ISIT'2013 1569718957

On the Correlation Between Polarized BECs

Mani Bastani Parizi and Emre Telatar
EPFL, Lausanne, Switzerland
Email: {mani.bastaniparizi,emre.telatar}@epfl.ch

***Abstract*—We consider the 2^n channels synthesized by the n-fold application of Arıkan's polar transform to a binary erasure channel (BEC). The synthetic channels are BECs themselves, and we show that, asymptotically for almost all these channels, the pairwise correlations between their erasure events are extremely small: the correlation coefficients vanish faster than any exponential in n. Such a fast decay of correlations allows us to conclude that the union bound on the block error probability of polar codes is very tight.**

I. Introduction

Channel Polarization is a technique recently introduced by Arıkan [1] as a means of constructing capacity achieving codes for binary discrete memoryless channels (B-DMCs). The underlying principle of channel polarization is the following: Let $W : \mathcal{X} \longrightarrow \mathcal{Y}$ be a B-DMC with input alphabet $\mathcal{X} = \mathbb{F}_2$. From two independent copies of W synthesize $W^- : \mathcal{X} \longrightarrow \mathcal{Y}^2$ and $W^+ : \mathcal{X} \longrightarrow \mathcal{Y}^2 \times \mathcal{X}$ as:

$$W^-(y_1, y_2|u_1) = \sum_{u_2 \in \mathcal{X}} \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2),$$

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2).$$

This transform can be repeated n times to get $N = 2^n$ B-DMCs $W_n^{(\mathbf{s})}, \mathbf{s} \in \{-,+\}^n$. Arıkan shows that (i) the transformation preserves the mutual information, (ii) $W_n^{(\mathbf{s})}$s approach to "extremal" channels, i.e., either noiseless or useless channels. In particular, the fraction of almost noiseless channels is equal to the symmetric capacity of the original B-DMC W. Based on these properties Arıkan constructs *polar codes* by sending uncoded data bits only on (almost) noiseless channels and arbitrary (but known to receiver) bits on the remaining channels. The channels used to transmit information are referred to as "information" channels and the rest are called "frozen" channels. A successive cancellation decoder has been proposed by Arıkan to decode the information bits with complexity $O(N \log N)$ and shown to have a block error probability that behaves roughly as $O\big(2^{-\sqrt{N}}\big)$ (cf. [2]).

The set of Binary Erasure Channels (BECs) is stable under Polarization in the sense that if W is a BEC, then W^+ and W^- are also BECs. We denote a BEC with erasure probability ϵ as $\mathsf{BEC}(\epsilon)$. One can establish a one-to-one relationship between a $\mathsf{BEC}(\epsilon)$ and an "erasure indicator" random variable E such that $E \in \{0,1\}$ and $\mathbb{P}[E=1] = \epsilon$. The polar transform of a BEC is hence equivalent to taking two independent copies of E and creating the erasure indicators of W^- and W^+.

Lemma 1 (Polar Transform of BEC [1, Proposition 6])**.** *If W is a BEC with erasure probability ϵ, applying the polar transform $(W, W) \mapsto (W^-, W^+)$ produces two BECs: W^- with erasure probability $2\epsilon - \epsilon^2$ and W^+ with erasure probability ϵ^2. Moreover, W^- erases iff either copy of W erases, and W^+ erases iff both copies of W erase.*

Corollary 1. *The erasure indicators of W^- and W^+, denoted by E^- and E^+, are constructed from two independent copies of E, denoted by E and E', as:*

$$E^- = \max\{E, E'\} = E + E' - EE' \tag{1a}$$

$$E^+ = \min\{E, E'\} = EE'. \tag{1b}$$

While E and E' are independent (and hence uncorrelated), E^+ and E^- are correlated: $E^+ = 1$ implies $E^- = 1$. On the other side, by polarization $W_n^{(\mathbf{s})}$s (and equivalently $E_n^{(\mathbf{s})}$s) become deterministic as $n \to \infty$. Hence it looks like $E_n^{(\mathbf{s})}$ and $E_n^{(\mathbf{t})}$ would become uncorrelated for $\mathbf{s} \neq \mathbf{t}$, where $\mathbf{s}$ and $\mathbf{t}$ are sign sequences of length n used for indexing the channels. In particular it is easy to see that $\mathbb{E}\big[E_n^{(\mathbf{s})} E_n^{(\mathbf{t})}\big] - \mathbb{E}\big[E_n^{(\mathbf{s})}\big]\mathbb{E}\big[E_n^{(\mathbf{t})}\big]$ is small for almost every $\mathbf{s}, \mathbf{t}$.

In this paper we provide upper-bounds on correlation *coefficients* defined as:

$$\rho_n^{(\mathbf{s},\mathbf{t})} \triangleq \frac{\mathbb{E}\big[E_n^{(\mathbf{s})} E_n^{(\mathbf{t})}\big] - \mathbb{E}\big[E_n^{(\mathbf{s})}\big]\mathbb{E}\big[E_n^{(\mathbf{t})}\big]}{\sqrt{\mathrm{var}\big[E_n^{(\mathbf{s})}\big]\,\mathrm{var}\big[E_n^{(\mathbf{t})}\big]}} \tag{2}$$

and exploit these bounds and the inclusion–exclusion principle to find lower bounds on the block error probability of polar codes. In particular, our bounds are strong enough to show that the sum of the Bhattacharyya parameters of the information channels is a tight estimate of the block error probability.

II. Notation

Throughout this manuscript, we use uppercase letters (like X) to indicate a random variable, and its lowercase version (x) for a realization of that random variable. The boldface letters denote matrices, vectors or sequences which will be clear from the context.

We denote the sets by script-style uppercase letters like $\mathcal{S}$ and by $|\mathcal{S}|$ we mean the cardinality of $\mathcal{S}$.

We abbreviate $1 - x$ to $\overline{x}$.

