Refined Rate of Channel Polarization

Toshiyuki Tanaka and Ryuhei Mori
Graduate School of Informatics, Kyoto University, Kyoto-shi, Kyoto, 606-8501 Japan.
e-mail: tt@i.kyoto-u.ac.jp, rmori@sys.i.kyoto-u.ac.jp

Abstract—A rate-dependent upper bound of the best achievable block error probability of polar codes with successive-cancellation decoding is derived.

I. INTRODUCTION

Channel polarization [1] is a method which allows us to construct a family of error-correcting codes, called polar codes. Polar codes have been attracting theoretical interest because they are capacity achieving for binary-input symmetric memoryless channels (B-SMCs), which are also achieving symmetric capacity for general binary-input memoryless channels (B-MCs), whereas computational complexity of encoding and decoding is polynomial in the block length. Soon after the first proposal [1], one can find in the literature a number of contributions regarding channel polarization and polar codes [2]–[10].

Of particular theoretical interest is analysis of how fast the best achievable block error probability $P_e$ of polar codes decays toward zero as the block length $N$ tends to infinity. Arıkan [1] has shown that $P_e$ tends to zero as $N \to \infty$ whenever the code rate $R$ is less than the symmetric capacity of the underlying channel. The upper bound he obtained is proportional to a negative power of $N$, which means that guaranteed speed of the convergence to zero is very slow. His result has subsequently been improved by Arıkan and Telatar [4], who have obtained a much tighter upper bound, which scales as exponential in $-N^{\beta}$ for $\beta < 1/2$. Both of these bounds, however, do not depend on the code rate $R$. A rate-dependent bound is more desirable, since one naturally expects a smaller error probability from a smaller code rate, which might in turn suggest that the rate-independent bounds are not tight.

In this paper, we present an analysis of the rate of channel polarization. The argument basically follows that of Arıkan and Telatar [4], but extends it to obtain rate-dependent bounds of the best achievable error probability.

II. PROBLEM

Let $W : \mathcal{X} \mapsto \mathcal{Y}$ be an arbitrary binary-input memoryless channel (B-MC) with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y}$, and channel transition probabilities $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$. Let $I(W)$ be the symmetric capacity of $W$, which is defined as the mutual information between the input and output of $W$ when the input is uniformly distributed over $\mathcal{X}$. It is an upper bound of achievable rates over $W$ with codes that use input symbols with equal frequency. Let the Bhattacharyya parameter $Z(W)$ of the channel $W$ be defined as

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

It is an upper bound of the maximum-likelihood estimation error for a single channel usage.

Polar codes are constructed on the basis of recursive application of channel combining and splitting operation. In this operation, two independent copies of a channel $W$ is combined and then split to generate two different channels $W^- : \mathcal{X} \mapsto \mathcal{Y}^2$ and $W^+ : \mathcal{X} \mapsto \mathcal{Y} \times \mathcal{X}$. The operation, in its most basic form, is defined as

$$W^-(y_1, y_2|x_1) = \sum_{x \in \mathcal{X}} \frac{1}{2} W(y_1|x_1)W(y_2|x_2),$$
$$W^+(y_1, y_2, x_1|x_2) = \frac{1}{2} W|1\rangle(x_1)W|2\rangle(x_2),$$

with

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad x = (x_1, x_2).$$

It has been shown [1] that

$$Z(W^+) = Z(W)^2,$$

$$Z(W) \leq Z(W^-) \leq 2Z(W) - Z(W)^2. \quad (3)$$

In constructing polar codes, we recursively generate channels with the channel combining and splitting operation, starting with the given channel $W$, as

$$W \to \{W^-, W^+\} \to \{W^-, W^+, W^+, W^+\} \to \cdots,$$

where we have adopted the shorthand notation $W^{--} = (W^-)^{-}$, etc.

