Fast Polarization for Processes with Memory

Boaz Shuval, Ido Tal
Department of Electrical Engineering,
Technion, Haifa 32000, Israel.
Email: {bshuval@campus, idotal@ee}.technion.ac.il

***Abstract*—Fast polarization is crucial for the performance guarantees of polar codes. In the memoryless setting, the rate of polarization is known to be exponential in the square root of the block length. A complete characterization of the rate of polarization for models with memory has been missing. We consider polar codes for processes with memory that are characterized by an underlying aperiodic and irreducible finite state Markov chain. We show that the rate of polarization for these processes is the same as in the memoryless setting, both to the high and to the low-entropy sets. Thus, polar codes achieve the Markov capacity in many information-theoretic applications.**

I. Introduction

Memory is prevalent in many communication scenarios. In this research we show that polar codes can be used directly for a large class of scenarios with memory. This enables leveraging the attractive properties of polar codes — such as low complexity encoding and decoding, vanishing error performance, and versatility — to scenarios with memory.

Polar codes [1] were first developed for binary-input, symmetric, memoryless, channels. They employ *successive cancellation* (SC) decoding, which consists of N successive decoding operations. The polarization phenomenon is that for large enough N, the decoding operations polarize to two sets: a 'low-entropy' set and a 'high-entropy' set. The vanishing error performance of polar codes is due to polarization happening sufficiently fast. Fast polarization to the low-entropy set for the memoryless setting was established in [1], [2].

Polar codes were extended to many other memoryless scenarios, e.g., non-binary channels [3], [4], source coding [5], [6], asymmetric channels and sources [7], and more. Many of these applications are contingent upon fast polarization to the high-entropy set; for memoryless settings, this was established in [5] (see also [8] for a different proof that is closer to ours).

The study of polar codes for scenarios with memory began in [4, Chapter 5], which showed that polarization occurs for a certain class of processes with memory. In [9], polarization was established for a more general class of processes with memory. That paper further showed that polarization to the low-entropy set is fast even for processes with memory. Fast polarization to the high-entropy set was not addressed.

A practical decoding algorithm for polar codes for finite-state channels was suggested in [10]. This algorithm is an extension of SC decoding, taking into account an underlying state structure. Its increase in complexity relative to the complexity of SC decoding is polynomial with the number of states. The authors also showed [10, Theorem 3] that their elegant scheme from [7] can be applied to models with memory. To this end, they required the additional (then unproved) assumption of fast polarization both to the low and high-entropy sets.

This paper completes the picture. We show that for a large class of processes with memory, polarization is fast both to the low-entropy and high-entropy sets. Fast polarization to the low-entropy set will follow from a specialization of [9]. Fast polarization to the high-entropy set, Theorem 10, is the main result of this paper. Consequently, polar codes can be used in settings with memory with vanishing error probability.

Specifically, we consider stationary processes whose memory can be encompassed by an underlying (hidden) aperiodic and irreducible finite-state Markov chain. This family of processes includes, as special cases, finite state Markov channels with an ergodic state sequence, discrete ergodic sources with finite memory, and many input-constrained systems (e.g., (d,k)-runlength limited (RLL) constraint, with and without noise).

Due to space constraints, some proofs are omitted/shortened. The full version of this paper [11] contains detailed proofs.

II. Preliminaries

A. Distribution Parameters

Let random variables (U,Q) have joint distribution $P_{\mathsf{U},\mathsf{Q}}(u,q) = P_{\mathsf{Q}}(q)P_{\mathsf{U}|\mathsf{Q}}(u|q)$. For simplicity, we assume that U is binary. The random variable Q is some observation dependent on U that takes values in a finite alphabet $\mathcal{Q}$.

In the following equations, we define the *Bhattacharyya parameter* of U given Q, $\mathcal{Z}(\mathsf{U}|\mathsf{Q})$; the *total variation distance* of U given Q, $\mathcal{K}(\mathsf{U}|\mathsf{Q})$; and the *conditional entropy* of U given Q, $\mathcal{H}(\mathsf{U}|\mathsf{Q})$.

$$\mathcal{Z}(\mathsf{U}|\mathsf{Q}) = \sum\nolimits_q 2\sqrt{P_{\mathsf{U},\mathsf{Q}}(0,q)P_{\mathsf{U},\mathsf{Q}}(1,q)}, \tag{1}$$

$$\mathcal{K}(\mathsf{U}|\mathsf{Q}) = \sum\nolimits_q \left|P_{\mathsf{U},\mathsf{Q}}(0,q) - P_{\mathsf{U},\mathsf{Q}}(1,q)\right|, \tag{2}$$

$$\mathcal{H}(\mathsf{U}|\mathsf{Q}) = \sum\nolimits_q P_{\mathsf{Q}}(q) \sum\nolimits_u P_{\mathsf{U}|\mathsf{Q}}(u|q)\log_2 P_{\mathsf{U}|\mathsf{Q}}(u|q). \tag{3}$$

The three parameters take values in $[0,1]$.

