac{3}{4}]).
$$

(141)

and

$$2^{-j} \Pr(Z_{n-j} \in [x_j, x_{2j+2}^2]) \leq \Pr(Z_n \in [\frac{1}{4}, \frac{3}{4}]).
$$

(142)

It is easy to see that these two relations combined with (140) will result in (124). Firstly, note that for $j = 2$ the relations (141) and (142) are trivial. Also, for $j \geq 3$ because of the fact that $x_{2j+2}^2 \geq \frac{3}{4}$, then (141) will be a direct consequence of (142), and hence it is enough to prove (142).

To prove (142), we show that

$$\Pr(Z_n \in [\frac{1}{4}, \frac{3}{4}] \mid Z_{n-j} \in [x_j, x_{2j+2}^2]) \geq 2^{-j},
$$

(143)

and from this we can conclude (142) by writing

$$\Pr(Z_n \in [\frac{1}{4}, \frac{3}{4}]).$$
It thus remains to show (143). A moment of thought reveals that (143) is an immediate consequence of the following statement: For any value $y$ s.t. $y \in [x_j, x_j^{2^{-j}}]$, there exists a $j$-tuple $(b_1, ..., b_j) \in \Omega_j$ such that $t_{b_1} \circ \cdots \circ t_{b_j} (y) \in [t_0, y_1]$. We show this last statement by constructing the binary values $b_1, ..., b_j$ in terms of $y$ (we use a similar approach as in Step 1). Consider the following algorithm: start with $y_0 = y$ and for $1 \leq i \leq j$, we recursively construct $b_i$ from $y_{i-1}$ by the following rule: If $t_0(y_{i-1}) \leq \frac{3}{4}$, then $b_i = 0$ and $y_i = t_0(y_{i-1})$. Otherwise, let $b_i = 1$ and $y_i = t_1(y_{i-1})$. To show that this algorithm succeeds in the sense that $y_j \in \left[\frac{1}{4}, \frac{3}{4}\right]$ we first observe that once the value of $y_i$ lies in the interval $[\frac{1}{4}, \frac{3}{4}]$ (for some $1 \leq i \leq j$), then for all $i \leq f \leq j$ we have $y_f \in [\frac{1}{4}, \frac{3}{4}]$. Hence, we only need to show that by the above algorithm, the exists an index $i$, s.t. $1 \leq i \leq j$, and the value of $y_i$ lies in the interval $[\frac{1}{4}, \frac{3}{4}]$. On the one hand, assume $y \in [x_j, \frac{1}{4}]$. We can then write

\[
\frac{j \text{ times}}{t_0 \circ \cdots \circ t_0 (y)} = 1 - (1 - y)^{2^j} \geq 1 - (1 - x_j)^{2^j} \geq \frac{1}{2},
\]

where the last steps follows from the fact that $x_j = 2^{-j}$. On the other hand, assume $y \in (\frac{3}{4}, x_j^{2^{-j}}]$. We can write

\[
\frac{j \text{ times}}{t_1 \circ \cdots \circ t_1 (y)} = x_j < \frac{3}{4}.
\]

As a result, the above algorithm always succeeds and the lemma is proved for $a = 1 - b = \frac{1}{4}$.

2) Proof of Lemma [17]: Recall that for a realization of a Markov process $Z_n = (b_k)_{k \in \mathbb{N}} \in \Omega$, we define $\phi_n = (b_1, ..., b_n)$. The maps $t_0$ and $t_1$, hence the maps $\phi_n$, are strictly increasing maps on $[0, 1]$. Thus $\phi_n(z) \to 0$ implies that $\phi_n(z') \to 0$ for $z' \leq z$ and $\phi_n(z) \to 1$ implies that $\phi_n(z') \to z'$ for $z' \geq z$. Moreover, we know that for almost every $z \in (0, 1)$, $\lim_{n \to \infty} \phi_n(z)$ is either 0 or 1 for almost every realization $(\phi_n)_{n \in \mathbb{N}}$. Hence, it suffices to let

\[
z^*_n = \inf \{ z : \phi_n(z) \to 1 \}.
\]

To prove the second part of the lemma, notice that

\[
\begin{align*}
z &= \Pr(Z_\infty = 1) \\
 &= \Pr(\phi_n(z) \to 1) \\
 &= \Pr(\inf \{ z : \phi_n(z) \to 1 \} \leq z) \\
 &= \Pr(z^*_n < z).
\end{align*}
\]

Which shows that $z^*_n$ is uniformly distributed on $[0, 1]$.