For sign sequences $\mathbf{s} \in \{-,+\}^*$ and $\mathbf{t} \in \{-,+\}^*$, $\mathrm{CP}[\mathbf{s}, \mathbf{t}]$ denotes their common prefix. Furthermore let $|\mathbf{s}|$ denote the length of a sequence $\mathbf{s}$.

III. Properties of Correlation Coefficients

As we mentioned in Section I, we are interested in analyzing the matrix of correlation coefficients of the erasure indicator

vector $\mathbf{E}_n = \left[E_n^{(\mathbf{s})} : \mathbf{s} \in \{-,+\}^n\right]$. It is more convenient to index the $N = 2^n$ elements of that vector using sign sequences $\mathbf{s} \in \{-,+\}^n$ instead of mapping the sign sequences to integers and using the natural indexing. We will use the same indexing for the N^2 elements of the correlation coefficient matrix.

Arıkan has already shown that the vector $\mathbf{Z}_n = \mathbb{E}\left[\mathbf{E}_n\right]$ can be computed via a single-step recursion. More precisely, having $\mathbf{Z}_{n-1}$ we can compute the elements of $\mathbf{Z}_n$ as:

$$Z_n^{(\mathbf{s}-)} = 2Z_{n-1}^{(\mathbf{s})} - \left(Z_{n-1}^{(\mathbf{s})}\right)^2 \tag{3a}$$

$$Z_n^{(\mathbf{s}+)} = \left(Z_{n-1}^{(\mathbf{s})}\right)^2 \tag{3b}$$

for $\forall \mathbf{s} \in \{-,+\}^{n-1}$ with $Z_0 = \epsilon$.

Interestingly, the correlation coefficients matrix $\rho_n = \left[\rho_n^{(\mathbf{s},\mathbf{t})} : \mathbf{s},\mathbf{t} \in \{-,+\}^n\right]$ can also be computed via a single-step recursion as follows:

Lemma 2. *The correlation coefficients matrix of the random vector $\mathbf{E}_n$, ρ_n can be computed in terms of ρ_{n-1} and $\mathbf{Z}_{n-1}$ recursively as follows: Let $f_+(z) \triangleq \sqrt{z/(1+z)}$, $f_-(z) \triangleq f_+(\overline{z})$, $g_+(z) \triangleq \sqrt{\overline{z}/(1+z)}$ and $g_-(z) \triangleq -g_+(\overline{z})$. Then for $s,t \in \{-,+\}$ and $\mathbf{s},\mathbf{t} \in \{-,+\}^{n-1}$,*

$$\rho_n^{(\mathbf{s}s,\mathbf{t}t)} = \rho_{n-1}^{(\mathbf{s},\mathbf{t})}\Big[2f_s\big(Z_{n-1}^{(\mathbf{s})}\big)f_t\big(Z_{n-1}^{(\mathbf{t})}\big) + g_s\big(Z_{n-1}^{(\mathbf{s})}\big)g_t\big(Z_{n-1}^{(\mathbf{t})}\big)\rho_{n-1}^{(\mathbf{s},\mathbf{t})}\Big]. \tag{4}$$

Clearly $\rho_0 = 1$ by definition.

The property of being computable by a single-step recursion generalizes to the higher order statistics (cf. [3]).

One can derive the properties stated in the sequel on $\rho_n^{(\mathbf{s},\mathbf{t})}$ according to the aforementioned recursions. We omit the proof of Lemma 2 as well as the proofs of these properties as they are straightforward but tedious and can be found in [3].

Property 1. $0 \le \rho_n^{(\mathbf{s},\mathbf{t})} \le \min\left\{\sqrt{\frac{\overline{Z_n^{(\mathbf{s})}}\,Z_n^{(\mathbf{t})}}{Z_n^{(\mathbf{s})}\overline{Z_n^{(\mathbf{t})}}}}, \sqrt{\frac{Z_n^{(\mathbf{s})}\overline{Z_n^{(\mathbf{t})}}}{\overline{Z_n^{(\mathbf{s})}}Z_n^{(\mathbf{t})}}}\right\}$

Property 2. *For $\mathbf{s},\mathbf{t} \in \{-,+\}^{n-1}$ and $s_n, t_n \in \{-,+\}$ $\rho_n^{(\mathbf{s}s_n,\mathbf{t}t_n)} \le \rho_{n-1}^{(\mathbf{s},\mathbf{t})}$ with equality iff*

(i) $\rho_{n-1}^{(\mathbf{s},\mathbf{t})} = 0$, or

(ii) $s_n = t_n$ and $\rho_{n-1}^{(\mathbf{s},\mathbf{t})} = 1$ and $Z_{n-1}^{(\mathbf{s})} = Z_{n-1}^{(\mathbf{t})}$, or

(iii) $Z_{n-1}^{(\mathbf{s})} = b_{s_n}$ and $Z_{n-1}^{(\mathbf{t})} = b_{t_n}$ where $b_+ = 1$ and $b_- = 0$.

Property 3. *If $\mathbf{s} \ne \mathbf{t}$ then $\rho_n^{(\mathbf{s},\mathbf{t})} \le \frac{1}{3}$.*

IV. Convergence of Correlation Coefficients

In the previous section we showed how correlation coefficients can be computed efficiently by single-step recursions and derived some algebraic properties of them. In this section we show that correlation coefficients converge to zero.