Lemma 1. *We have*

$$\mathcal{K}(\mathsf{U}|\mathsf{Q}) \geq \sqrt{1-\mathcal{H}(\mathsf{U}|\mathsf{Q})} \geq \sqrt{1-\mathcal{Z}(\mathsf{U}|\mathsf{Q})},$$
$$\mathcal{K}(\mathsf{U}|\mathsf{Q}) \leq \sqrt{1-\mathcal{Z}(\mathsf{U}|\mathsf{Q})^2} \leq \sqrt{1-\mathcal{H}(\mathsf{U}|\mathsf{Q})^2}.$$

Informally, as a consequence of the above lemma,

$$\begin{aligned}\mathcal{Z}(\mathsf{U}|\mathsf{Q}) \approx 0 &\iff \mathcal{H}(\mathsf{U}|\mathsf{Q}) \approx 0 \iff \mathcal{K}(\mathsf{U}|\mathsf{Q}) \approx 1,\\ \mathcal{Z}(\mathsf{U}|\mathsf{Q}) \approx 1 &\iff \mathcal{H}(\mathsf{U}|\mathsf{Q}) \approx 1 \iff \mathcal{K}(\mathsf{U}|\mathsf{Q}) \approx 0.\end{aligned} \tag{4}$$

The definitions in Equations (1) to (3) naturally extend to the case where there are multiple random variables related to U. For example, consider a triplet of random variables $(\mathsf{U},\mathsf{Q},\mathsf{S})$ with joint distribution $P_{\mathsf{U},\mathsf{Q},\mathsf{S}}(u,q,s)$ such that U is binary and Q,S take values in finite alphabets $\mathcal{Q},\mathcal{S}$. Then, $\mathcal{K}(\mathsf{U}|\mathsf{Q},\mathsf{S}) = \sum_{q,s} |P_{\mathsf{U},\mathsf{Q},\mathsf{S}}(0,q,s) - P_{\mathsf{U},\mathsf{Q},\mathsf{S}}(1,q,s)|$; the remaining parameters are similarly extended.

Lemma 2. *The triplet* $(\mathsf{U},\mathsf{Q},\mathsf{S})$ *satisfies* $\mathcal{K}(\mathsf{U}|\mathsf{Q}) \leq \mathcal{K}(\mathsf{U}|\mathsf{Q},\mathsf{S})$*;* $\mathcal{Z}(\mathsf{U}|\mathsf{Q}) \geq \mathcal{Z}(\mathsf{U}|\mathsf{Q},\mathsf{S})$*; and* $\mathcal{H}(\mathsf{U}|\mathsf{Q}) \geq \mathcal{H}(\mathsf{U}|\mathsf{Q},\mathsf{S})$.

B. Polar Construction

Let $(\mathsf{X}_j,\mathsf{Y}_j)$, $j=1,2,\ldots$ be a stationary process, such that X_j are binary and $\mathsf{Y}_j \in \mathcal{Y}$, where $\mathcal{Y}$ is a finite alphabet. Random variables X_j are to be estimated from observations Y_j.

We denote Arıkan's polarization matrix by G_N, where $N = 2^n$. Following [9], we define for $i = 1,2,\ldots,N$:

$$\mathsf{U}_1^N = \mathsf{X}_1^N G_N, \tag{5a}$$
$$\mathsf{V}_1^N = \mathsf{X}_{N+1}^{2N} G_N, \tag{5b}$$
$$\mathsf{Q}_i = (\mathsf{U}_1^{i-1}, \mathsf{Y}_1^N), \tag{5c}$$
$$\mathsf{R}_i = (\mathsf{V}_1^{i-1}, \mathsf{Y}_{N+1}^{2N}). \tag{5d}$$