Following Arıkan [1], this process of recursive generation of channels can be dealt with by introducing a channel-valued stochastic process, defined as follows. Let $\{B_1, B_2, \ldots\}$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables with $P(B_1 = 0) = P(B_1 = 1) = 1/2$. Given a channel $W$, we define a sequence of channel-valued random variables $\{W_0, W_1, \ldots\}$ as

$$W_0 = W, \quad W_{n+1} = \begin{cases} W_n^- & \text{if } B_{n+1} = 1 \\ W_n^+ & \text{if } B_{n+1} = 0 \end{cases}. \quad (5)$$

We also define a real-valued random process $\{Z_0, Z_1, \ldots\}$ via $Z_n = Z(W_n)$. 
Conceptually, a polar code is constructed by picking up channels with good quality, among \(N = 2^n\) realizations of \(W_n\). We use these selected channels for transmitting data, while some predetermined values are transmitted over the remaining unselected channels. Thus, the rate of the resulting polar code is \(R\) if we pick up \(NR\) channels. We are interested in performance of polar codes under successive cancellation (SC) decoding, which is defined in \cite{1}. Let \(P_e(N, R)\) be the best achievable block error probability of polar codes of block length \(N\) and rate \(R\) under successive cancellation decoding. Since the Bhattacharyya parameter \(Z(W)\) serves as an upper bound of bit error probability in each step of successive cancellation decoding, an inequality of the form

\[
P(Z_n \leq \gamma) \geq R
\]

implies \(P_e(N, R) \leq NR\gamma\) via union bound.

It has been proved \cite{1} that, for any \(R < I(W)\), there exists a polar code with block length \(N = 2^n\), whose block error probability \(P_e(N, R)\) is arbitrarily close to 0. The proof is based on showing the condition \(6\) to hold for \(\gamma \in o(N^{-1})\).

III. Main Result

The main contribution of this paper is to prove the following theorem, which improves the results in \cite{1}, \cite{3}, giving a rate-dependent upper bound of the block error probability.

**Theorem 1:** Let \(W\) be any B-MC with \(I(W) > 0\). Let \(R \in (0, I(W))\) be fixed. Then, for \(N = 2^n, n \in \mathbb{N}\), the best achievable block error probability \(P_e(N, R)\) satisfies,

\[
P_e(N = 2^n, R) = o\left(2^{-2^{(n+t+\delta)/2}}\right),
\]

for any \(t\) satisfying \(t < Q^{-1}(R/I(W))\), where \(Q(x) = \int_{x}^{\infty} e^{-u^2/2} du/\sqrt{2\pi}\).

IV. Proof

A. Outline

The proof basically follows that of Arikan and Telatar \cite{3} but extends it in several respects. It consists of three stages, which we call polarization, concentration, and bootstrapping, respectively. In the first stage, it will be argued that realizations of \(Z_n\) are in \((0, \zeta)\) for some \(\zeta > 0\) with probability arbitrarily close to \(I(W)\) as \(n\) becomes large. This corresponds to the fundamental result of channel polarization \cite{1}. In the second stage, concentration will be argued, that is, again with probability arbitrarily close to \(I(W)\) as \(n\) gets large, realizations of \(Z_n\) are in \((0, f_n)\) for some \(f_n\) approaching zero exponentially in \(n\). In the last stage, we will argue that, once \(Z_m\) for some \(m\) enters the interval \([0, f_n]\), the sequence \(Z_{m+1}, \ldots, Z_n\) is rapidly decreasing with overwhelming probability, which is a refinement of the “bootstrapping argument” of \cite{3}. The last stage is further divided into two substages, the rate-independent bootstrapping stage and the rate-dependent bootstrapping stage, the latter of which is crucial in order to see dependence on the code rate.

B. Preliminaries

For \(m, n \in \mathbb{N}\) with \(m < n\), define

\[
S_{m,n} = \sum_{i=m+1}^{n} B_i,
\]

which follows a binomial distribution, since it is a sum of i.i.d. Bernoulli random variables.

**Definition 1:** For a fixed \(\gamma \in [0, 1]\), let \(G_{m,n}(\gamma)\) be the event defined by

\[
G_{m,n}(\gamma) = \{S_{m,n} \geq \gamma(n-m)\}.
\]

From the law of large numbers,

\[
\lim_{n-m \to \infty} P(G_{m,n}(\gamma)) = 1
\]

holds if \(\gamma < 1/2\).