3) Proof of Lemma [12]: In order to compute $\lim_{n \to \infty} \mathbb{E}[\frac{1}{n} \log(\phi_{\omega_n}^{-1}(b) - \phi_{\omega_n}^{-1}(a))]$, we first define the process $\{Z_n\}_{n \in \mathbb{N}}$ with $Z_0 = z \in [0, 1]$ and

\[
Z_{n+1} = \begin{cases} 
\sqrt{Z_n}, & \text{w.p. } \frac{1}{2}, \\
1 - \sqrt{1 - Z_n}, & \text{w.p. } \frac{1}{2}. 
\end{cases}
\tag{144}
\]

We can think of $\tilde{Z}_n$ as the reverse stochastic process of $Z_n$. Equivalently, we can also define $\tilde{Z}_n$ via the inverse maps $t_0^{-1}, t_1^{-1}$. Consider the sequence of i.i.d. symmetric Bernoulli random variables $B_1, B_2, ..., B_n$ and define $\tilde{Z}_n = \psi_{\omega_n}(z)$ where $\omega_n = (b_1, ..., b_n) \in \Omega_n$ and

\[
\psi_{\omega_n} = t_0^{-1} \circ t_1^{-1} \circ \cdots \circ t_0^{-1}.
\tag{145}
\]

We show that the Lebesgue measure (or the uniform probability measure) on $[0, 1]$, denoted by $\nu$, is the unique, hence ergodic, invariant measure for the Markov process $\tilde{Z}_n$. To prove this result, first note that if $\tilde{Z}_n$ is distributed according to the Lebesgue measure, then

\[
\Pr(\tilde{Z}_{n+1} < x) = \frac{1}{2} \Pr(\tilde{Z}_n < t_0(x)) + \frac{1}{2} \Pr(\tilde{Z}_n < t_1(x)) = \frac{1}{2} x^2 - \frac{1}{2} (2x - x^2) = x.
\]

Thus, $\tilde{Z}_{n+1}$ is also distributed according to the Lebesgue measure and this implies the invariance of the Lebesgue measure for $\tilde{Z}_n$. In order to prove the uniqueness, we will show that for any $z \in (0, 1)$, $\tilde{Z}_n$ converges weakly to a uniformly distributed random point in $[0, 1]$, i.e.,

\[
\tilde{Z}_n = \psi_{\omega_n}(z) \xrightarrow{\text{d}} \nu.
\tag{146}
\]

Note that with (146) the uniqueness of $\nu$ is proved since for any invariant measure $\rho$ assuming $\tilde{Z}_n$ is distributed according to $\rho$, we have

\[
\rho(\cdot) = \Pr(\tilde{Z}_n \in \cdot) = \int \Pr(\tilde{Z}_n \in \cdot, \rho)(dz) \xrightarrow{\text{d}} \nu(\cdot).
\tag{147}
\]

To prove (146), note that $\psi_{\omega_n}$ has the same (probability law as $\phi_{\omega_n}^{-1}$ and we know that $\phi_{\omega_n}^{-1}(z) \to \omega_n^*$ almost surely and hence weakly. Also, $z^*_n$ is distributed according to $\nu$, which proves (146). We are now ready to show that

\[
\lim_{n \to \infty} \mathbb{E}\left[ \frac{1}{n} \log(\phi_{\omega_n}^{-1}(b) - \phi_{\omega_n}^{-1}(a)) \right] = \frac{1}{2 \ln 2}.
\tag{148}
\]