Note that these definitions apply to two consecutive blocks of length N; this will be useful in the sequel. By (5), we can write

$$(\mathsf{U}_i,\mathsf{Q}_i) = f(\mathsf{X}_1^N,\mathsf{Y}_1^N), \quad (\mathsf{V}_i,\mathsf{R}_i) = f(\mathsf{X}_{N+1}^{2N},\mathsf{Y}_{N+1}^{2N}),$$

where function f depends solely on i. Due to stationarity, $P_{\mathsf{U}_i,\mathsf{Q}_i} = P_{\mathsf{V}_i,\mathsf{R}_i}$. Denoting $\mathsf{T}_i = \mathsf{U}_i + \mathsf{V}_i$, we obtain

$$P_{\mathsf{T}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i}(t,v,q,r) = P_{\mathsf{U}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i}(t+v,v,q,r). \tag{6}$$

Let $\mathsf{B}_1,\mathsf{B}_2,\ldots$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli-$1/2$ random variables. Set $i = 1 + (\mathsf{B}_1\mathsf{B}_2\cdots\mathsf{B}_n)_2$, where $(\mathsf{B}_1\mathsf{B}_2\cdots\mathsf{B}_n)_2 = \sum_{j=1}^n \mathsf{B}_j 2^{n-j}$. Thus, i is a random variable that assumes any value in $\{1,2,\ldots,N\}$ with equal probability. Define the random variables

$$\mathsf{K}_n = \mathcal{K}(\mathsf{U}_i|\mathsf{U}_1^{i-1},\mathsf{Y}_1^N) = \mathcal{K}(\mathsf{U}_i|\mathsf{Q}_i),$$
$$\mathsf{Z}_n = \mathcal{Z}(\mathsf{U}_i|\mathsf{U}_1^{i-1},\mathsf{Y}_1^N) = \mathcal{Z}(\mathsf{U}_i|\mathsf{Q}_i),$$
$$\mathsf{H}_n = \mathcal{H}(\mathsf{U}_i|\mathsf{U}_1^{i-1},\mathsf{Y}_1^N) = \mathcal{H}(\mathsf{U}_i|\mathsf{Q}_i)$$

whenever $i = 1 + (\mathsf{B}_1\mathsf{B}_2\cdots\mathsf{B}_n)_2$. They denote the relevant distribution parameters for a uniformly chosen index after n polarization steps.

By the properties of G_N [1, Section VII],

$$\mathsf{K}_{n+1} = \begin{cases} \mathcal{K}(\mathsf{U}_i+\mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i), & \text{if } \mathsf{B}_{n+1}=0 \\ \mathcal{K}(\mathsf{V}_i|\mathsf{U}_i+\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i), & \text{if } \mathsf{B}_{n+1}=1. \end{cases} \tag{7}$$

Similar relationships hold for H_{n+1} and Z_{n+1}.

C. Polarization

Polarization occurs when the fraction of indices with moderate conditional entropy $|\{i : \mathcal{H}(\mathsf{U}_i|\mathsf{Q}_i) \in (\epsilon, 1-\epsilon)\}|/N$ vanishes for large enough n, for any $\epsilon > 0$.

Definition 1. Let A_n, $n = 1,2,\ldots$ be a sequence of random variables that take values in $[0,1]$.

1) The sequence A_n *polarizes* if it converges almost surely to a $\{0,1\}$-random variable A_∞ as $n \to \infty$.
2) The sequence A_n *polarizes fast to* 0 with $\beta > 0$ if it polarizes and $\lim_{n\to\infty} \mathbb{P}(\mathsf{A}_n < 2^{-2^{n\beta}}) = \mathbb{P}(\mathsf{A}_\infty = 0)$.
3) The sequence A_n *polarizes fast to* 1 with $\beta > 0$ if it polarizes and $\lim_{n\to\infty} \mathbb{P}(\mathsf{A}_n > 1 - 2^{-2^{n\beta}}) = \mathbb{P}(\mathsf{A}_\infty = 1)$.

The following lemma [2], [4], is an important tool for establishing fast polarization (see also [12]).