Based on the properties we already obtained we have the following result:

Lemma 3. *Let $\mathbf{s}$ and $\mathbf{t}$ be infinite sign sequences such that $\mathbf{s} \ne \mathbf{t}$ and $\mathbf{s}^n$ and $\mathbf{t}^n$ be the subsequences corresponding to their first n elements respectively. Then $\lim_{n\to\infty} \rho_n^{(\mathbf{s}^n,\mathbf{t}^n)} = 0$.*

Proof: Let $m = |\mathrm{CP}\,[\mathbf{s},\mathbf{t}]|$ and $a_n \triangleq \rho_n^{(\mathbf{s}^n,\mathbf{t}^n)}$. For $n > m$, by Properties 1 and 3 we know $a_n \in [0, 1/3]$ and by Property 2 it is decreasing. Hence, a_n is a convergent sequence. Suppose its limit is $a^* > 0$. This implies for every $\varepsilon > 0$ there exist a n_0 such that for $n > n_0$, $a_n/a_{n-1} \ge 1 - \varepsilon$. By the continuity of (4), we must have $|Z_{n-1}^{(\mathbf{s}^{n-1})} - b_{s_n}| < \delta$ and $|Z_{n-1}^{(\mathbf{t}^{n-1})} - b_{t_n}| < \delta$ for all $n > n_0$ (according to the equality condition (iii) of Property 2) where δ is a quantity approaching zero as ε gets small. This implies $s_n = s^*$ and $t_n = t^*$ for all $n > n_0$ because the evolutions of Z do not allow Z to jump from one extreme to the other. Without loss of generality, assume $s^* = +$ which in turn requires $Z_{n-1}^{(\mathbf{s}^{n-1})} > 1 - \delta$. Now we have an incompatible situation: $s_n = +$ for all $n > n_0$ will drive $Z_n^{(\mathbf{s}^n)}$ to 0. This shows a_n cannot converge to a non-zero value. ∎

We can further show that the average of the elements of the correlation coefficients matrix is exponentially small in n.

Lemma 4. *For any $\mathbf{s},\mathbf{t} \in \{-,+\}^{n-1}$,*

$$\tfrac{1}{4}\textstyle\sum_{(s,t)\in\{-,+\}^2} \rho_n^{(\mathbf{s}s,\mathbf{t}t)} \le \tfrac{2}{3}\rho_{n-1}^{(\mathbf{s},\mathbf{t})}.$$

Proof: Let $a = Z_{n-1}^{(\mathbf{s})}$, $b = Z_{n-1}^{(\mathbf{t})}$, $f(x) \triangleq \frac{1}{\sqrt{2}}\left[\sqrt{\frac{x}{1+x}} + \sqrt{\frac{\overline{x}}{1+\overline{x}}}\right]$, and $g(x) \triangleq \frac{1}{2}\left[\sqrt{\frac{\overline{x}}{1+x}} - \sqrt{\frac{x}{1+\overline{x}}}\right]$. Using (4) one can easily verify that:

$$\begin{aligned}\tfrac{1}{4}\textstyle\sum_{(s,t)\in\{-,+\}^2} \rho_n^{(\mathbf{s}s,\mathbf{t}t)} &= f(a)f(b)\rho_{n-1}^{(\mathbf{s},\mathbf{t})} + g(a)g(b)\rho_{n-1}^{(\mathbf{s},\mathbf{t})^2} \\ &= \left[f(a)f(b) + g(a)g(b)\rho_{n-1}^{(\mathbf{s},\mathbf{t})}\right]\rho_{n-1}^{(\mathbf{s},\mathbf{t})}.\end{aligned}$$

Now, observe that both sides of the above are positive and:

$$\begin{aligned}&\left[f(a)f(b) + g(a)g(b)\rho_{n-1}^{(\mathbf{s},\mathbf{t})}\right]^2 \\ &\quad\overset{(*)}{\le} \left[f(a)^2 + \rho_{n-1}^{(\mathbf{s},\mathbf{t})}g(a)^2\right]\left[f(b)^2 + \rho_{n-1}^{(\mathbf{s},\mathbf{t})}g(b)^2\right] \\ &\quad\le \left[f(a)^2 + g(a)^2\right]\left[f(b)^2 + g(b)^2\right]\end{aligned}$$

where (*) follows from the Cauchy-Schwarz inequality. It is easy to see $f(x)^2 + g(x)^2 = \frac{1}{2}\left(1 + \sqrt{\frac{x\overline{x}}{(1+x)(1+\overline{x})}}\right)$ which is maximized at $x = \frac{1}{2}$ (for $x \in [0,1]$) with value $\frac{2}{3}$. ∎

Corollary 2. *The average of the normalized correlation matrix elements satisfies:*

$$\frac{1}{4^n}\sum_{\mathbf{s},\mathbf{t}\in\{-,+\}^n} \rho_n^{(\mathbf{s},\mathbf{t})} \le \left(\frac{2}{3}\right)^n \tag{5}$$

V. Rate of Convergence

Corollary 2 implies that for large enough n, almost all of non-diagonal entries of ρ_n are small. However, the bound it gives is not strong enough to show the asymptotic tightness of the union bound on the block error probability of polar codes. For that, one has to show (i) that the correlations decay like $O\left(2^{-(1+\alpha)n}\right)$ for some $\alpha > 0$, and (ii) that this bound applies not just to the average value of $\rho_n^{(\mathbf{s},\mathbf{t})}$ but to $\max_{\mathbf{t}\ne\mathbf{s}} \rho_n^{(\mathbf{s},\mathbf{t})}$ for the $\mathbf{s}$'s and $\mathbf{t}$'s which index the information channels.

To this end, we establish a probabilistic framework similar to that used in [1] for proving the channel polarization theorem.

Let $S_1, S_2, \ldots,$ be i.i.d Bernoulli $\left(\frac{1}{2}\right)$ random variables such that $S_i \in \{-,+\}$, define $\mathbf{S}^n \triangleq (S_1, S_2, \ldots, S_n)$ and $\mathcal{F}_n \triangleq \sigma(\mathbf{S}^n)$ as the σ-algebra generated by random vector $\mathbf{S}^n$. We consider the random variables $Z_n^{(\mathbf{S})} = \mathbb{E}\big[E_n^{(\mathbf{S}^n)} | \mathbf{S}^n\big]$ and $\rho_n^{(\mathbf{S}^n, \mathbf{t}^n)}$ for $\mathbf{t}^n \in \{-,+\}^n$ which are all $\mathcal{F}_n$ measurable.