Using the mean-value theorem, we can write

\[
\psi_{\omega_n}(a) - \psi_{\omega_n}(b) = \psi'_{\omega_n}(c)(b - a),
\tag{149}
\]

for some $c \in (a, b)$. And by chain rule,

\[
\psi_{\omega_n}(c) = t_{b_0}^{-1} \circ t_{b_{n-1}}^{-1} \circ \cdots \circ t_{b_1}^{-1}(c)
\]

\[
= t_{b_1}^{-1}(c) \times t_{b_2}^{-1}(t_{b_1}^{-1}(c)) \times \cdots \times t_{b_n}^{-1}(t_{b_{n-1}}^{-1} \circ \cdots \circ t_{b_1}^{-1}(c))
\]

\[
= t_{b_1}^{-1}(\psi_{\omega_0}(c)) \times t_{b_2}^{-1}(\psi_{\omega_1}(c)) \times \cdots \times t_{b_n}^{-1}(\psi_{\omega_{n-1}}(c)),
\]

and after applying $\frac{1}{n} \log(\cdot)$ to both sides we obtain

\[
\frac{1}{n} \log(\psi_{\omega_n}(c)) = \frac{1}{n} \sum_{j=1}^{n} \log t_{b_j}^{-1}(\psi_{\omega_{j-1}}(c)).
\tag{150}
\]

By the ergodic theorem, the last expression converges almost surely to the expectation of $\log t_{b_j}^{-1}(U)$, where $U$ is assumed...
to be distributed according to \( \nu \). Hence, the asymptotic value of (150) can be computed as

\[
E[\log t^{-1}(U)] = \frac{1}{2} \int_{0}^{1} \log(\sqrt{x})' dx + \frac{1}{2} \int_{0}^{1} \log(1 - \sqrt{1 - x})' dx
\]

\[
= \frac{1}{2 \ln 2} - 1. \tag{151}
\]

The proof now follows as a result of (148), (149), (150), and (151).

4) Proof of Lemma 17: The proof is by contradiction. Let us assume the contrary, i.e., we assume there exists \( n \geq m \) s.t.,

\[
\Pr(H_n \leq 2^{-n/2}) > I(W) - \beta 2^{-n/2}. \tag{152}
\]

In the following, we show that with such an assumption we reach to a contradiction. We have

\[
E[H_n(1 - H_n)] = E[H_n(1 - H_n) | H_n \leq 2^{-n/2}] \Pr(H_n \leq 2^{-n/2}) + E[H_n(1 - H_n) | H_n > 2^{-n/2}] \Pr(H_n > 2^{-n/2}). \tag{153}
\]

It is now easy to see that

\[
E[H_n(1 - H_n) | H_n \leq 2^{-n/2}] \leq 2^{-n/2},
\]

and since \( E[H_n(1 - H_n)] \geq 2^{-n/2} \), by using (153) we get

\[
E[H_n(1 - H_n) | H_n > 2^{-n/2}] \Pr(H_n > 2^{-n/2}) \geq 2^{-n/2}(\gamma - \alpha). \tag{154}
\]

We can further write

\[
E[(1 - H_n)] = E[H_n(1 - H_n) | H_n \leq 2^{-n/2}] \Pr(H_n \leq 2^{-n/2}) + E[H_n(1 - H_n) | H_n > 2^{-n/2}] \Pr(H_n > 2^{-n/2}), \tag{155}
\]

and noticing fact that \( 1 - H_n \geq H_n(1 - H_n) \) we can plug (154) in (155) to obtain

\[
E[(1 - H_n)] \geq E[H_n(1 - H_n) | H_n \leq 2^{-n/2}] \Pr(H_n \leq 2^{-n/2}) + 2^{-n/2}(\gamma - \alpha). \tag{156}
\]

We now continue by using (152) in (156) to obtain

\[
E[(1 - H_n)] > (1 - \alpha 2^{-n/2})(I(W) - \beta 2^{-n/2}) + 2^{-n/2}(\gamma - \alpha) \geq I(W) + 2^{-n/2}(\gamma - \alpha(1 + I(W)) - \beta),
\]

and since \( 2\alpha + \beta = \gamma \), we get \( E[1 - H_n] > I(W) \). This is a contradiction since \( H_n \) is a martingale and \( E[1 - H_n] = I(W) \).