Lemma 3. *[2], [4] Let* B_n, $n = 1,2,\ldots$ *be an i.i.d. Bernoulli-*$1/2$ *process and* A_n, $n = 1,2,\ldots$ *be a* $[0,1]$*-valued process that polarizes to* A_∞*. Assume that there exist* $k \geq 1$ *and* $d_0, d_1 > 0$ *such that* $\mathsf{A}_{n+1} \leq k\mathsf{A}_n^{d_i}$ *if* $\mathsf{B}_{n+1} = i$*, for* $i = 0,1$*. Then, for any* $0 < \beta < E = (\log_2 d_0 + \log_2 d_1)/2$*, we have*

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{A}_n < 2^{-2^{n\beta}}) = \mathbb{P}(\mathsf{A}_\infty = 0). \tag{8}$$

Arıkan showed in [1] that in the memoryless case (i.e., $P_{\mathsf{X}_1^N,\mathsf{Y}_1^N}(x_1^N,y_1^N) = \prod_{j=1}^N P_{\mathsf{X},\mathsf{Y}}(x_j,y_j)$) the process H_n polarizes. Fast polarization to the low-entropy set was established in [2] by showing that $\mathsf{Z}_{n+1} \leq 2\mathsf{Z}_n$ if $\mathsf{B}_{n+1} = 0$ and $\mathsf{Z}_{n+1} = \mathsf{Z}_n^2$ if $\mathsf{B}_{n+1} = 1$, and using Lemma 3 and Equation (4).

Fast polarization to the high-entropy set is important for many applications of polar codes. For example, it is integral to source coding applications [5] and to channel coding without symmetry assumptions [7]. Lemma 3 will be useful for establishing fast polarization results to the high entropy set.

III. Finite-State Aperiodic Irreducible Markov Processes

We now introduce a class of processes with memory, described using a hidden state sequence. We call them *Finite-state Apreiodic Irreducible Markov processes* (FAIM processes).

Our model applies to many problems in information theory that can be described using states. Examples include compression of finite-memory sources and coding for input constrained channels. Additionally, our model may be applied to finite-state channels and channels with intersymbol interference; in this case, the FAIM state sequence describes both the channel state and input state. That is, FAIM processes enable us to model non-i.i.d. input sequences.

A. Definition

Let $(\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j)$, $j \in \mathbb{Z}$ be a stationary process, where X_j is binary, $\mathsf{Y}_j \in \mathcal{Y}$, and $\mathsf{S}_j \in \mathcal{S}$. Alphabets $\mathcal{Y}$ and $\mathcal{S}$ are finite, and $\mathcal{S} = \{1,2,\ldots,|\mathcal{S}|\}$. We call $\mathsf{S}_j, j \in \mathbb{Z}$ the *state sequence*. It encompasses the memory of the process.

The process $(\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j)$, $j \in \mathbb{Z}$ is described by the conditional distribution $P_{\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j|\mathsf{S}_{j-1}}$, which, by stationarity, is independent of j. We call it a FAIM process if

$$P_{\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j|\mathsf{S}_{j-1}} = P_{\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j|\mathsf{S}_{-\infty}^{j-1},\mathsf{X}_{-\infty}^{j-1},\mathsf{Y}_{-\infty}^{j-1}},$$

and the state sequence S_j, $j \in \mathbb{Z}$ is a finite-state, homogeneous, aperiodic, and irreducible Markov chain. For any $N > M > 0$, with b denoting the value of the middle state S_M,

$$\begin{aligned} P_{\mathsf{X}_1^N,\mathsf{Y}_1^N,\mathsf{S}_N|\mathsf{S}_0} &= \sum\nolimits_b P_{\mathsf{X}_1^M,\mathsf{Y}_1^M,\mathsf{S}_M,\mathsf{X}_{M+1}^N,\mathsf{Y}_{M+1}^N,\mathsf{S}_N|\mathsf{S}_0} \\ &= \sum\nolimits_b P_{\mathsf{X}_{M+1}^N,\mathsf{Y}_{M+1}^N,\mathsf{S}_N|\mathsf{S}_M} \cdot P_{\mathsf{X}_1^M,\mathsf{Y}_1^M,\mathsf{S}_M|\mathsf{S}_0}. \end{aligned} \tag{9}$$

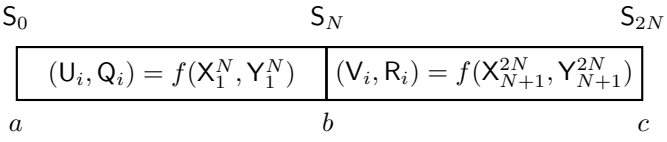

Fig. 1. Two adjacent length-N blocks of a FAIM process. Here, $\mathsf{S}_0 = a$, $\mathsf{S}_N = b$, and $\mathsf{S}_{2N} = c$, where $a, b, c \in \mathcal{S}$.