We show that for any $\alpha > 0$, $\max_{\mathbf{t}^n \neq \mathbf{S}^n} \rho_n^{(\mathbf{S}^n, \mathbf{t}^n)} \leq 2^{-(1+\alpha)n}$ with very high probability for sufficiently large n.

A. Closely related $\mathbf{s}$ and $\mathbf{t}$

Let us first focus on $\rho_n^{(\mathbf{s},\mathbf{t})}$ for $\mathbf{s}$ and $\mathbf{t}$ sharing a long common prefix. Recall that $|\operatorname{CP}[\mathbf{s},\mathbf{t}]|$ denotes the length of this prefix.

Lemma 5. *Fix $\alpha > 0$. Set $m_n \triangleq 4\log\big(2(1+\alpha)n - 1\big)$. Then:*

$$\lim_{n\to\infty} \mathbb{P}\left[\max_{\mathbf{t}^n \neq \mathbf{S}^n : |\operatorname{CP}[\mathbf{S}^n, \mathbf{t}^n]| \geq m_n} \rho_n^{(\mathbf{S}^n, \mathbf{t}^n)} \leq 2^{-(1+\alpha)n}\right] = 1$$

Proof: Let $\mathbf{P} = \operatorname{CP}[\mathbf{S}^n, \mathbf{t}^n]$ and $n_0 = |\mathbf{P}|$. Observe that $\mathbf{P}$ is a uniformly chosen sign sequence in $\{-,+\}^{n_0}$. According to Property 2, $\rho_{n_0}^{(\mathbf{P},\mathbf{P})} = 1$ and:

$$\rho_n^{(\mathbf{S}^n, \mathbf{t}^n)} < \rho_{n_0+1}^{(\mathbf{P}S_{n_0+1}, \mathbf{P}t_{n_0+1})} = \sqrt{\frac{Z_{n_0}^{(\mathbf{P})}\overline{Z_{n_0}^{(\mathbf{P})}}}{2 + Z_{n_0}^{(\mathbf{P})}\overline{Z_{n_0}^{(\mathbf{P})}}}}$$
$$\leq \min\left\{\sqrt{\frac{1}{2}Z_{n_0}^{(\mathbf{P})}}, \sqrt{\frac{1}{2}\overline{Z_{n_0}^{(\mathbf{P})}}}\right\}.$$

Results of [2] show that for any fixed $0 < \beta < 1/2$ and $\delta > 0$ there exist a m_0 such that for $n_0 \geq m_0$

$$\mathbb{P}\left[Z_{n_0}^{(\mathbf{P})} \in [2^{-N_0^\beta}, 1 - 2^{-N_0^\beta}]\right] < \delta$$

where $N_0 = 2^{n_0}$.

In particular we take $\beta = \frac{1}{4}$ in the above bound and take n large enough so that $m_n \geq m_0$. Hence $n_0 \geq m_n \geq m_0$, and with probability at least $1-\delta$, $Z_{n_0}^{(\mathbf{P})}$ is extremal. Together with $2^{-N_0^{1/4}} \leq 2^{-2(1+\alpha)n+1}$ we get

$$\mathbb{P}\left[\rho_n^{(\mathbf{S}^n, \mathbf{t}^n)} \leq 2^{-(1+\alpha)n}\right] \geq 1 - \delta. \qquad \blacksquare$$

B. Distantly related $\mathbf{s}$ and $\mathbf{t}$

A more involved task is find an upper-bound on $\rho_n^{(\mathbf{s},\mathbf{t})}$ when $\mathbf{s}$ and $\mathbf{t}$ do not have a long common prefix. For this purpose we first seek an upper-bound on $\rho_n^{(\mathbf{S}^n, \mathbf{t}^n)} / \rho_{n-1}^{(\mathbf{S}^{n-1}, \mathbf{t}^{n-1})}$ only in terms of $\mathbf{S}^{n-1}$, S_n and $p_n = |\operatorname{CP}[\mathbf{S}^n, \mathbf{t}^n]|$, denoted as $\chi\left(\mathbf{S}^{n-1}, S_n, p_n\right)$.

To this end, let:

$$M\Big(S_n, t_n, \rho_{n-1}^{(\mathbf{S}^{n-1}, \mathbf{t}^{n-1})}, Z_{n-1}^{(\mathbf{S})}, Z_{n-1}^{(\mathbf{t})}\Big) \triangleq \frac{\rho_n^{(\mathbf{S}^n, \mathbf{t}^n)}}{\rho_{n-1}^{(\mathbf{S}^{n-1}, \mathbf{t}^{n-1})}}.$$

$M(s,t,r,a,b)$ takes four possible forms according to (4) each of which can be bounded as:

$$M(+,t,r,a,b) \leq \min\left\{1, \sqrt{2a} + r\right\}$$
$$M(-,t,r,a,b) \leq \min\left\{1, \sqrt{2\overline{a}} + r\right\}$$

using Lemma 6 (and triangle inequality if $s \neq t$):

Lemma 6. *Let $f(x) \triangleq \sqrt{\frac{x}{1+x}}$ and $g(x) \triangleq \sqrt{\frac{\overline{x}}{1+x}}$. Define*

$$F(r,a,b) \triangleq 2f(a)f(b) + g(a)g(b)r.$$

Then

$$F(r,a,b) \leq \min\left\{1, \sqrt{2a} + r\right\}, \tag{6}$$

for all $0 \leq r \leq 1, 0 \leq a \leq 1, 0 \leq b \leq 1$.