**APPENDIX B**

**Auxiliary Lemmas**

**Lemma 21:** Consider a channel \( W \) with its Bhattacharyya process \( Z_n = Z(W_n) \) and assume that for \( n \in \mathbb{N} \)

\[
E[(Z_n(1 - Z_n))^\alpha] \leq \beta 2^{-n/2}, \tag{157}
\]

where \( \alpha, \beta, \rho \) are positive constants with \( \alpha < 1 \). We then have for \( n \in \mathbb{N} \)

\[
\Pr(Z_n \leq \frac{1}{2}) \geq I(W) - c_1 2^{-n/2}, \tag{158}
\]

where \( c_1 \) is a positive constant that depends on \( \alpha, \beta, \rho \).

**Proof:** The proof consists of three steps. First, consider an arbitrary BMS channel \( W \) and let \( Z_n = Z(W_n) \). Also, consider the process \( Y_n = 1 - Z_n^2 \). By using the relations (13) and (14), it can easily be checked that the process \( Y_n \) has the form of (161) and hence Lemma 22 is applicable to \( Y_n \). We thus have from (162) that for \( n \in \mathbb{N} \)

\[
\Pr(Y_n > \frac{1}{2}) \leq c_2 2^{n/2} \tag{159}
\]

As a consequence

\[
I(W) = \lim_{n \to \infty} \Pr(Y_n > \frac{1}{2}) \leq c_2(1 - Z(W)^2)(1 + \log \frac{1}{1 - Z(W)^2}). \tag{159}
\]

In the second step, we consider a channel \( W \) for which (157) holds for \( n \in \mathbb{N} \). By using (157), it is easy to see that for \( n \in \mathbb{N} \)

\[
E[(Z_n^2(1 - Z_n^2))^\alpha] \mathbb{1}_{(z_n > \frac{1}{2})} \leq \sup_{z \in [1, 1]} (z(1 + z))^\alpha E[(Z_n(1 - Z_n))^\alpha] \mathbb{1}_{(z_n > \frac{1}{2})} \leq 2\alpha \beta 2^{-n/2} \leq 2^{1-n/2}. \tag{160}
\]

In the final step, we consider a number \( n \in \mathbb{N} \) and let \( N = 2^n \). We then define the set \( A \) as

\[
A = \{ i \in \{0, 1, \ldots, N - 1\} : Z(W_n^i) \leq \frac{1}{2} \},
\]

with \( A_n \) being its complement. We have

\[
\sum_{i \in A_n} I(W_n^i) \tag{a) \sum_{i \in A_n} c_2(1 - Z(W_n^i)^2)(1 + \log \frac{1}{1 - Z(W_n^i)^2}) \tag{b) \sum_{i \in A_n} 4c_2c_3(Z(W_n^i)^2(1 - Z(W_n^i)^2))^\alpha \tag{c) 4c_2c_3N} \tag{161}
\]

Here (a) follows from (159), (b) follows from Lemma 23 and the fact that for \( x \leq \frac{1}{2} \) we have \( 1 + \log \frac{1}{1 - x} \leq 4 \log \frac{1}{1 - x} \), and (c) follows from (160). Now, as a consequence of the above chain of inequalities we have

\[
|A| \geq \sum_{i \in A} I(W_n^i) \tag{a) NI(W) - \sum_{i \in A} I(W_n^i) \tag{b) N(I(W) - 8c_2c_3\beta 2^{-n/2}),}
\]

and consequently

\[
\Pr(Z_n \leq \frac{1}{2}) = \frac{|A|}{N} \geq I(W) - 8c_2c_3\beta 2^{-n/2}. \tag{161}
\]

Hence, the proof follows by letting \( c_1 = 8c_2c_3\beta \). □

**Lemma 22:** Consider a generic stochastic process \( \{X_n\}_{n \geq 0} \) s.t. \( X_0 = x \), where \( x \in (0, 1) \), and for \( n \geq 1 \)

\[
X_n \leq \begin{cases} 
X_{n-1}^2 & : \text{if } B_n = 1, \\
2X_{n-1} & : \text{if } B_n = 0.
\end{cases} \tag{161}
\]
Here, $\{B_n\}_{n \geq 1}$ is a sequence of iid random variables with distribution $\text{Bernoulli}(\frac{1}{2})$. We then have for $n \in \mathbb{N}$

$$
\Pr(X_n \leq 2^{-2\sum_{i=1}^{n} b_i}) \geq 1 - c_2 x (1 + \log \frac{1}{x}),
$$

where $c_2$ is a positive constant.