We use the following shorthand for the stationary distribution of the state sequence: $\pi_N(a) = P_{\mathsf{S}_N}(a)$, $\pi_{N|M}(b|a) = P_{\mathsf{S}_N|\mathsf{S}_M}(b|a)$, and $\pi_{N,M}(b,a) = P_{\mathsf{S}_N,\mathsf{S}_M}(b,a)$. By stationarity, $\pi_N(a) = \pi_0(a)$. By irreducibility and aperiodicity, $\pi_0(a) > 0$ for all $a \in \mathcal{S}$. Aperiodicity is assumed because periodic processes may not polarize [9, Theorem 3].

B. Blocks of a FAIM Process

Typically, the state sequence is not observed by the encoder or decoder. The joint distribution of $(\mathsf{X}_1^N, \mathsf{Y}_1^N)$ is given by

$$P_{\mathsf{X}_1^N,\mathsf{Y}_1^N}(x_1^N, y_1^N) = \sum\nolimits_{b,a} P_{\mathsf{X}_1^N,\mathsf{Y}_1^N,\mathsf{S}_N|\mathsf{S}_0}(x_1^N, y_1^N, b|a)\pi_0(a).$$

Definition 2 (Block)**.** Let $(\mathsf{X}_j, \mathsf{Y}_j, \mathsf{S}_j)$, $j \in \mathbb{Z}$ be a FAIM process. We call $(\mathsf{X}_{L+1}^M, \mathsf{Y}_{L+1}^M)$ a *block* of the FAIM process, with length $M - L$. State S_L is the *initial* state of the block and state S_M is the *final* state of the block.

We emphasize that for block $(\mathsf{X}_{L+1}^M, \mathsf{Y}_{L+1}^M)$, the initial state is S_L and *not* S_{L+1}.

The following lemma establishes that FAIM processes are a special case of the family of processes considered in [9].

Lemma 4. *If $(\mathsf{X}_j, \mathsf{Y}_j, \mathsf{S}_j)$, $j \in \mathbb{Z}$ is a FAIM process, there exists a non-increasing sequence $\psi(N)$ with $\psi(0) < \infty$ and $\psi(N) \to 1$ as $N \to \infty$, such that for any $N > M \geq L \geq 1$,*

$$P_{\mathsf{X}_1^L,\mathsf{Y}_1^L,\mathsf{X}_{M+1}^N,\mathsf{Y}_{M+1}^N} \leq \psi(M-L) \cdot P_{\mathsf{X}_1^L,\mathsf{Y}_1^L} \cdot P_{\mathsf{X}_{M+1}^N,\mathsf{Y}_{M+1}^N}. \quad (10)$$

To see this, define

$$\psi(N) = \begin{cases} \max_{a,b} \pi_{N|0}(b|a)/\pi_0(b), & \text{if } N > 0 \\ \max_a 1/\pi_0(a), & \text{if } N = 0, \end{cases} \quad (11)$$

and use (9). A process satisfying (10) with $\psi(N) \to 1$ as $N \to \infty$ is called *ψ-mixing*.

Two adjacent blocks share a state: the final state of the first block is the initial state of the second block. By (9), for any $N > M \geq 1$,

$$P_{\mathsf{X}_1^M,\mathsf{Y}_1^M,\mathsf{X}_{M+1}^N,\mathsf{Y}_{M+1}^N|\mathsf{S}_0,\mathsf{S}_M,\mathsf{S}_N} = P_{\mathsf{X}_1^M,\mathsf{Y}_1^M|\mathsf{S}_0,\mathsf{S}_M} P_{\mathsf{X}_{M+1}^N,\mathsf{Y}_{M+1}^N|\mathsf{S}_M,\mathsf{S}_N}. \quad (12)$$

We will use ascending letters to denote values of ordered states. In Figure 1 we illustrate a useful case. A block of length $2N$ comprises two adjacent blocks of length N. State S_0, the initial state of the first block, takes value a. State S_N, at the end of the first block and the beginning of the second block, takes value b. State S_{2N}, at the end of the second block, takes value c.