C. Random Process

We now consider the random process \(X_n \in [0, 1]\) satisfying the following properties.

1) \(X_n\) converges to a random variable \(X_{\infty}\) almost surely.
2) Conditional on \(X_n\), if \(X_n \neq 0, 1\),

\[
X_{n+1} \begin{cases} \in [X_n, qX_n] & \text{if } B_{n+1} = 1 \\ = X_n^2 & \text{if } B_{n+1} = 0 \end{cases}
\]

for a constant \(q \geq 1\), and \(X_{n+1} = X_n\) with probability 1 for \(X_n = 0\) or 1.

Equation \(3\) implies that the random process \(Z_n\) satisfies the above properties with \(q = 2\). It should be noted that the properties 1 and 2 imply \(P(X_{\infty} \in \{0, 1\}) = 1\).

**Definition 2:** For \(\zeta \in (0, 1)\) and \(n \in \mathbb{N}\), define an event \(T_n(\zeta)\) with

\[
T_n(\zeta) = \{X_i \leq \zeta; \forall i \geq n\}.
\]

The following lemma is an immediate consequence of the above definition.

**Lemma 1:** For any fixed \(\zeta \in (0, 1)\),

\[
\lim_{n \to \infty} P(T_n(\zeta)) = P(X_{\infty} = 0).
\]

D. Concentration

For large enough \(n\), one can expect that \(X_n\) is exponentially small in \(n\) with probability arbitrarily close to \(P(X_{\infty} = 0)\). In other words, a \(P(X_{\infty} = 0)\)-fraction of realizations of \(X_n\) “concentrates” toward zero. To formalize the above statement, we introduce the following definition.

**Definition 3:** Let \(\rho \in (0, 1)\) and \(\beta \in (0, 1/2)\). The events \(C_n(\rho)\) and \(D_n(\beta)\) are defined as

\[
C_n(\rho) = \{X_n \leq \rho^n\},
\]

\[
D_n(\beta) = \{X_n \leq 2^{-2^{n-\beta}}\},
\]

respectively.

We will first prove that the event \(C_n\) has a probability arbitrarily close to \(P(X_{\infty} = 0)\) as \(n\) tends to infinity, on the basis of which we will next prove that the event \(D_n(\beta)\) has a probability arbitrarily close to \(P(X_{\infty} = 0)\) as \(n \to \infty\).
Letting such that holds. conditional on obtains consecutive 0s followed by consideration of properties of a process follows the same idea. leading to the "bootstrapping argument" of \[4\]. We basically converge to zero exponentially fast. This is the basic idea (0 the condition). We take \(D\) proposition, on the basis of which the result for the event \(C_{\alpha n}(\rho(\gamma))\) in Lemma 2, and let \(\rho(\gamma)\) satisfying the condition \((1 - \alpha)\) of \[4\]. We basically follow the same idea.

The proof of the bootstrapping stage is based on a consideration of properties of a process \(\{L_i\}\), defined on the basis of \(\{X_i\}\) as

\[
L_i = \log_2 X_i, \quad i = 0, \ldots, m, \quad (12)
\]

\[
L_{i+1} = \begin{cases} 2L_i & \text{if } B_{i+1} = 1 \\ L_i + \log_2 q & \text{if } B_{i+1} = 0 \end{cases}, \quad i \geq m \quad (13)
\]

for a fixed \(m\). The inequality \(X_i \leq 2^{L_i}\) holds on the sample-path basis for all \(i \geq 0\).