Proof: See [3]. $\blacksquare$

Observe that the upper-bounds on M depend only $Z_{n-1}^{(\mathbf{S})}$ and $\rho_{n-1}^{(\mathbf{S}^{n-1}, \mathbf{t}^{n-1})}$. Let us also define

$$\rho_{n,p}^{(\mathbf{s}^n, *)} \triangleq \max_{\mathbf{t}^n \neq \mathbf{s}^n : |\operatorname{CP}[\mathbf{s}^n, \mathbf{t}^n]| \leq p} \rho_n^{(\mathbf{s}^n, \mathbf{t}^n)}.$$

Consequently we may choose:

$$\chi\left(\mathbf{S}^{n-1}, +, p_n\right) = \min\left\{1, \sqrt{2Z_{n-1}^{(\mathbf{S})}} + \rho_{n-1,p_n}^{(\mathbf{S}^{n-1}, *)}\right\} \tag{7a}$$

$$\chi\left(\mathbf{S}^{n-1}, -, p_n\right) = \min\left\{1, \sqrt{2\overline{Z_{n-1}^{(\mathbf{S})}}} + \rho_{n-1,p_n}^{(\mathbf{S}^{n-1}, *)}\right\} \tag{7b}$$

Now we would like to show that $\min_{s_n} \chi\left(\mathbf{S}^{n-1}, s_n, p_n\right)$ gets arbitrarily small with very high probability. For this, we first need the following lemma:

Lemma 7. *For any sequence p_n such that $\lim_{n\to\infty} \frac{n}{2} - p_n = \infty$ and any fixed $\gamma > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\forall i \geq \frac{n}{2} : \rho_{i,p_n}^{(\mathbf{S}^i, *)} \leq \gamma\right] = 1. \tag{8}$$

Proof: Observe that for fixed p, $\rho_{i,p}^{(\mathbf{s}^i, *)}$ is decreasing in i (if $i > p$). Hence $\rho_{n/2,p_n}^{(\mathbf{s}^{n/2}, *)} \leq \gamma$ implies $\rho_{i,p_n}^{(\mathbf{s}^n, *)} \leq \gamma$ for all $i \geq n/2$.

Suppose $\mathbf{s}$ is a sequence such that for some $\mathbf{t} \neq \mathbf{s}$ with $|\operatorname{CP}[\mathbf{s},\mathbf{t}]| \leq p_n$, $\rho_{n/2}^{(\mathbf{s}^{n/2}, \mathbf{t}^{n/2})} > \gamma$. Recall that $\mathbf{s}^i$ (resp. $\mathbf{t}^i$) denotes the subsequence of $\mathbf{s}$ (resp. $\mathbf{t}$) including its first i elements.

Define $a_i \triangleq \rho_i^{(\mathbf{s}^i, \mathbf{t}^i)}$ and $m_i \triangleq a_i / a_{i-1}$. It is clear that $a_{p_n+1} \leq \frac{1}{3}$ and a_i is decreasing for $i > p_n$ by Properties 3 and 2.

For any $0 < \varepsilon < 1$, $a_{n/2} > \gamma$ implies that the number of indices $i \in \{p_n + 2, p_n + 3, \ldots, \frac{n}{2}\}$ for which $m_i \leq 1 - \varepsilon$ is at most $\frac{\log(3\gamma)}{\log(1-\varepsilon)}$.

Let $l = \frac{n}{2} - p_n - 1$, take $\varepsilon = 1/\sqrt{l}$, and observe that the number of indices for which $m_i \leq 1 - 1/\sqrt{l}$ is at most $c_\gamma \sqrt{l}$ where c_γ is a constant that depends on γ only. These indices partition the interval $[p_n + 2 : \frac{n}{2}]$ into at most $c_\gamma\sqrt{l}$ segments, one of those must have a length at least $c_\gamma^{-1}\sqrt{l}$. Let us only consider this "long" segment:

The fact that $m_i \geq 1 - 1/\sqrt{l}$ on this segment implies the sign sequence $s_{p_n+2}, \ldots, s_{n/2}$ must be constant on this segment (cf. Proof of Lemma 3). The set of sequences of length l which have a run of the same sign for an interval of length $c_\gamma^{-1}\sqrt{l}$ has probability at most $2l \cdot 2^{-c_\gamma^{-1}\sqrt{l}}$. However, by assumption

$l = \frac{n}{2} - p_n - 1$ goes to infinity as n gets large. Hence the probability of having such a $\mathbf{s}$ sequence gets arbitrarily small when n gets large. ∎

Lemma 8. *For any sequence p_n such that $\lim_{n\to\infty} \frac{n}{2} - p_n = \infty$ and any fixed $\alpha > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left[\forall i > \frac{n}{2} : \min_{s_i} \chi\left(\mathbf{S}^{i-1}, s_i, p_n\right) \le 2^{-4(1+\alpha)}\right] = 1.$$

Proof: Let

$$\mathcal{G}_R(n) \triangleq \left\{\forall i \ge \frac{n}{2} : \rho_{i,p_n}^{(\mathbf{S}^i,*)} \le 2^{-(5+4\alpha)}\right\}.$$

Observe that Lemma 7 implies for any $\delta > 0$ there exist n_0 such that $\mathbb{P}[\mathcal{G}_R(n)] \ge 1 - \delta/2$ for $n \ge n_0$.

Let

$$\mathcal{G}_Z(n) \triangleq \left\{\forall i \ge \frac{n}{2} : Z_i^{(\mathbf{S})} \notin \left[2^{-(11+8\alpha)}, 1 - 2^{-(11+8\alpha)}\right]\right\}.$$

Likewise, the convergence of Z process implies that there exist a n_1 such that for any $n \ge n_1$ $\mathbb{P}[\mathcal{G}_Z(n)] \ge 1 - \delta/2$.

