A Simple Proof of Fast Polarization

Ido Tal, *Member, IEEE*

***Abstract*—Fast polarization is a key property of polar codes. It was proved for the binary polarizing 2×2 kernel by Arıkan and Telatar. The proof was later adapted to the general case by Şaşoğlu. We give a simplified proof.**

***Index Terms*—polar codes, fast polarization**

I. Introduction

Polar codes are a novel family of error correcting codes, invented by Arıkan [1]. The seminal definitions and assumptions in [1] were soon expanded and generalized. Key to almost all the results involving polar codes is the concept of fast polarization. The essence of fast polarization is the phenomenon stated in the following lemma. The lemma was used implicitly by Korada, Şaşoğlu, and Urbanke [2, proof of Theorem 11], and is a generalization of a result by Arıkan and Telatar [3, Theorem 3]. Its explicit formulation and full proof appear in a monograph by Şaşoğlu [4, Lemma 5.9].

Lemma 1: Let $T_0, T_1, \ldots$ be an i.i.d. process where T_0 is uniformly distributed over $\{1, 2, \ldots, \ell\}$. Let $Z_0, Z_1, \ldots$ be a $[0,1]$-valued random process such that

$$Z_{m+1} \leq K \cdot Z_m^{d_t} \; , \quad \text{whenever } T_m = t \; . \tag{1}$$

We assume $K \geq 1$ and $d_1, d_2, \ldots, d_\ell > 0$. Suppose also that Z_m converges almost surely to a $\{0,1\}$-valued random variable Z_∞. Then, for any

$$0 < \beta < E \triangleq \frac{1}{\ell} \sum_{t=1}^{\ell} \log_\ell d_t$$

we have

$$\lim_{m \to \infty} \Pr[Z_m \leq 2^{-\ell^{\beta \cdot m}}] = \Pr[Z_\infty = 0] \; . \tag{2}$$

The lemma is used to prove that the Bhattacharyya parameter associated with a random variable that underwent polarization (for example, a synthesized channel) polarizes to 0 at a rate faster than polynomial [4, Theorem 5.4]. A similar claim holds in the case of polarization of the Bhattacharyya parameter to 1 [5, Theorem 16].

The original proof [4, Lemma 5.9] of Lemma 1 is somewhat involved. To summarize, if K were equal to 1, the proof would follow almost directly from the weak law of large numbers. However, for $K > 1$, a sequence of bootstrapping arguments is applied to strengthen the bound gradually in each step.

The main aim of this paper is to give a simpler proof of Lemma 1. Thus, we hopefully give insight into the simple mechanics that are at play. Our simpler proof also leads to a stronger result. That is, we will prove the following, which implies Lemma 1.

Lemma 2: Let $\{T_m\}_{m=0}^{\infty}$, $\{Z_m\}_{m=0}^{\infty}$, K, and E be as in Lemma 1. Then, for $0 < \beta < E$,

$$\lim_{m_0 \to \infty} \Pr[Z_m \leq 2^{-\ell^{\beta \cdot m}} \text{ for all } m \geq m_0] = \Pr[Z_\infty = 0] \; . \tag{3}$$

Note that Lemma 2 has an "almost sure flavor" [6, page 69, Equation (2)], while Lemma 1 has an "in probability flavor" [6, page 70, Equation (5)]. We prove Lemma 2 in Section II and show that it implies Lemma 1 in Section III.

II. Proof of Lemma 2

Let $\epsilon_a, \epsilon_b > 0$ and $m_a < m_b$ be parameters. We now define three events, denoted A, B, and C.

$$A: \qquad |Z_m - Z_\infty| \leq \epsilon_a \; , \quad \text{for all } m \geq m_a \; . \tag{4}$$

$$B: \qquad \left| \frac{|\{m_a \leq i < m : T_i = t\}|}{m - m_a} - \frac{1}{\ell} \right| \leq \epsilon_b \; , \quad \text{for all } m \geq m_b \text{ and all } 1 \leq t \leq \ell \; . \tag{5}$$

$$C: \qquad Z_\infty = 0 \; . \tag{6}$$

We first claim that for any fixed $\epsilon_a > 0$,

$$\lim_{m_a \to \infty} \Pr[A] = 1 \; . \tag{7}$$

The above follows immediately from [6, Theorem 4.1.1], but let us elaborate for completeness. By definition of almost sure convergence, the event of Z_m converging to Z_∞ has probability 1. Thus, the event "there exists an m_a for which (4) holds" must have probability 1 as well, since it contains the former event. We now emphasize that the event A is dependent on m_a by adopting to notation $A = A(m_a)$, and note that the previous sentence can be written succinctly as

$$\Pr\left[\bigcup_{m_a=0}^{\infty} A(m_a)\right] = 1 \; .$$

Since we clearly have $A(0) \subseteq A(1) \subseteq A(2) \subseteq \cdots$, we deduce (7) from the above and the monotone property of measures [6, page 21, property (ix)].