If we fix \(L_m\) and \(S_{m, n}\), the largest value of \(L_n\) is achieved by the sequence \(\{B_{m+1}, B_{m+2}, \ldots\}\) of \((n - m - S_{m, n})\) consecutive 0s followed by \(S_{m, n}\) consecutive 1s. One therefore obtains

\[
L_n \leq 2^{S_{m, n}} [L_m + (n - m - S_{m, n}) \log_2 q]. \quad (14)
\]

**Lemma 2:** Fix \(\gamma \in (0, 1)\) and \(\varepsilon > 0\), and let \(\rho = \rho(\gamma)\) be such that \(\log_2 \rho = -(1 - \gamma)(n - m) \log_2 q/m - \varepsilon\) holds. Then, conditional on \(C_{\alpha n}(\rho(\gamma)) \cap G_{m, n}(\gamma)\), the inequality

\[
L_n \leq 2^{\gamma(n-m)\varepsilon m}
\]

holds.

**Proof:** Conditional on \(C_{\alpha n}(\rho) \cap G_{m, n}(\gamma)\), one has, from (14), the inequality

\[
L_n \leq 2^{\gamma \log_2 \rho + (1 - \gamma)(n - m) \log_2 q}. \quad (15)
\]

Letting \(\rho = \rho(\gamma)\) completes the proof.  

**Proposition 2:** For an arbitrary fixed \(\beta \in (0, 1/2)\), let \(D_{\alpha n}(\beta)\) be the event defined in (11). Then,

\[
\lim_{n \to \infty} P(D_{\alpha n}(\beta)) = P(X_\infty = 0).
\]

**Proof:** Since \(\beta \in (0, 1/2)\), there exists \((\gamma, \alpha) \in (0, 1/2) \times (0, 1) \) satisfying the condition \((1 - \alpha) = \beta\) (e.g., letting \(\gamma = (1 + 2\beta)/4\) and \(\alpha = (1 - 2\beta)/(1 + 2\beta)\) satisfies the condition). We take \(m = \alpha n\) in Lemma 2 and let \(\{L_i^{(1)}\}\) denote the process defined by (12) and (13) with \(m = \alpha n\). Then, for any \(\varepsilon > 0\), one obtains by applying Lemma 2 that, conditional on the event \(C_{\alpha n}(\rho(\gamma)) \cap G_{m, n}(\gamma)\) with \(\rho(\gamma)\) defined in Lemma 2 the inequality

\[
L_{n}^{(1)} \leq 2^{\gamma \alpha n} \cdot \varepsilon \alpha n 
\]

holds, which in turn implies

\[
\{X_n \leq 2^{\gamma \alpha n} \varepsilon \alpha n\} \supseteq C_{\alpha n}(\rho(\gamma)) \cap G_{m, n}(\gamma). \quad (15)
\]

For any \(n \geq (\varepsilon \alpha)^{-1}, \beta n \leq (1 - \alpha) n + \log_2 \varepsilon \alpha n\) holds, so that one obtains

\[
D_{\alpha n}(\beta) \supseteq \{X_n \leq 2^{\gamma \alpha n} \varepsilon \alpha n\}. \quad (16)
\]

From (15) and (16), as well as the independence of \(C_{\alpha n}(\rho(\gamma))\) and \(G_{m, n}(\gamma)\), one consequently has

\[
P(D_{\alpha n}(\beta)) \geq P(G_{m, n}(\gamma)) P(C_{\alpha n}(\rho(\gamma))). \quad (17)
\]

Hence, using (18) and Proposition 1

\[
\lim_{n \to \infty} P(D_{\alpha n}(\beta)) \geq \lim_{n \to \infty} P(G_{m, n}(\gamma)) P(C_{\alpha n}(\rho(\gamma))) = P(X_\infty = 0). \quad (19)
\]

**F. Bootstrapping: Rate-Dependent Stage**

So far, our treatment of the random variable \(S_{m, n}\) is restricted to that within regimes of the law of large numbers. In order to obtain a rate-dependent bound, we have to go further and treat \(S_{m, n}\) within regimes of the central limit theorem.

**Definition 4:** For \(t \in \mathbb{R}\) and for a function \(f(n) = o(\sqrt{n})\), the event \(\mathcal{H}_{m, n}(t)\) is defined as

\[
\mathcal{H}_{m, n}(t) = \left\{ S_{m, n} \geq \frac{1}{2} (n - m + t\sqrt{n - m}) + f(n) \right\}. \quad (19)
\]

Noting that the random variable \(S_{m, n}\) is a sum of \((n - m)\) i.i.d. Bernoulli random variables, and that the mean and the variance are \((n - m)/2\) and \((n - m)/4\), respectively, the following lemma is a direct consequence of the central limit theorem.