**Proof:** We begin by recalling some related notation. Assuming $\{B_n\}_{n \in \mathbb{N}}$ is a sequence of iid $\text{Bernoulli}(\frac{1}{2})$ random variables, we denote by $(\mathcal{F}, \Omega, \Pr)$ the probability space generated by this sequence. We also let $(\mathcal{F}_n, \Omega_n, \Pr_n)$ be the probability space generated by $(B_1, \ldots, B_n)$. Finally, we denote by $\theta_n$ the natural embedding of $\mathcal{F}_n$ into $\mathcal{F}$, i.e., for every $F \in \mathcal{F}_n$

$$
\theta_n(F) = \{ (b_1, b_2, \ldots, b_n, b_{n+1}, \ldots) \in \Omega \mid (b_1, \ldots, b_n) \in F \}.
$$

We thus have $\Pr_n(F) = \Pr(\theta_n(F))$.

We slightly modify $X_n$ to start with $X_0 = x$, where $x \in (0,1)$, and for $n \geq 1$

$$
X_n = \left\{ \begin{array}{ll}
X_{n-1}^2 & \text{if } B_n = 1, \\
2X_{n-1} & \text{if } B_n = 0.
\end{array} \right.
$$

(163)

It is easy to see that if we prove the lemma for this version of $X_n$, then the result of the lemma is valid for any generic $X_n$ that satisfies (161).

Equivalently, we can analyze the process $A_n = -\log X_n$, i.e., $A_0 = -\log x \neq a_0$ and

$$
A_{n+1} = \left\{ \begin{array}{ll}
2A_n & \text{if } B_n = 1, \\
A_n - 1 & \text{if } B_n = 0.
\end{array} \right.
$$

(164)

Note that in terms of the process $A_n$, the statement of the lemma can be phrased as

$$
\Pr(A_n \geq 2^{\sum_{i=1}^{n} b_i}) \geq 1 - c_2 \frac{1 + a_0}{2a_0}.
$$

(165)

Let us first explain how to associate to each $(b_1, \ldots, b_n) \in \omega_n \in \Omega_n$ a sequence of “runs” $(r_1, \ldots, r_{k(\omega_n)})$. This sequence is constructed by the following procedure. Each of the $r_i$’s is a positive integer. We construct the integers $r_i$ one by one starting from $r_1$. We define $r_1$ as the smallest index $i \in \mathbb{N}$ so that $b_{i+1} \neq b_i$. In general, $r_i$ is constructed from the previous $r_i$’s, $1 \leq i < k$, in the following way. If $\sum_{j=1}^{k-1} r_j < n$ then

$$
r_k = \min\{i \mid \sum_{j=1}^{k-1} r_j < i \leq n, b_{i+1} \neq b_{\sum_{j=1}^{k-1} r_j} \} - \sum_{j=1}^{k-1} r_j.
$$

(166)

The process stops whenever the sum of the runs equals $n$ (i.e., whenever $\sum_{i=1}^{k} r_i$ is exactly equal to $n$). Denote the stopping time of the process by $k(\omega_n)$. In words, the sequence $(b_1, \ldots, b_n)$ starts with $b_1$. It then repeats $b_1$, $r_1$ times. Next follow $r_2$ instances of $b_1$ ($b_1 \triangleq 1 - b_1$), followed again by $r_3$ instances of $b_1$, and so on. We see that $b_1$ and $(r_1, \ldots, r_{k(\omega_n)})$ fully describe $\omega_n = (b_1, \ldots, b_n)$. Therefore, there is a one-to-one map

$$
(b_1, \ldots, b_n) \leftrightarrow (b_1, (r_1, \ldots, r_{k(\omega_n)})).
$$

(167)

As an example, for the sequence $\omega_8 = (b_1, b_2, \ldots, b_8) = (1,0,0,1,0,0,0,1)$, we have $k(\omega_8) = 5$, and the corresponding sequence of runs is $(r_1, r_2, r_3, r_4, r_5) = (1,2,1,3,1)$. Also, the knowledge of the sequence $(r_1, r_2, r_3, r_4, r_5)$ and the fact that $b_1 = 1$ will uniquely determine the sequence $(b_1, b_2, \ldots, b_8)$.