Event B is concerned with the frequency of t in the subsequence of i.i.d. random variables $T_{m_a}, T_{m_a+1}, \ldots, T_{m-1}$, each of which is uniform over $\{1, 2, \ldots, \ell\}$. We claim that for any fixed $\epsilon_b > 0$ and $m_a \geq 0$,

$$\lim_{m_b \to \infty} \Pr[B] = 1 \; . \tag{8}$$

To see this, we use the strong law of large numbers. Denote $B = B(m_b)$. Next, we abuse notation and denote by $B(m_b, t)$ the event of (5) holding, but for t fixed (we remove the

This work was supported by the Israel Science Foundation under grant 1769/13.

The author is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000 (email:idotal@ee.technion.ac.il).

sentence "and all $1 \le t \le \ell$" from the definition). Thus, $B(m_b) = \bigcap_{t=1}^{\ell} B(m_b, t)$. Hence, (8) will follow from proving that for $1 \le t \le \ell$ fixed but arbitrary,

$$\lim_{m_b \to \infty} \Pr[B(m_b, t)] = 1 \ . \tag{9}$$

Fix t, and assign to each T_i, $i \ge m_a$, an indicator equalling 1 if and only if T_i equals t. The indicators are i.i.d. and equal 1 with probability $1/\ell$. By the strong law of large numbers [6, Theorem 5.4.2], the fraction of indicators equaling 1 approaches $1/\ell$ almost surely. Hence, so does the fraction of T_i equalling t. We now invoke [6, Theorem 4.1.1], deduce (9), and consequently (8).

By (7) and (8), we deduce that for any $\delta_a, \delta_b > 0$ there exist $m_a < m_b$ such that

$$\Pr[A] \ge 1 - \delta_a \tag{10}$$

and

$$\Pr[B] \ge 1 - \delta_b \ . \tag{11}$$

Hence,

$$\Pr[A \cap B \cap C] \ge \Pr[Z_\infty = 0] - \delta_a - \delta_b \ . \tag{12}$$

Equation (12) will be used towards the end of the proof.

We now focus on the implications of the event $A \cap B \cap C$. Define the shorthand

$$\theta \triangleq -\log_{\epsilon_a} K \ .$$

Note that θ is non-negative, and approaches 0 as ϵ_a approaches 0. By the definition of the events A and C, we have that $Z_m \le \epsilon_a$ when $m \ge m_a$. Thus, $K \le Z_m^{-\theta}$ when $m \ge m_a$. Hence, under the event $A \cap B \cap C$, we can simplify (1) to

$$Z_{m+1} \le Z_m^{d_t - \theta} \ , \quad \text{whenever } m \ge m_a \text{ and } T_m = t \ . \tag{13}$$

The above equation is the heart of the proof: we have effectively managed to "make K equal 1" — the simple case discussed earlier. We have "paid" for this simplification by having the exponents be $d_t - \theta$ instead of the original d_t. However, since θ can be made arbitrarily close to 0, this will not be a problem. Essentially, all that remains is some simple algebra, followed by taking the relevant parameters small/large enough. We do this now.

Events A and C have been put to use and have yielded (13). We will now call on event B. We take ϵ_a small enough such that $d_t - \theta > 0$ for all $1 \le t \le \ell$, and further require that $\epsilon_b < 1/\ell$. Recalling that $Z_{m_a} \in [0,1]$, we repeatedly apply (13) and deduce the following under $A \cap B \cap C$. For all $m \ge m_b$,

$$Z_m \le Z_{m_a}^{\prod_{t=1}^{\ell} (d_t - \theta)^{(m-m_a)\cdot\left(\frac{1}{\ell} \pm \epsilon_b\right)}} \ , \tag{14}$$

where the above "$\pm$" notation is in fact a function of t, defined as

$$\pm \triangleq \begin{cases} + & \text{if } d_t - \theta \le 1, \\ - & \text{otherwise.} \end{cases}$$

By the definition of event A, we have that $Z_{m_a} \le \epsilon_a$. We take $\epsilon_a \le 1/2$. Hence, (14) simplifies to the claim that under $A \cap B \cap C$, for all $m \ge m_b$,

$$Z_m \le 2^{-\prod_{t=1}^{\ell} (d_t - \theta)^{(m-m_a)\cdot\left(\frac{1}{\ell} \pm \epsilon_b\right)}} = 2^{-\ell^{(E-\Delta)m}} \ , \tag{15}$$

where

$$\Delta = \sum_{t=1}^{\ell} \frac{1}{\ell} \log_\ell \left(\frac{d_t}{d_t - \theta} \right) - \sum_{t=1}^{\ell} \pm\epsilon_b \log_\ell (d_t - \theta) + \sum_{t=1}^{\ell} \frac{m_a}{m} \left(\frac{1}{\ell} \pm \epsilon_b \right) \log_\ell (d_t - \theta) \ . \tag{16}$$

