On the rate of channel polarization

Erdal Arıkan
Department of Electrical-Electronics Engineering
Bilkent University
Ankara, TR-06800, Turkey
Email: arikan@ee.bilkent.edu.tr

Emre Telatar
Information Theory Laboratory
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: emre.telatar@epfl.ch

***Abstract*—A bound is given on the rate of channel polarization. As a corollary, an earlier bound on the probability of error for polar coding is improved. Specifically, it is shown that, for any binary-input discrete memoryless channel W with symmetric capacity $I(W)$ and any rate $R < I(W)$, the polar-coding block-error probability under successive cancellation decoding satisfies $P_e(N,R) \leq 2^{-N^\beta}$ for any $\beta < \frac{1}{2}$ when the block-length N is large enough.**

I. Results

Channel polarization is a method introduced in [1] for constructing capacity-achieving codes on symmetric binary-input memoryless channels. Both the construction and the probability of error analysis of polar codes, as these codes were called, are centered around a random process $\{Z_n : n \in \mathbb{N}\}$ which keeps track of the Bhattacharyya parameters of the channels that arise in the course of channel polarization. The aim here is to give an asymptotic convergence result on $\{Z_n\}$ in as simple a setting as possible. For further background on the problem, we refer to [1].

For the purposes here, the polarization process can be modeled as follows. Suppose B_i, $i = 1, 2, \ldots$, are i.i.d., $\{0,1\}$-valued random variables with

$$P(B_1 = 0) = P(B_1 = 1) = \frac{1}{2}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$. Set $\mathcal{F}_0 = \{\emptyset, \Omega\}$ as the trivial σ-algebra and set $\mathcal{F}_n$, $n \geq 1$, to be the σ-algebra generated by $(B_1, \ldots, B_n)$. We may assume that $\mathcal{F} = \bigcup_{n\geq 0} \mathcal{F}_n$.

Suppose further that a stochastic process $\{Z_n : n \in \mathbb{N}\}$ is defined on this probability space with the following properties:

(z.1) For each $n \in \mathbb{N}$, Z_n takes values in the interval $[0,1]$ and is measurable with respect to $\mathcal{F}_n$. That is, Z_0 is constant, and Z_n is a function of $B_1, \ldots, B_n$.

(z.2) For some constant q and for each $n \in \mathbb{N}$,

$$\begin{aligned} Z_{n+1} &= Z_n^2 && \text{when } B_{n+1} = 1, \\ Z_{n+1} &\leq q Z_n && \text{when } B_{n+1} = 0. \end{aligned}$$

(z.3) $\{Z_n\}$ converges a.s. to a $\{0,1\}$-valued random variable Z_∞ with $P(Z_\infty = 0) = I_0$ for some $I_0 \in [0,1]$.

The main result of this note is that whenever $\{Z_n\}$ converges to zero, this converges is almost surely fast:

Theorem 1: For any $\beta < 1/2$,

$$\lim_{n\to\infty} P\big(Z_n < 2^{-2^{n\beta}}\big) = I_0. \tag{1}$$

Remark 1: The random process $\{Z_n : n \in \mathbb{N}\}$ considered in [1] satisfies the properties (z.1)–(z.3) with $q = 2$ and $I_0 = I(W)$ where $I(W)$ denotes the symmetric capacity of the underlying channel W. The framework in this note is held more general than in [1] in anticipation of the results here being applicable to more general channel polarization scenarios.

Remark 2: Clearly, the statement of the theorem remains valid if we replace $2^{-2^{n\beta}}$ with $\alpha^{-2^{n\beta}}$ for any $\alpha > 1$.

Remark 3: As a corollary to Theorem 1, the result of [1] on the probability of block-error for polar coding under successive cancellation decoding is strengthened as follows.

Theorem 2: Let W be any B-DMC with $I(W) > 0$. Let $R < I(W)$ and $\beta < \frac{1}{2}$ be fixed. Then, for $N = 2^n$, $n \geq 0$, the block error probability for polar coding under successive cancellation decoding at block length N and rate R satisfies

$$P_e(N,R) = \mathcal{O}\big(2^{-N^\beta}\big).$$

In comparison, the result in [1] was that for $R < I(W)$

$$P_e(N,R) = \mathcal{O}(N^{-\frac{1}{4}}).$$

Remark 4: The polarization process $\{Z_n\}$ considered in [1] satisfies the additional condition that $Z_{n+1} \geq Z_n$ when $B_{n+1} = 0$. Under this condition, Theorem 1 has the following converse.