Now (7a) and (7b) imply that for $\mathbf{S} \in \mathcal{G}_R(n) \cap \mathcal{G}_Z(n)$, $\forall i > \frac{n}{2}$, either $\chi\left(\mathbf{S}^{i-1}, +, p_n\right) \le 2^{-4(1+\alpha)}$ or $\chi\left(\mathbf{S}^{i-1}, -, p_n\right) \le 2^{-4(1+\alpha)}$. For $n \ge \max\{n_0, n_1\}$, $\mathbb{P}[\mathcal{G}_R(n) \cap \mathcal{G}_Z(n)] \ge 1 - \delta$ which proves the claim. ∎

Lemma 9. *Fix $\alpha > 0$ and let $m_n \triangleq 4\log\big(2(1+\alpha)n - 1\big)$ (as in Lemma 5). Then:*

$$\lim_{n\to\infty} \mathbb{P}\left[\max_{\mathbf{t}\neq\mathbf{S}:|\mathrm{CP}[\mathbf{S},\mathbf{t}]|<m_n} \rho_n^{(\mathbf{S},\mathbf{t})} \le 2^{-(1+\alpha)n}\right] = 1$$

Proof: For any p, let us define the random variable $B_{n,p} \triangleq \mathbb{1}\left[S_n = \arg\min_s \chi\left(\mathbf{S}^{n-1}, s, p\right)\right]$. It is easy to see that $\mathbb{P}[B_{n,p} = 1|\mathcal{F}_{n-1}] = \mathbb{P}[B_{n,p} = 0|\mathcal{F}_{n-1}] = \frac{1}{2}$.

Fix $\varepsilon > 0$ and let

$$\mathcal{G}_B(n, p, \varepsilon) \triangleq \left\{\frac{1}{n/2}\sum_{i=n/2+1}^{n} B_{i,p} \ge \frac{1-\varepsilon}{2}\right\}.$$

Observe that $\mathbb{P}[\mathcal{G}_B(n, p, \varepsilon)]$ is independent of p and by the Weak Law of Large Numbers for any $\delta > 0$ there exist a n_0 such that $\mathbb{P}[\mathcal{G}_B(n, p, \varepsilon)] \ge 1 - \delta/2$ for $n \ge n_0$.

Fix $\alpha' > 0$ and define

$$\mathcal{G}_\chi(n) \triangleq \left\{i \ge \frac{n}{2} : \min_{s_i} \chi(\mathbf{S}^{i-1}, s_i, m_n) \le 2^{-4(1+\alpha')}\right\}$$

As $\lim_{n\to\infty} \frac{n}{2} - m_n = \infty$, in view of Lemma 8, there exist n_1 such that $\mathbb{P}[\mathcal{G}_\chi(n)] \ge 1 - \delta/2$ for $n \ge n_1$.

For $n \ge \max\{n_0, n_1\}$, $\mathbb{P}[\mathcal{G}_B(n, m_n, \varepsilon) \cap \mathcal{G}_\chi(n)] \ge 1 - \delta$ and for $\mathbf{S}^n \in \mathcal{G}_B(n, m_n, \varepsilon) \cap \mathcal{G}_\chi(n)$ and any $\mathbf{t}^n \neq \mathbf{S}^n$ such that $|\mathrm{CP}[\mathbf{S}^n, \mathbf{t}^n]| < m_n$ we have:

$$\begin{aligned}
\log\left(\rho_n^{(\mathbf{S}^n,\mathbf{t}^n)}\right) &\le \log\left(\rho_{n/2}^{(\mathbf{S}^{n/2},\mathbf{t}^{n/2})}\right) \\
&\quad + \sum_{i=n/2+1}^{n} \log\left(\chi\left(\mathbf{S}^{i-1}, S_i, m_n\right)\right) \\
&\overset{(*)}{\le} \sum_{i=n/2+1}^{n} -4(1+\alpha')B_{i,m_n} \\
&\le -n(1-\varepsilon)(1+\alpha').
\end{aligned}$$

In the above, (*) follows from the fact that $0 \ge \rho_n^{(\mathbf{s},\mathbf{t})} \le 1$ and observing that if $B_{i,m_n} = 1$ then $\chi(\mathbf{S}^{i-1}, S_i, m_n) \le 2^{-4(1+\alpha')}$ (as $\mathbf{S} \in \mathcal{G}_\chi(n)$), otherwise $\chi(\mathbf{S}^{i-1}, S_i, m_n) \le 1$ hence:

$$\log\left(\chi\left(\mathbf{S}^{i-1}, S_i, m_n\right)\right) \le -4(1+\alpha')B_{i,m_n}.$$

For $\mathbf{S} \in \mathcal{G}_B(n, m_n, \varepsilon)$, $\sum_{i=n/2+1}^{n} B_{i,m_n} \ge \frac{n(1-\varepsilon)}{4}$.

Choosing α' and ε such that$(1-\varepsilon)(1+\alpha') \ge (1+\alpha)$ proves the claim. ∎

Theorem 1. *For any $\alpha > 0$.*

$$\lim_{n\to\infty} \mathbb{P}\left[\max_{\mathbf{t}\neq\mathbf{S}} \rho_n^{(\mathbf{S},\mathbf{t})} \le 2^{-n(1+\alpha)}\right] = 1. \tag{9}$$

Proof: The proof follows by combining the results of Lemma 5 and Lemma 9. ∎

VI. Lower-Bound on The Block-Error Probability of Polar Codes

In this section, we use our results on correlations among polarized BECs to give lower-bounds on block error probability of Polar Codes over BEC. Recall the analysis of error of the code: The error event $\mathcal{E}$ is the union of error events in each of the information channels: $\mathcal{E} = \bigcup_{\mathbf{s}\in\mathcal{A}} \mathcal{E}_\mathbf{s}$ where $\mathcal{A} \subset \{-,+\}^n$ is the set of information bits and $\mathcal{E}_\mathbf{s}$ denotes the error in the synthetic channel $W_n^{(\mathbf{s})}$.