In light of (3), our task is now the following. Given $0 < \beta < E$ and $\delta_a, \delta_b > 0$, we must show that there exists a choice of $m_a < m_b$ and $\epsilon_a, \epsilon_b > 0$ such that (12) holds and $\Delta < E - \beta$. Equation (12) will follow from choosing parameters for which (10) and (11) hold. We show that the inequality on Δ holds by showing that each of the three sums in (16) can be made smaller than $(E-\beta)/3$. Recalling that θ goes to 0 as ϵ_a tends to 0, we deduce that the first sum can be made smaller than $(E-\beta)/3$ by taking ϵ_a small enough. Similarly, we can make the second sum smaller than $(E-\beta)/3$ by taking ϵ_b small enough. For the third sum, we first fix m_a large enough such that (10) holds (note that event A is a function of ϵ_a, which is by now fixed). Lastly, we take m_b large enough such that the third sum is smaller than $(E-\beta)/3$ for all $m \ge m_b$, and (11) holds (again, note that event B is a function of m_a and ϵ_b, which have been fixed).

We have just proven the following. Fix $0 < \beta < E$ and $\delta_a, \delta_b > 0$. Denote the event cardinal to (3) as

$$D: \qquad Z_m \le 2^{-\ell^{\beta \cdot m}} \ , \quad \text{for all } m \ge m_0 \ .$$

Then, for $m_a < m_b$ and $\epsilon_a, \epsilon_b > 0$ as above, setting $m_0 = m_b$ results in D containing $A \cap B \cap C$. Thus, by (12),

$$P[D] \ge \Pr[Z_\infty = 0] - \delta_a - \delta_b \ , \quad \text{for } m_0 = m_b \ .$$

Since the probability of D increases with m_0,

$$\lim_{m_0 \to \infty} \Pr[D] \ge \Pr[Z_\infty = 0] - \delta_a - \delta_b \ .$$

The above inequality holds for all $\delta_a, \delta_b > 0$, and so must also hold for $\delta_a = \delta_b = 0$. Thus, to prove (3), all that remains to show is

$$\lim_{m_0 \to \infty} \Pr[D] \le \Pr[Z_\infty = 0] \ .$$

Indeed,

$$\Pr[D] \le \Pr[\lim_{m \to \infty} Z_m = 0] = \Pr[Z_\infty = 0] \ .$$

Thus, the claim is true when taking m_0 to infinity as well.

III. Proof of Lemma 1

We now explain why Lemma 2 implies Lemma 1. That is, why (3) implies (2). Clearly, (3) implies

$$\liminf_{m \to \infty} \Pr[Z_m \le 2^{-\ell^{\beta \cdot m}}] \ge \Pr[Z_\infty = 0] \ .$$

Thus, the claim will follow if we prove that

$$\limsup_{m \to \infty} \Pr[Z_m \le 2^{-\ell^{\beta \cdot m}}] \le \Pr[Z_\infty = 0] \ .$$

Assume to the contrary that there exists $0 < \beta < E$ such that

$$\limsup_{m \to \infty} \Pr[Z_m \le 2^{-\ell^{\beta \cdot m}}] > \Pr[Z_\infty = 0] \ .$$

The above implies that the Z_m cannot converge in probability to Z_∞ [6, page 70, Equation (5)]. This contradicts [6, Theorem 4.1.2], by which almost sure convergence implies convergence in probability.

We end this section with the following observation: in both lemmas, we assume that the T_i are uniformly distributed over $1/\ell$. This is in line with how polar codes are defined, and has afforded us some notational convenience. However, both lemmas still hold if this assumption is not met. That is, in the more general case, we define E as the expected value of $\log_\ell D$, where D equals d_t if $T_0 = t$. To show this, the current proofs need only slight and superficial amendments.

ACKNOWLEDGMENTS

The author thanks Eren Şaşoğlu, Boaz Shuval, and the anonymous reviewers for helpful comments.

REFERENCES

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

[2] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6253–6264, December 2010.

[3] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2009)*, Seoul, South Korea, 2009, pp. 1493–1495.

[4] E. Şaşoğlu, "Polarization and polar codes," in *Found. and Trends in Commun. and Inform. Theory*, vol. 8, no. 4, 2012, pp. 259–381.

[5] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010.

[6] K. L. Chung, *A Course in Probability Theory*, 3rd ed. San Diego: Academic Press, 2001.

Ido Tal (S'05–M'08) was born in Haifa, Israel, in 1975. He received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Technion-Israel Institute of Technology, Haifa, Israel, in 1998, 2003 and 2009, respectively. During 2010–2012 he was a postdoctoral scholar at the University of California at San Diego. In 2012 he joined the Electrical Engineering Department at Technion. His research interests include constrained coding and error-control coding. He received the IEEE Joint Communications Society/Information Theory Society Paper Award (jointly with Alexander Vardy) for the year 2017.