C. Boundary State-Informed Parameters for FAIM Processes

Let $(\mathsf{X}_1^N, \mathsf{Y}_1^N)$ be a block of a FAIM process and let $(\mathsf{U}, \mathsf{Q}) = f(\mathsf{X}_1^N, \mathsf{Y}_1^N)$, where function $f(\cdot,\cdot)$ is independent of the state sequence and U is binary. We denote

$$P_a^b(u,q) \triangleq P_{\mathsf{U},\mathsf{Q}|\mathsf{S}_N,\mathsf{S}_0}(u,q|b,a) = \frac{P_{\mathsf{U},\mathsf{Q},\mathsf{S}_N|\mathsf{S}_0}(u,q,b|a)}{\pi_{N|0}(b|a)}. \quad (13)$$

and further define $P_a^b(q) = P_{\mathsf{Q}|\mathsf{S}_N,\mathsf{S}_0}(q|b,a)$. We denote by $\mathcal{Z}_a^b(\mathsf{U}|\mathsf{Q})$, $\mathcal{K}_a^b(\mathsf{U}|\mathsf{Q})$, and $\mathcal{H}_a^b(\mathsf{U}|\mathsf{Q})$ the results of replacing $P_{\mathsf{U},\mathsf{Q}}(u,q)$ with $P_a^b(u,q)$ in equations (1) to (3), respectively. For example, $\mathcal{K}_a^b(\mathsf{U}|\mathsf{Q}) = \sum_q \left|P_a^b(0,q) - P_a^b(1,q)\right|$.

As $P_{\mathsf{U},\mathsf{Q},\mathsf{S}_N,\mathsf{S}_0}(u,q,b,a) = P_a^b(u,q) \cdot \pi_{N,0}(b,a)$, we respectively define the *boundary state-informed* (BSI) total variation distance, Bhattacharyya parameter, and conditional entropy, as

$$\begin{aligned} \mathcal{K}(\mathsf{U}|\mathsf{Q},\mathsf{S}_N,\mathsf{S}_0) &= \sum\nolimits_{a,b} \pi_{N,0}(b,a)\mathcal{K}_a^b(\mathsf{U}|\mathsf{Q}), \\ \mathcal{Z}(\mathsf{U}|\mathsf{Q},\mathsf{S}_N,\mathsf{S}_0) &= \sum\nolimits_{a,b} \pi_{N,0}(b,a)\mathcal{Z}_a^b(\mathsf{U}|\mathsf{Q}), \\ \mathcal{H}(\mathsf{U}|\mathsf{Q},\mathsf{S}_N,\mathsf{S}_0) &= \sum\nolimits_{a,b} \pi_{N,0}(b,a)\mathcal{H}_a^b(\mathsf{U}|\mathsf{Q}). \end{aligned}$$

BSI parameters are defined for blocks of the process; they depend on the initial and final states of the block.

IV. Fast Polarization for FAIM Processes

We use the notation of Section II-B for FAIM processes. That is, $\mathsf{U}_1^N, \mathsf{V}_1^N, \mathsf{Q}_i, \mathsf{R}_i$, $i = 1, \ldots, N$ are defined using (5). The random variables $\mathsf{B}_1, \ldots, \mathsf{B}_n$ are Bernoulli-1/2 and i.i.d., and $i = 1 + (\mathsf{B}_1\mathsf{B}_2 \cdots \mathsf{B}_n)_2$. Via i, we define the random variables $\mathsf{K}_n = \mathcal{K}(\mathsf{U}_i|\mathsf{Q}_i)$, $\mathsf{H}_n = \mathcal{H}(\mathsf{U}_i|\mathsf{Q}_i)$, and $\mathsf{Z}_n = \mathcal{Z}(\mathsf{U}_i|\mathsf{Q}_i)$.

Let $\hat{\mathsf{K}}_n, \hat{\mathsf{H}}_n$, and $\hat{\mathsf{Z}}_n$ denote the BSI versions of $\mathsf{K}_n, \mathsf{Z}_n$, and H_n, respectively. That is, with i chosen randomly as above,

$$\begin{aligned} \hat{\mathsf{K}}_n &= \mathcal{K}(\mathsf{U}_i|\mathsf{Q}_i,\mathsf{S}_N,\mathsf{S}_0), \\ \hat{\mathsf{Z}}_n &= \mathcal{Z}(\mathsf{U}_i|\mathsf{Q}_i,\mathsf{S}_N,\mathsf{S}_0), \\ \hat{\mathsf{H}}_n &= \mathcal{H}(\mathsf{U}_i|\mathsf{Q}_i,\mathsf{S}_N,\mathsf{S}_0). \end{aligned} \quad (14)$$