**Lemma 3:** Let \(m < n\). Then, for any \(t \in \mathbb{R}\),

\[
\lim_{n \to \infty} \inf_{t \in \mathbb{R}} P(\mathcal{H}_{m, n}(t)) = Q(t). \quad (20)
\]

**Proposition 3:** For an arbitrary function \(f(n) = o(\sqrt{n})\),

\[
\lim_{n \to \infty} P \left( X_n \leq 2^{\gamma \alpha n} \varepsilon \alpha n \right) \geq Q(t) P(X_\infty = 0). \quad (21)
\]

**Proof:** For a fixed \(\beta \in (0, 1/2)\), we take \(m = \frac{1}{\beta} \log_2 n\) in Lemma 2 and let \(\{L_i^{(2)}\}\) denote the process defined by (12) and (13) with this choice of \(m\).

Conditional on the event \(D_{\alpha n}(\beta)\), one obtains, from (14), the inequality

\[
L_{n}^{(2)} \leq 2^{S_{m, n}} \left[ -2^{\beta m} + (n - m - S_{m, n}) \log_2 q \right]. \quad (20)
\]

Let \(\mathcal{H}_{m, n}(t)\) be the event defined in Definition 4 for a fixed \(t \in \mathbb{R}\) and for an arbitrarily chosen function \(f(n) = o(\sqrt{n})\).
Conditional on $D_m(\beta) \cap \mathcal{H}_{m,n}(t)$, $L_n^{(2)}$ is bounded from above as
\[
L_n^{(2)} \leq 2^{\frac{1}{2}((m+n) + \frac{1}{4} H \sqrt{m-n} + f(m-n))} \times \left[ -2^{\beta m} + \left( \frac{n-m-\frac{1}{2} \sqrt{n-m} - f(n-m)}{2} \right) \log_2 q \right]
\]
which implies that there exists $n_0$ such that for all $n \geq n_0$, the condition
\[
\{ X_n \leq 2^{\frac{1}{2}((m+n) + \frac{1}{4} H \sqrt{m-n} + f(m-n))} \} \cap D_m(\beta) \cap \mathcal{H}_{m,n}(t)
\]
is satisfied. From this observation, as well as the independence of $D_m(\beta)$ and $\mathcal{H}_{m,n}(t)$, one has
\[
P \left( X_n \leq 2^{\frac{1}{2}((m+n) + \frac{1}{4} H \sqrt{m-n} + f(m-n))} \right) \leq P(D_m(\beta))P(\mathcal{H}_{m,n}(t)).
\]
Thus,
\[
\liminf_{n \to \infty} P \left( X_n \leq 2^{\frac{1}{2}((m+n) + \frac{1}{4} H \sqrt{m-n} + f(m-n))} \right) \geq \liminf_{n \to \infty} P(D_m(\beta))P(\mathcal{H}_{m,n}(t)) = P(X_{\infty} = 0)Q(t).
\]
Since $m = o(\sqrt{n})$, one can safely absorb possible effects of $m$ into the function $f$. This completes the proof. \hfill $\blacksquare$

G. Converse

In this subsection, we discuss the converse, in which probabilities that $X_n$ takes small values are bounded from above.