Theorem 3: If the condition (z.2) in the definition of $\{Z_n : n \in \mathbb{N}\}$ is replaced with the condition that

$$\begin{aligned} Z_{n+1} &= Z_n^2 && \text{when } B_{n+1} = 1, \\ Z_{n+1} &\geq Z_n && \text{when } B_{n+1} = 0, \end{aligned}$$

and if $Z_0 > 0$, then for any $\beta > 1/2$,

$$\lim_{n\to\infty} P\big(Z_n < 2^{-2^{n\beta}}\big) = 0. \tag{2}$$

In the rest of this note, we prove Theorems 1 and 3. We leave out the proof of Theorem 2 since it follows readily from the existing results in [1].

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

II. Proof of Theorem 1

Lemma 1: Let $A : \mathbb{R} \to \mathbb{R}$, $A(x) = x + 1$ denote adding one, and $D : \mathbb{R} \to \mathbb{R}$, $D(x) = 2x$ denote doubling. Suppose a sequence of numbers $a_0, a_1, \dots, a_n$ is defined by specifying a_0 and the recursion

$$a_{i+1} = f_i(a_i)$$

with $f_i \in \{A, D\}$. Suppose $\big|\{0 \le i \le n-1 : f_i = D\}\big| = k$ and $\big|\{0 \le i \le n-1 : f_i = A\}\big| = n-k$, i.e., during the first n iterations of the recursion we encounter doubling k times and adding-one $n-k$ times. Then

$$a_n \le D^{(k)}\big(A^{(n-k)}(a_0)\big) = 2^k(a_0 + n - k).$$

Proof: Observe that the upper bound on a_n corresponds to choosing

$$f_0 = \cdots f_{n-k-1} = A \quad \text{and} \quad f_{n-k} = \cdots = f_{n-1} = D.$$

We will show that any other choice of $\{f_i\}$ can be modified to yield a higher value of a_n. To that end suppose $\{f_i\}$ is not chosen as above. Then there exists $j \in \{1, \dots, n-1\}$ for which $f_{j-1} = D$ and $f_j = A$. Define $\{f'_i\}$ by swapping f_j and f_{j-1}, i.e.,

$$f'_i = \begin{cases} A & i = j-1 \\ D & i = j \\ f_i & \text{else} \end{cases}$$

and let $\{a'_i\}$ denote the sequence that results from $\{f'_i\}$. Then

$$\begin{aligned} a'_i &= a_i \quad \text{for } i < j \\ a'_j &= a_{j-1} + 1 \\ a'_{j+1} &= 2a'_j = 2a_{j-1} + 2 \\ &> 2a_{j-1} + 1 = a_{j+1}. \end{aligned}$$

Since the recursion from $j+1$ onwards is identical for the $\{f_i\}$ and $\{f'_i\}$ sequences, and since both A and D are order preserving, $a'_{j+1} > a_{j+1}$ implies that $a'_n > a_n$. ∎

Lemma 2: For any $\epsilon > 0$ there exists an m such that

$$P\big(Z_n \le 1/q^2 \text{ for all } n \ge m\big) > I_0 - \epsilon.$$

Proof: Let $\Omega_0 = \{\omega : Z_n(\omega) \to 0\}$. Recall that by (z.3) $P(\Omega_0) = I_0$. Since for non-negative sequences, "$a_n \to 0$" is the same as "for all $k \ge 1$ there exists n_0 such that for all $n \ge n_0$, $a_n < 1/k$," we have

$$\Omega_0 = \bigcap_{k \ge 1} \bigcup_{n_0 \ge 1} A_{n_0,k}$$

where $A_{n_0,k} := \big\{\omega : \text{for all } n \ge n_0,\ Z_n(\omega) < 1/k\big\}$. Thus, for any choice of k, Ω_0 is included in $\bigcup_{n_0} A_{n_0,k}$, and for $k = q^2$,

$$I_0 = P(\Omega_0) \le P\Bigg(\bigcup_{n_0 \ge 1} A_{n_0,q^2}\Bigg).$$

Since A_{n_0,q^2} is increasing in n_0, for any $\epsilon > 0$ there is an m so that

$$P\big(A_{m,q^2}\big) > P\Bigg(\bigcup_{n_0 \ge 1} A_{n_0,q^2}\Bigg) - \epsilon \ge I_0 - \epsilon. \qquad ∎$$

Lemma 3: For any $\epsilon > 0$ there is an n_0 such that whenever $n \ge n_0$

$$P\big(\log_q Z_n \le -n/10\big) > I_0 - \epsilon.$$

Proof: Define $S_n = \sum_{i=1}^n B_i$. Define $G_{m,n,\alpha}$ as the event

$$S_n - S_m \ge \alpha(n-m)$$

i.e., the event that the slice $\{B_i : i = m+1, \dots, n\}$ contains more than an α fraction of ones. Note that for any $\alpha < 1/2$, whenever $n - m$ is large, this event has probability close to 1; formally, for any $\alpha < 1/2$ and $\epsilon > 0$ there is $n_0 = n_0(\epsilon, \alpha)$ such that $P(G_{m,n,\alpha}) > 1 - \epsilon$ whenever $n - m \ge n_0$. Let $A_m := \{\omega : Z_n(\omega) < 1/q^2 \text{ for all } n \ge m\}$. Given $\epsilon > 0$, find $m = m(\epsilon)$ such that $P(A_m) > I_0 - \epsilon/2$. Such an m exists by Lemma 2.