For a BEC — with a pessimistic assumption on decoder — a decision error happens exactly when an erasure happens. [1] Therefore, $\mathcal{E}_\mathbf{s} = \left\{E_n^{(\mathbf{s})} = 1\right\}$ and the union bound gives us:

$$\mathbb{P}[\mathcal{E}] \le \sum_{\mathbf{s}\in\mathcal{A}} Z_n^{(\mathbf{s})} \tag{10}$$

A trivial lower-bound on the probability of decoding error is obtained by observing that $\mathcal{E} \supseteq \mathcal{E}_\mathbf{s}$, hence, $\mathbb{P}[\mathcal{E}] \ge \mathbb{P}[\mathcal{E}_\mathbf{s}]$ for any $\mathbf{s} \in \mathcal{A}$. In particular,

$$\mathbb{P}[\mathcal{E}] \ge \max_{\mathbf{s}\in\mathcal{A}} \mathbb{P}[\mathcal{E}_\mathbf{s}] = \max_{\mathbf{s}\in\mathcal{A}} Z_n^{(\mathbf{s})}. \tag{11}$$

However, having the second order statistics, one can use the inclusion–exclusion principle to obtain a much tighter lower-bound on probability of error.

Lemma 10. *Let W be a $\mathsf{BEC}(\epsilon)$ and $\mathcal{C}_n$ be a polar code of block-length $N = 2^n$ with information bits $\mathcal{A}_n$. The block error probability of such a code, $P_e(\mathcal{C}_n)$ is lower-bounded as:*

$$\begin{aligned}
P_e(\mathcal{C}_n) \ge \sum_{\mathbf{s}\in\mathcal{A}_n} Z_n^{(\mathbf{s})} - \frac{1}{2}\sum_{\substack{\mathbf{s},\mathbf{t}\in\mathcal{A}_n:\\ \mathbf{s}\neq\mathbf{t}}} \Bigg[& Z_n^{(\mathbf{s})} Z_n^{(\mathbf{t})} \\
& + \rho_n^{(\mathbf{s},\mathbf{t})} \sqrt{Z_n^{(\mathbf{s})}\overline{Z_n^{(\mathbf{s})}}}\sqrt{Z_n^{(\mathbf{t})}\overline{Z_n^{(\mathbf{t})}}}\Bigg]
\end{aligned} \tag{12}$$

where $\mathbf{Z}_n$ vector and ρ_n matrix can be computed via single-step recursions explained in Section III.

[1] A practical decoder can break the ties randomly which increases the chance of correctly decoding the bit to $\frac{1}{2}$. An analysis analogous to the one we do in this section applies to such a decoder.

Proof: The result follows by applying the inclusion–exclusion principle to lower-bound the probability of $\bigcup_{\mathbf{s}\in\mathcal{A}_n}\mathcal{E}_{\mathbf{s}}$. ■

While the lower-bound given by Lemma 10 is already useful in practice (see Section VII), we seek for a lower-bound that is theoretically more significant.

Theorem 2. *Let W be a $\mathsf{BEC}(\epsilon)$ and $R<1-\epsilon$. Let $\mathcal{C}_n$ be a polar code of block length $N=2^n$ with information bits $\mathcal{A}_n$ such that $|\mathcal{A}_n|=\lceil NR\rceil$. Let $P(N,R,\epsilon)$ be the sum of $\lceil NR\rceil$ smallest elements of the vector $\mathbf{Z}_n$. Then, for any fixed $\delta>0$ and sufficiently large n:*

$$(1-\delta)\,P\left(N,(1-\delta)R,\epsilon\right)\leq P_e(\mathcal{C}_n)\leq P(N,R,\epsilon).$$

Proof: The upper-bound is already known and we only need to prove the lower-bound. Let

$$\mathcal{D}_n=\left\{\mathbf{s}\in\{-,+\}^n:\max_{\mathbf{t}\neq\mathbf{s}}\rho_n^{(\mathbf{s},\mathbf{t})}\leq\delta 2^{-n}\right\}$$

By Theorem 1 we know that $\lim_{n\to\infty}\frac{|\mathcal{D}_n|}{N}=1$. Let, $\mathcal{C}'_n$ be the polar code defined by the information bits $\mathcal{A}'_n=\mathcal{A}_n\cap\mathcal{D}_n$ and $S'_n\triangleq\sum_{\mathbf{s}\in\mathcal{A}'_n}Z_n^{(\mathbf{s})}$. It is clear that $\lim_{n\to\infty}\frac{|\mathcal{A}'_n|}{|\mathcal{A}_n|}=1$, $S'_n\leq P(N,R,\epsilon)$ (as $\mathcal{A}_n$ contains $\lceil NR\rceil$ smallest elements of $\mathbf{Z}_n$), and $P_e(\mathcal{C}'_n)\leq P_e(\mathcal{C}_n)$ as $\mathcal{C}'_n$ is a sub-code of $\mathcal{C}_n$.

Choose n large enough such that $\frac{|\mathcal{A}'_n|}{|\mathcal{A}_n|}\geq 1-\delta$ and $P(N,R,\epsilon)\leq\delta$ (note that this is possible since $R<1-\epsilon$ and the results of [2] suggest that $P(N,R,\epsilon)=O\left(2^{-\sqrt{N}}\right)$). By (12):

$$\begin{aligned}S'_n-P_e(\mathcal{C}_n)&\leq S'_n-P_e(\mathcal{C}'_n)\\&\leq\frac{1}{2}\sum_{\substack{\mathbf{s},\mathbf{t}\in\mathcal{A}'_n:\\ \mathbf{s}\neq\mathbf{t}}}\left[Z_n^{(\mathbf{s})}Z_n^{(\mathbf{t})}+\rho_n^{(\mathbf{s},\mathbf{t})}\sqrt{Z_n^{(\mathbf{s})}\overline{Z_n^{(\mathbf{s})}}}\sqrt{Z_n^{(\mathbf{t})}\overline{Z_n^{(\mathbf{t})}}}\right].\end{aligned}$$