**Proposition 4:** For an arbitrary function $f(n) = o(\sqrt{n})$
\[
\limsup_{n \to \infty} P \left( X_n \leq 2^{\frac{1}{2}((m+n) + \frac{1}{4} H \sqrt{m-n} + f(m-n))} \right) \leq Q(t)P(X_{\infty} = 0)
\]

**Proof:** Fix a process $\{X_n\}$. Let $\{\tilde{X}_n\}$ be the random process defined as
\[
\tilde{X}_i = X_i, \quad \text{for } i = 0, \ldots, m
\]
\[
\tilde{X}_i = \begin{cases} \tilde{X}_{i-1}, & \text{if } B_i = 1 \\ \tilde{X}_{i-1}, & \text{if } B_i = 0 \end{cases}, \quad \text{for } i > m
\]
The inequality $X_i \geq \tilde{X}_i$ holds on the sample-path basis for all $i \geq 0$, which implies
\[
P(X_n \leq a) \leq P(\tilde{X}_n \leq a),
\]
for any $a$. One also has
\[
\log_2 \log_2 (1/\tilde{X}_{m+k}) = S_{m+k}^n + \log_2 \log_2 (1/X_m)
\]

The central limit theorem dictates that $\frac{1}{\beta_n}(S - k/2)$ asymptotically follows the standard Gaussian distribution, so that, for any fixed $m$ and for an arbitrary function $f(k) = o(\sqrt{k})$, one has
\[
P \left( \tilde{X}_{m+k} \leq 2^{\frac{1}{2}(k+1/\beta_n)+f(k)} \right) \left| X_m \right. = 2^\frac{k}{2} + \frac{t \sqrt{k}}{2} + f(k) \left| X_m \right. = Q(t) + o(1)
\]
as $k \to \infty$. For any fixed $\delta \in (0,1)$, and $m \geq 0$
\[
\limsup_{n \to \infty} P \left( X_{m+k} \leq 2^{\frac{1}{2}(m+k+1/\beta_n)+f(m+k)} \right) \leq \limsup_{k \to \infty} P \left( X_{m+k} \leq 2^{\frac{1}{2}(m+k+1/\beta_n)+f(m+k)} \right)
\]
\[
\leq \limsup_{k \to \infty} \left\{ P \left( \tilde{X}_{m+k} \leq 2^{\frac{1}{2}(m+k+1/\beta_n)+f(m+k)} \right) \left| X_m \leq \delta \right. \right\} + P \left( X_{m+k} \leq \delta \right).
\]
From Fatou’s lemma,
\[
\limsup_{k \to \infty} P \left( X_{m+k} \leq \frac{\delta}{2}, X_m > \delta \right) \leq P \left( X_{\infty} \leq \frac{\delta}{2}, X_m > \delta \right).
\]
On the basis of (26), (27), and (28), one obtains
\[
\limsup_{n \to \infty} P \left( X_n \leq 2^{\frac{1}{2}(n+1/\beta_n)+f(n)} \right) \leq Q(t)P(X_{m} \leq \delta) + P \left( X_{\infty} \leq \frac{\delta}{2}, X_m > \delta \right).
\]
Since this is true for all $m$, we conclude that
\[
\limsup_{n \to \infty} P \left( X_n \leq 2^{\frac{1}{2}(n+1/\beta_n)+f(n)} \right) \leq \limsup_{m \to \infty} \left\{ Q(t)P(X_{m} \leq \delta) + P \left( X_{\infty} \leq \frac{\delta}{2}, X_m > \delta \right) \right\} = Q(t)P(X_{\infty} = 0),
\]
holds, where we have used the almost-sure convergence of $X_m$ to $X_{\infty}$ (property 1 in Sect. IV-C). Putting Propositions 3 and 4 together, we arrive at the following theorem.

**Theorem 2:** For an arbitrary function $f(n) = o(\sqrt{n})$
\[
\limsup_{n \to \infty} P \left( X_n \leq 2^{\frac{1}{2}(n+1/\beta_n)+f(n)} \right) = Q(t)P(X_{\infty} = 0).
\]

In applying Theorem 2 to $\{Z_n\}$, it should be noted that $P(\mathcal{Z}_{\infty} = 0) = I(W)$ holds. Theorem 1 is proved straightforwardly on the basis of Theorem 2 via the argument at the end of Sect. IV-C.