Note that for $\omega \in A_m$, and $n \ge m$, we have

$$\begin{aligned} Z_{n+1} &= Z_n^2 \le Z_n/q^2 && \text{when } B_{n+1} = 1, \\ Z_{n+1} &\le qZ_n && \text{when } B_{n+1} = 0. \end{aligned}$$

Considering $\log_q Z_n$, we get

$$\begin{aligned} \log_q Z_{n+1} &\le \log_q Z_n - 2 && \text{when } B_{n+1} = 1, \\ \log_q Z_{n+1} &\le \log_q Z_n + 1 && \text{when } B_{n+1} = 0. \end{aligned}$$

Consequently,

$$\begin{aligned} \log_q Z_n &\le \log_q Z_m - 2(S_n - S_m) + (n - m - (S_n - S_m)) \\ &\le -3(S_n - S_m) + (n - m). \end{aligned}$$

Now find $n_0 \ge 2m$ such that whenever $n \ge n_0$, $P(G_{m,n,2/5}) > 1 - \epsilon/2$. Then for any $n \ge n_0$, for $\omega \in A_m \cap G_{m,n,2/5}$ we have

$$\log_q Z_n \le -(n-m)/5 \le -n/10.$$

Noting that $P\big(A_m \cap G_{m,n,2/5}\big) > I_0 - \epsilon$, the proof is completed. ∎

Proof of Theorem 1. Given $\beta < 1/2$, fix $\beta' \ge 1/3$ and $\beta' \in (\beta, 1/2)$. Choose $n_3(\epsilon)$ such that with $n_2(\epsilon) := 3\log_2 n_3(\epsilon)$ and $n_1(\epsilon) := 20\, n_2(\epsilon)$, we have

(i) $n_1(\epsilon) \ge 40$ and $n_1(\epsilon) \ge n_0(\epsilon/3)$ where n_0 is as in Lemma 3,
(ii) $P\big(G_{n_1(\epsilon), n_1(\epsilon)+n_2(\epsilon), \beta'}\big) > 1 - \epsilon/3$,
(iii) $P\big(G_{n_1(\epsilon)+n_2(\epsilon), n_3(\epsilon), \beta'}\big) > 1 - \epsilon/3$,
(iv) $\beta'(n_3(\epsilon) - n_1(\epsilon) - n_2(\epsilon)) \ge \beta n_3(\epsilon) + \log_2(\log_q(2))$.

Given $n \ge n_3(\epsilon)$ set $n_2 = 3\log_2 n$ and $n_1 = 20\, n_2$. Observe that (i)–(iv) are satisfied with (n_1, n_2, n) in place of $(n_1(\epsilon), n_2(\epsilon), n_3(\epsilon))$. Let

$$G = \big\{\log_q Z_{n_1} \le -n_1/10\big\} \cap G_{n_1, n_1+n_2, \beta'} \cap G_{n_1+n_2, n, \beta'}.$$

Note that $P(G) > I_0 - \epsilon$. Observe that the process $\{\log_q Z_i : i \ge n_1\}$ is upper bounded by the process $\{L_i : i \ge n_1\}$ defined by $L_{n_1} = \log_q Z_{n_1}$ and for $i \ge n_1$

$$\begin{aligned} L_{i+1} &= 2L_i && \text{when } B_{i+1} = 1, \\ L_{i+1} &= L_i + 1 && \text{when } B_{i+1} = 0. \end{aligned}$$

For $\omega \in G$ we have

(a) $L_{n_1} \leq -n_1/10$,
(b) during the evolution of L_i from time n_1 to n_1+n_2 there are at least $\beta' n_2$ doublings,
(c) during the evolution of L_i from time n_1+n_2 to n there are at least $\beta'(n-n_1-n_2)$ doublings.

By Lemma 1 we obtain

$$\begin{aligned} L_{n_1+n_2} &\leq 2^{\beta' n_2}(L_{n_1}+n_2) \\ &\leq 2^{\beta' n_2}(-n_1/10+n_2) \\ &\leq -2^{\beta' n_2} n_1/20 \end{aligned}$$

and

$$\begin{aligned} L_n &\leq 2^{\beta'(n-n_1-n_2)}\big(L_{n_1+n_2}+(n-n_1-n_2)\big) \\ &\leq 2^{\beta'(n-n_1-n_2)}\big(-2^{\beta' n_2} n_1/20+n\big) \\ &\leq 2^{\beta'(n-n_1-n_2)}\big(-2^{n_2/3} n_1/20+n\big) \\ &\leq 2^{\beta'(n-n_1-n_2)}\big(-n(n_1/20-1)\big) \\ &\leq -n2^{\beta'(n-n_1-n_2)} \\ &\leq -2^{\beta'(n-n_1-n_2)} \\ &\leq -(\log_q(2))^{\beta n}. \end{aligned}$$

This implies that $Z_n \leq 2^{-2^{\beta n}}$ on a set of probability at least $I_0 - \epsilon$ whenever $n \geq n_3(\epsilon)$, completing the proof.