Observe that $\rho_n^{(\mathbf{s},\mathbf{t})}\leq\delta/N$ for all $\mathbf{s},\mathbf{t}$ in the above summation, $\sum_{\mathbf{s},\mathbf{t}\in\mathcal{A}'_n:\mathbf{s}\neq\mathbf{t}}Z_n^{(\mathbf{s})}Z_n^{(\mathbf{t})}\leq\sum_{\mathbf{s},\mathbf{t}\in\mathcal{A}'_n}Z_n^{(\mathbf{s})}Z_n^{(\mathbf{t})}={S'_n}^2$, and

$$\begin{aligned}&\sum_{\mathbf{s},\mathbf{t}\in\mathcal{A}'_n:\mathbf{s}\neq\mathbf{t}}\sqrt{Z_n^{(\mathbf{s})}\overline{Z_n^{(\mathbf{t})}}}\sqrt{Z_n^{(\mathbf{t})}\overline{Z_n^{(\mathbf{t})}}}\\&\leq\sum_{\mathbf{s},\mathbf{t}\in\mathcal{A}'_n:\mathbf{s}\neq\mathbf{t}}\sqrt{Z_n^{(\mathbf{s})}}\sqrt{Z_n^{(\mathbf{t})}}\leq\sum_{\mathbf{s},\mathbf{t}\in\mathcal{A}'_n}\sqrt{Z_n^{(\mathbf{s})}}\sqrt{Z_n^{(\mathbf{t})}}\\&=\left[\sum_{\mathbf{s}\in\mathcal{A}'_n}\sqrt{Z_n^{(\mathbf{s})}}\right]^2\overset{(*)}{\leq}|\mathcal{A}'_n|\sum_{\mathbf{s}\in\mathcal{A}'_n}Z_n^{(\mathbf{s})}\leq NS'_n,\end{aligned}$$

where (*) follows by the Cauchy-Schwarz inequality [2].

Therefore,

$$S'_n-P_e(\mathcal{C}_n)\leq\frac{1}{2}\left[{S'_n}^2+\delta S'_n\right]\leq\delta S'_n,$$

[2]For any set of m numbers $x_i, i=1,2,\cdots,m$:

$$\left(\textstyle\sum_{i=1}^m x_i\right)^2\leq m\sum_{i=1}^m x_i^2$$

.

R	$\sum_{\mathbf{s}\in\mathcal{A}_n}Z_n^{(\mathbf{s})}$	$\max_{\mathbf{s}\in\mathcal{A}_n}Z_n^{(\mathbf{s})}$	Lower-bound (12)
0.2	$4.04\cdot10^{-18}$	$3.43\cdot10^{-19}$	$4.04\cdot10^{-18}$
0.25	$1.87\cdot10^{-11}$	$9.25\cdot10^{-13}$	$1.87\cdot10^{-11}$
0.3	$5.4\cdot10^{-7}$	$2.29\cdot10^{-8}$	$5.4\cdot10^{-7}$
0.35	$8.14\cdot10^{-4}$	$2.11\cdot10^{-5}$	$8.12\cdot10^{-4}$
0.4	0.17	$3.49\cdot10^{-3}$	0.14

(a) $N=4096$

R	$\sum_{\mathbf{s}\in\mathcal{A}_n}Z_n^{(\mathbf{s})}$	$\max_{\mathbf{s}\in\mathcal{A}_n}Z_n^{(\mathbf{s})}$	Lower-bound (12)
0.2	$9.32\cdot10^{-36}$	$4.72\cdot10^{-37}$	$9.32\cdot10^{-36}$
0.25	$1.32\cdot10^{-22}$	$3.54\cdot10^{-24}$	$1.32\cdot10^{-22}$
0.3	$2.32\cdot10^{-13}$	$5.4\cdot10^{-15}$	$2.32\cdot10^{-13}$
0.35	$2.63\cdot10^{-7}$	$3.61\cdot10^{-9}$	$2.63\cdot10^{-7}$
0.4	$5.47\cdot10^{-3}$	$4.91\cdot10^{-5}$	$5.43\cdot10^{-3}$

(b) $N=16384$

TABLE I: Bounds on Block Error Probability of Polar Code on $\mathsf{BEC}(0.5)$

where the last inequality follows by observing that $S'_n\leq P(N,R,\epsilon)\leq\delta$. As a result,

$$(1-\delta)S'_n\leq P_e(\mathcal{C}_n)$$

Observe that $\mathcal{C}'_n$ is a code of rate $R'\geq(1-\delta)R$ and by definition $S'_n\geq P(N,R',\epsilon)\geq P\left(N,(1-\delta)R,\epsilon\right)$. Hence we can lower-bound the LHS of the above by substituting S'_n with $P\left(N,(1-\delta)R,\epsilon\right)$ which completes the proof. ■

VII. Numerical Results

In this section we provide a numerical example which confirms our theoretical results. We have considered Polar Codes of different rates on a $\mathsf{BEC}(0.5)$ and computed the upper-bound of (10), the trivial lower-bound of (11) and the tighter lower-bound of (12). We emphasize that we have exactly computed the lower-bound on the error probability by computing the correlation coefficients. We did the computations for block lengths of $N=4096$ ($n=12$) and $N=16384$ ($n=14$).

As shown in Table I, the proposed lower bound is much tighter than the trivial one. Moreover, the results show that the lower bound is very close to the upper bound of (10). This confirms that $P(N,R,\epsilon)$ (as defined in Theorem 2) is indeed a very good estimation for the block error probability of Polar Codes over BEC.

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, no. 7, pp. 3051 –3073, July 2009.

[2] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *IEEE International Symposium on Information Theory (ISIT), 2009*, July 2009, pp. 1493 –1495.

[3] M. Bastani Parizi and E. Telatar, "On the correlation between polarized BECs," *ArXiv e-prints*, Jan. 2013. [Online]. Available: http://arxiv.org/abs/1301.5536