V. DISCUSSION

A. Extension to Construction with a Larger Matrix

Polar codes can be constructed on the basis of a matrix larger than the $2 \times 2$ matrix $F$ in [3]. Korada, Şasgoğlu, and Urbanke [6] have provided a full characterization of whether a matrix induces channel polarization. They have shown that if
an \( \ell \times \ell \) matrix \( G \) is polarizing, then given a symmetric B-MC \( W \),
\[
\lim_{n \to \infty} P \left( Z_n \leq 2^{-\ell \cdot n^{\beta}} \right) = I(W)
\]
holds for any \( \beta < E(G) \), where \( E(G) \) is the exponent of the matrix \( G \) defined in \([6]\). For a non-polarizing matrix, the exponent \( E(G) \) is zero.

Our analysis can be extended to obtain a rate-dependent result for channel polarization using a larger matrix. The extension includes introduction of a sequence \( \{B_i\} \) of i.i.d. random variables with \( P(B_1 = k) = 1/\ell \) for \( k = 1, 2, \ldots, \ell \). Let \( \{D_1, D_2, \ldots, D_\ell\} \) be “partial distances” of the matrix \( G \) defined in \([6]\). The exponent \( E(G) \) is given by the mean of the random variable \( \log_\ell D_{B_i} \). Let \( V(G) \) be the variance of the random variable \( \log_\ell D_{B_i} \). Our result in this direction is the following:
\[
\lim_{n \to \infty} P \left( Z_n \leq 2^{-\ell \cdot (n^{\beta} + \log_\ell n(n^{\beta})^2)} \right) = Q(t)I(W). \tag{31}
\]

The worst case of polarizing partial distances is given by the case where only one of \( \{D_1, D_2, \ldots, D_\ell\} \) is equal to 2 and the rest are equal to 1. Since \( E(G) = \frac{\log_\ell 2}{\ell} \) and \( V(G) = \left( \frac{\log_\ell 2}{\ell} \right)^2 (\ell - 1) \) for the worst case, a universal bound is obtained as
\[
\lim_{n \to \infty} P \left( Z_n \leq 2^{-\ell \cdot (n^{\beta} + \log_\ell n(n^{\beta})^2)} \right) \geq Q(t)I(W), \tag{32}
\]
which can be regarded as a refinement of Theorem 8 in \([6]\).

### B. Minimum Distance and ML Decoding

We return to the original construction of polar codes on the basis of the \( 2 \times 2 \) matrix \( F \). Polar codes are linear codes, and their generator matrices are obtained from the matrices of the form \( F^{\otimes n} \) via removal of some rows (corresponding to “shortening”) and reordering of the remaining rows. Hussami, Korada, and Urbanke \([2]\) studied the class of linear codes constructed from \( F^{\otimes n} \) via shortening, and showed using minimum distance analysis that the error probability of such codes is \( \omega(2^{-2^{n\beta}}) \) (in the standard Landau notation) for \( \beta > \frac{1}{2} \). This fact means that polar codes with SC decoding achieve the best performance as \( n \to \infty \) up to the dominant term in the double exponent of the error probability. In this subsection, it is shown that the minimum distance analysis does not give the second dominant term in the double exponent of polar codes with SC decoding. This fact implies that SC decoding is not necessarily optimal in the second dominant term.

**Proposition 5:** For any codes whose generator matrix consists of \( 2^\alpha \) \( R \) distinct rows of \( F^{\otimes n} \) and any fixed \( t > Q^{-1}(R) \), the error probability of ML decoding is \( \omega(2^{-2^{n \cdot \log_2 n} + \log_2 n}) \).

**Proof:** Let \( \mathcal{I} \subseteq \{0, 1, \ldots, 2^n - 1\} \) denote the set of indices of rows of \( F^{\otimes n} \) chosen to form the generator matrix. The minimum distance of the codes is given by \( \min_{i \in \mathcal{I}} 2^w(i) \), where \( w(i) \) denotes the Hamming weight of the binary expansion of \( i \). Let the minimum distance of a code be \( 2^d \). Since the number of rows with weight \( 2^d \) of the matrix \( F^{\otimes n} \) is \( \binom{n}{d} \), one obtains the inequality
\[
\sum_{i=d}^{n} \binom{n}{i} \geq 2^{2^d R}, \tag{33}
\]
or equivalently,
\[
P(S_n \geq d) \geq R, \tag{34}
\]
where \( S_n \) is a sum of \( n \) i.i.d. Bernoulli random variables with probability one half. Let \( d = n/2 + t \sqrt{n}/2 \) for any fixed \( t \in \mathbb{R} \). Then,
\[
P \left( S_n - \frac{n}{2} \geq t \right) \geq R, \tag{35}
\]
From the central limit theorem, the left-hand side converges to \( Q(t) \) as \( n \to \infty \). Hence, the condition \( t \leq Q^{-1}(R) \) is necessary for asymptotic existence of the codes satisfying the conditions stated in the Proposition, completing the proof.