We think of $\omega_n = (b_1, \ldots, b_n)$ as a realization of the random vector $(B_1, \ldots, B_n)$. In this regard, each realization $(b_1, \ldots, b_n)$ is associated with a value $k(\omega_n)$ and a run sequence $(r_1, \ldots, r_{k(\omega_n)})$. Thus, $k(\omega_n)$ and $(r_1, \ldots, r_{k(\omega_n)})$ are similarly the corresponding realizations of random objects which we denote by $K$ and $(R_1, \ldots, R_K)$.

Note that for a generic sequence $(b_1, \ldots, b_n)$ we can either have $b_1 = 1$ or $b_1 = 0$. We start with the first case, i.e., we first condition ourselves on the event $B_1 = 1$.

Case I ($b_1 = 1$): It is easy to see that assuming $b_1 = 1$ we have:

$$
\sum_{i=1}^{n} b_i \geq \sum_{j \text{ odd} \leq k(\omega_n)} r_j,
$$

(168)

and

$$
n = \sum_{j=1}^{k(\omega_n)} r_j.
$$

For such a choice of $b_1$’s, we will now write the value of the process $A_n$ for several small values of $n$. In this way, it is easy to notice a simple pattern for the evolution of $A_n$ in terms of the sequence of runs. We have

$$
A_1 = a_0 r_1,
A_2 = a_0 2 r_1 - r_2,
A_4 = (a_0 2 r_1 - r_2) 2 r_3 = a_0 2^2 r_1 + r_3 - r_2 2 r_3,
A_7 = (a_0 2^2 r_1 - r_2) 2 r_3 = a_0 2^3 r_1 + r_3 - r_2 2^2 r_3 - r_4,
A_8 = ((a_0 \times 2^3 - r_2) \times 2 r_3 - r_4) \times 2 r_5
$$

$$
= a_0 2^3 r_1 + r_3 - r_2 2^2 r_3 - r_4 2 r_5
$$

$$
= 2^2 r_1 + r_3 - r_2 2 - 2(r_1 + r_3) r_4.
$$

(169)

In general, for a sequence $(b_1, \ldots, b_n)$ with the associated run sequence $(r_1, \ldots, r_{k(\omega_n)})$ we can write (note that $b_1 = 1$):

$$
A_n = a_0 2\sum_{i \text{ odd} \leq k(\omega_n)} r_i - \sum_{i \text{ even} \leq k(\omega_n)} r_i 2\sum_{j \text{ odd} \leq i} r_j + \sum_{i \text{ odd} \leq k(\omega_n)} r_i 2\sum_{j \text{ odd} \leq i} r_j
$$

$$
= [2\sum_{i \text{ odd} \leq k(\omega_n)} r_i] \left[ a_0 - \left( \sum_{i \text{ even} \leq k(\omega_n)} r_i 2\sum_{j \text{ odd} \leq i} r_j \right) \right].
$$

(167)

Here, by $\sum_{i \text{ even} \leq j} r_i$ we mean that the sum is over all the positive integers $i$ that are even and are also less than the given value $j$. Similarly, for example by $\sum_{i \text{ odd} \leq j} r_i$ we mean that the sum is over all integers $i$ that are odd and also
satisfy \( j < i \leq k(\omega_n) \). Now, if we consider the random vector \((B_1, B_2, \ldots, B_n)\) and its associated run sequence \((R_1, \ldots, R_K)\), we can write
\[
\Pr(A_n \geq 2\sum_{i} R_i) = \Pr(2\sum_{i} B_i(a_0 - \sum_{i \text{ even } \leq K} R_i) \geq 2\sum_{i} B_i(a_0 - 1)) = \Pr(a_0 - \sum_{i \text{ even } \leq K} R_i \geq a_0 - 1). \tag{170}
\]