III. Proof of Theorem 3

Let $\{Z_n : n \in \mathbb{N}\}$ be a process satisfying the hypothesis of Theorem 3. Observe that the random process $\{\log_2\big(-\log_2(Z_n)\big) : n \in \mathbb{N}\}$ is upper bounded by the process $\{K_n : n \in \mathbb{N}\}$ defined by $K_0 := \log_2(-\log_2(Z_0))$ and for $n \geq 1$

$$K_n := K_{n-1} + B_n = K_0 + \sum_{i=1}^{n} B_i.$$

So, we have

$$\begin{aligned} P\big(Z_n \leq 2^{-2^{\beta n}}\big) &= P\big(\log_2\big(-\log_2(Z_n)\big) \geq \beta n\big) \\ &\leq P(K_n \geq \beta n) \\ &= P\left(\sum_{i=1}^{n} B_i \geq n\beta - K_0\right). \end{aligned}$$

For $\beta > \frac{1}{2}$, this last probability goes to zero as n increases by the law of large numbers.

IV. Concluding remarks

In an earlier version of this note [2], Theorem 1 was proved using the following inequality due to Hajek [3] in place of Lemma 2.

Lemma 4: Suppose $\{Z_n : n \in \mathbb{N}\}$ satisfies the conditions (z.1)-z(3) with (z.2) replaced with:

(z.2) For each $n \in \mathbb{N}$,

$$\begin{aligned} Z_{n+1} &= Z_n^2 && \text{when } B_{n+1} = 1, \\ Z_{n+1} &= Z_n^2 - 2Z_n && \text{when } B_{n+1} = 0. \end{aligned}$$

Then $E\big[\sqrt{Z_n(1-Z_n)}\,\big] \leq \frac{1}{2}\big(\frac{3}{4}\big)^{n/2}$.

The present proof is more direct and simpler than the one in [2].

In recent work, Korada et al. generalized the above rate of channel polarization results as part of a study where they considered more general forms of polar code constructions [4]. There $\{B_i : i = 1, 2, \ldots\}$ were taken as i.i.d., $\{0, 1, \ldots, \ell-1\}$-valued random variables with

$$P(B_1 = i) = \frac{1}{\ell}, \qquad i = 0, \ldots, \ell - 1,$$

for some $\ell \geq 2$. The random process $\{Z_n : n \in \mathbb{N}\}$ was defined with the properties (z.1) and (z.3) as in here, but with (z.2) modified as:

(z.2) For each $n \in \mathbb{N}$ and $i = 0, \ldots, \ell - 1$,

$$Z_n^{D_i} \leq Z_{n+1} \leq 2^{\ell - i} Z_n^{D_i} \qquad \text{when } B_{n+1} = i$$

where $\{D_i : 0 \leq i \leq \ell - 1\}$ are a set of positive constants.

The following result was proved in [4].

Theorem 4: Let $E := \frac{1}{\ell}\sum_{i=0}^{\ell-1} \log_\ell D_i$. Then,

$$\begin{aligned} \lim_{n\to\infty} P(Z_n < 2^{-\ell^{n\beta}}) &= I_0 && \text{when } \beta < E, \\ \lim_{n\to\infty} P(Z_n < 2^{-\ell^{n\beta}}) &= 0 && \text{when } \beta > E. \end{aligned}$$

An open problem that remains is to obtain a more refined bound on the rate of channel polarization. Specifically, it would be of interest to find a function $\gamma : \mathbb{N} \times [0,1] \to [0,1]$ such that for any given $R \in [0,1]$

$$\lim_{n\to\infty} P(Z_n \leq \gamma(n, R)) = R.$$

Acknowledgment

This work was supported in part by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under contracts no. 105E065 and 107E216, and in part by the European Commission FP7 Network of Excellence NEWCOM++ (contract no. 216715).

References

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," submitted to *IEEE Trans. Inform. Theory*, Oct. 2007.
[2] E. Arıkan and E. Telatar, "On the rate of channel polarization," July 2008. [Online]. Available: arXiv:0807.3806v2 [cs.IT]
[3] B. Hajek, June 2007. Private communication.
[4] S. B. Korada, E. Şaşoğlu, R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," Jan 2009. [Online]. Available: arXiv:0901.0536v2 [cs.IT].