It should be noted that Proposition 5 also means that the minimum distance of the codes considered is asymptotically at most \( 2^{n \cdot Q^{-1}(R) \sqrt{n}/2} \).

The prefactor of the second dominant term \( \sqrt{n} \) in the double exponent is \( Q^{-1}(R) \) in Proposition 5 which is strictly larger than the prefactor \( Q^{-1}(R) I(W) \) in Theorem 1 whenever \( I(W) < 1 \). One can argue that it might be due to the channel-independent nature of the analysis leading to Proposition 5 which is reflected in the absence of the channel \( W \) in the result. In any case, whether polar codes with SC decoding are optimal in terms of the double exponent up to the second dominant term is an open problem, and thus needs further investigation.

### VI. Conclusion

We have derived a rate-dependent upper bound of the best achievable block error probability of polar codes with successive cancellation decoding. The derivation is based on that of the previous rate-independent results \([11, 13]\), which discusses channel polarization in regimes of the law of large numbers, extending it to regimes of the central limit theorem.

We would like to mention that the argument given in this paper can also be applied to the problem of lossy source coding discussed in \([7]\).

### References

[1] E. Arıkan, “Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. Info. Theory, vol. 55, no. 7, pp. 3051–3073, July 2009.

[2] E. Arıkan, “A performance comparison of polar codes and Reed-Muller codes,” IEEE Comm. Lett., vol. 12, no. 6, pp. 447–449, June 2008.

[3] R. Mori and T. Tanaka, “Performance of polar codes with construction using density evolution,” IEEE Comm. Lett., vol. 13, no. 7, pp. 519–521, July 2009.

[4] E. Arıkan and E. Telatar, “On the rate of channel polarization,” Proc. 2009 IEEE Int. Symp. Info. Theory, Seoul, Korea, pp. 1493–1495, June/July 2009; [online] arXiv:0807.3809v3 [cs.IT], 2008.

[5] S. B. Korada and E. Şaşoğlu, “A class of transformations that polarize binary-input memoryless channels,” Proc. 2009 IEEE Int. Symp. Info. Theory, Seoul, Korea, pp. 1478–1482, June/July 2009.

[6] S. B. Korada, E. Şaşoğlu, and R. Urbanke, “Polar codes: Characterization of exponent, bounds, and constructions,” Proc. 2009 IEEE Int. Symp. Info. Theory, Seoul, Korea, pp. 1483–1487, June/July 2009; [online] arXiv:0901.0536v2 [cs.IT], 2009.
[7] N. Hussami, S. B. Korada, and R. Urbanke, “Performance of polar codes for channel and source coding,” Proc. 2009 IEEE Int. Symp. Info. Theory, Seoul, Korea, pp. 1488–1492, June/July 2009; [online] arXiv:0901.2370v2 [cs.IT], 2009.

[8] R. Mori and T. Tanaka, “Performance and construction of polar codes on symmetric binary-input memoryless channels,” Proc. 2009 IEEE Int. Symp. Info. Theory, Seoul, Korea, pp. 1496–1500, June/July 2009; [online] arXiv:0901.2207v2 [cs.IT], 2009.

[9] S. B. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” [online] arXiv:0903.0307v1 [cs.IT], 2009.

[10] S. H. Hassani, S. B. Korada, and R. Urbanke, “The compound capacity of polar codes,” [online] arXiv:0907.3291v1 [cs.IT], 2009.