Our objective is to find a lower bound on the probability of \((165)\). In this regard, by using \((170)\), we can equivalently find an upper-bound on the probability of the complementary event:
\[
\Pr(\sum_{i \text{ even } \leq K} R_i \geq a_0 - 1). \tag{171}
\]

For \( n \in \mathbb{N} \), define the set \( U_n \in \mathcal{F}_n \) as
\[
U_n = \{ \omega_n \in \Omega_n | \exists l \leq k(\omega_n) : \sum_{i \text{ even } \leq l} r_i 2^{-\sum_{i \text{ odd } < i} R_j} \geq a_0 - 1 \}.
\]

Clearly we have:
\[
\Pr(\sum_{i \text{ even } \leq K} R_i 2^{-\sum_{i \text{ odd } < i} R_j} \geq a_0 - 1) \leq \Pr(U_n).
\]

Obtaining an upper bound on \(\Pr(U_n)\) for finite \( n \) seems to be a difficult task. This is because for finite \( n \) handling the distribution of the runs is cumbersome. The idea here is to show that we can obtain useful bounds on \(\Pr(U_n)\) (for any finite \( n \)) by considering the case when \( n \) tends to \( \infty \). In the infinite \( n \) limit, the run sequence \( \{R_i\}_{i \in \mathbb{N}} \) becomes an iid sequence (note that \( B_1 = 1 \)) and this makes the proofs much simpler.

In the following we show that if \((b_1, \ldots, b_n) \in U_n\), then for any choice of \(b_{n+1}\), it is true that \((b_1, \ldots, b_n, b_{n+1}) \in U_{n+1}\). The two bits \( b_n \) and \( b_{n+1} \) can and jointly take four possible values. Here, for the sake of brevity, we will only consider the case when \( b_n, b_{n+1} = 1 \), and the other three cases can be verified similarly. Let \( \omega_n = (b_1, \ldots, b_{n-1}, b_n = 1) \in U_n \). Hence, \( k(\omega_n) \) is an odd number (recall that \( b_1 = 1 \)) and the quantity \( \sum_{i \text{ even } \leq l} \sum_{i \text{ even } \leq l} r_i 2^{-\sum_{i \text{ odd } < i} R_j} \) does not depend on the value of \( r_k(\omega_n) \). Now consider the sequence \( \omega_n = (b_1, \ldots, b_n = 1, 1) \). Since the last bit \( (b_{n+1}) \) equals 1, then \( k(\omega_{n+1}) = k(\omega_n) \) (i.e. the two sequences \( \omega_n \) and \( \omega_{n+1} \) have the same number of runs). Therefore, it is easy to see that
\[
\sum_{i \text{ even } \leq l} r_i 2^{-\sum_{i \text{ odd } < i} R_j} = \sum_{i \text{ even } \leq l} r_i 2^{-\sum_{i \text{ odd } < i} R_j}.
\]

As a result \((b_1, \ldots, b_n, 1) \in U_{n+1}\). From above, we conclude that for any \( i \in \mathbb{N} \) we have \( \theta_i(U_i) \subseteq \theta_{i+1}(U_{i+1}) \) and as a result
\[
\Pr(U_i) = \Pr(\theta_i(U_i)) \leq \Pr(\theta_{i+1}(U_{i+1})) = \Pr(U_{i+1}).
\]

Hence, the quantity \( \lim_{n \to \infty} \Pr(U_n) = \lim_{n \to \infty} \Pr(\theta_n(U_n)) = \lim_{n \to \infty} \Pr(\sum_{i \in \mathcal{V}_n} \theta_i(U_i)) \) is an upper bound on \((171)\). Let us now consider the set
\[
V = \{ \omega \in \Omega | \exists l \in \mathbb{N} : \sum_{i \text{ even } \leq l} r_i 2^{-\sum_{i \text{ odd } < i} R_j} \geq a_0 - 1 \}.
\]

By the definition of \( V \) we have \( U_{\infty} \subseteq \theta_i(U_i) \subseteq V \), and as a result, \( \Pr(U_{\infty} \subseteq \theta_i(U_i)) \leq \Pr(V) \). In order to bound the probability of
\[\text{Note here that the } V \subseteq \Omega, \text{ while } U_n \subseteq \Omega_n.\]
applies for $A_n$. Firstly, note that by fixing the value of $n$ the distribution of $R_1$ is as follows: $\Pr(R_1 = i|B_1 = 0) = \frac{1}{2^i}$ for $1 \leq i \leq n - 1$ and $\Pr(R_1 = n|B_1 = 0) = \frac{1}{2^n}$. We have

$$\Pr(A_n \geq 2^{\sum_{j=1}^{n} B_i} | B_1 = 0)$$

$$= \sum_{j=1}^{\min(a_{0-1}, n)} \Pr(A_n \geq 2^{\sum_{j=1}^{n} B_i} | R_1 = j, B_1 = 0) \Pr(R_1 = j | B_1 = 0)$$

$$\leq \sum_{j=1}^{\min(a_{0-1}, n)} \Pr(A_n \geq 2^{\sum_{j=1}^{n} B_i} | R_1 = j, B_1 = 0) \Pr(R_1 = j | B_1 = 0)$$

$$+ \sum_{j=a_{0-1}}^{n} \Pr(R_1 = j | B_1 = 0)$$

$$\leq \sum_{j=1}^{\min(a_{0-1}, n)} \Pr(A_n - j \geq 2^{\sum_{j+1}^{n} B_i} | B_{j+1} = 1, A_j = a_0 - j) \times \frac{1}{2^j}$$

$$+ 2 \sum_{j=a_{0-1}}^{\infty} \frac{1}{2^j}$$

$$\leq \sum_{j=1}^{\min(a_{0-1}, n)} \frac{1}{2^j} \tilde{c} \cdot \frac{1}{2^{a_{0-1} - j}} + \frac{4}{2^{a_{0-1}}}$$

$$\leq \tilde{c} a_0 + \frac{4}{2^{a_{0-1}}}.$$

Thus, we can write (note by (174) that $\tilde{c} > 4$)

$$\Pr(A_n \geq 2^{\sum_{j=1}^{n} B_i} | B_1 = 0, A_0 = a_0) \geq 1 - \frac{\tilde{c}(1 + a_0)}{2^{a_0-1}}. \quad (176)$$

Finally, by considering the two cases together, we obtain from (175) and (176) the following:

$$\Pr(A_n \geq 2^{\sum_{j=1}^{n} B_i}) \geq 1 - \frac{2\tilde{c}(1 + a_0)}{2^{a_0}}.$$

Hence, the proof of the lemma follows with $c_2 = 2\tilde{c} = \frac{2}{(2^{\tilde{c}} - 1)^2}$.

**Lemma 23:** Let $\alpha < 1$ be a constant. We have for $x \in (0, \frac{4}{3}]$

$$x \log\left(\frac{1}{x}\right) \leq c_3 \left(x(1-x)^{\alpha}\right), \quad (177)$$

where

$$c_3 = \frac{2}{(1-\alpha) \ln 2}. \quad (178)$$

**Proof:** By applying the function $\log(\cdot)$ to both sides of (177) and some further simplifications, the inequality (177) is equivalent to the following: For $x \in (0, \frac{4}{3}]$

$$\log(\log\left(\frac{1}{x}\right)) \leq \log c_3 + (1 - \alpha) \log \frac{1}{x} + \alpha \log(1 - x).$$

As $x \leq \frac{3}{4}$, we have $\alpha \log(1 - x) \geq -\log 4$. Hence, in order for the above inequality to hold it is sufficient that for $x \in (0, \frac{3}{4}]$

$$\log(\log\left(\frac{1}{x}\right)) \leq \log \frac{c_3}{4} + (1 - \alpha) \log \frac{1}{x}.$$  

Now, by letting $u = \log\left(\frac{1}{x}\right)$, the last inequality becomes

$$(1 - \alpha)u - \log u + \log \frac{c_3}{4} \geq 0, \quad (179)$$

for $u \geq \log(\frac{4}{3})$. It is now easy to check that by the choice of $c_3$ as in (178), the minimum of the above expression over the range $u \geq \log(\frac{4}{3})$ is always non-negative and hence the proof follows.