By Lemma 2, $\mathsf{K}_n \leq \hat{\mathsf{K}}_n$, $\mathsf{Z}_n \geq \hat{\mathsf{Z}}_n$, and $\mathsf{H}_n \geq \hat{\mathsf{H}}_n$ for any n. Similar to (7), recalling that $2N = 2^{n+1}$, we have

$$\hat{\mathsf{K}}_{n+1} = \begin{cases} \mathcal{K}(\mathsf{U}_i + \mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_{2N}), & \text{if } \mathsf{B}_{n+1} = 0 \\ \mathcal{K}(\mathsf{V}_i|\mathsf{U}_i + \mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_{2N}), & \text{if } \mathsf{B}_{n+1} = 1. \end{cases} \quad (15)$$

Relationships akin to (15) hold for $\hat{\mathsf{Z}}_{n+1}$ and $\hat{\mathsf{H}}_{n+1}$, with $\mathcal{K}$ replaced with $\mathcal{Z}$ and $\mathcal{H}$, respectively.

A. Polarization of FAIM Processes

Let

$$\mathcal{H}_\star(\mathsf{X}|\mathsf{Y}) \triangleq \lim_{N\to\infty} \frac{1}{N}\mathcal{H}(\mathsf{X}_1^N|\mathsf{Y}_1^N).$$

This limit exists due to stationarity [13, Section 4.2] and the identity $\mathcal{H}(\mathsf{X}_1^N|\mathsf{Y}_1^N) = \mathcal{H}(\mathsf{X}_1^N, \mathsf{Y}_1^N) - \mathcal{H}(\mathsf{Y}_1^N)$. In [9], the following was shown:

Theorem 5. *[9] For a stationary ψ-mixing process $(\mathsf{X}_j, \mathsf{Y}_j)$, $j \in \mathbb{Z}$, with $\psi(0) < \infty$:*

1) *H_n polarizes to H_∞ with $\mathbb{P}(\mathsf{H}_\infty = 1) = \mathcal{H}_\star(\mathsf{X}|\mathsf{Y})$;*
2) *Z_n polarizes fast to 0 with any $\beta < 1/2$.*

Since FAIM processes are ψ-mixing, we obtain:

Corollary 6. *Let $(\mathsf{X}_j, \mathsf{Y}_j, \mathsf{S}_j)$, $j \in \mathbb{Z}$ be a FAIM process. Then,*

1) *Its conditional entropy process H_n polarizes to H_∞ with $\mathbb{P}(\mathsf{H}_\infty = 1) = \mathcal{H}_\star(\mathsf{X}|\mathsf{Y})$.*
2) *Its Bhattacharyya process Z_n polarizes fast to 0 with any $\beta < 1/2$.*

Proof: By Lemma 4, FAIM processes are ψ-mixing and satisfy the requirements of Theorem 5. ∎

B. Polarization of the BSI Distribution Parameters

This section is concerned with proving that the BSI distribution parameters polarize.

Theorem 7. *Let $(\mathsf{X}_j, \mathsf{Y}_j, \mathsf{S}_j)$, $j \in \mathbb{Z}$ be a FAIM process. The BSI conditional entropy process $\hat{\mathsf{H}}_n$ polarizes to $\hat{\mathsf{H}}_\infty$ and $\hat{\mathsf{H}}_\infty = \mathsf{H}_\infty$ almost surely.*

Proof: Consider two adjacent blocks of length $N = 2^n$ and let $i = 1 + (\mathsf{B}_1\mathsf{B}_2\cdots\mathsf{B}_n)_2$, as in Figure 1. By (12),

$$P_{\mathsf{U}_i,\mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_N,\mathsf{S}_{2N}} = P_{\mathsf{U}_i|\mathsf{Q}_i,\mathsf{S}_0,\mathsf{S}_N} P_{\mathsf{V}_i|\mathsf{R}_i,\mathsf{S}_N,\mathsf{S}_{2N}}. \tag{16}$$

Thus,

$$\begin{aligned}
\hat{\mathsf{H}}_n &\overset{(a)}{=} \frac{1}{2}\Big(\mathcal{H}(\mathsf{U}_i|\mathsf{Q}_i,\mathsf{S}_0,\mathsf{S}_N) + \mathcal{H}(\mathsf{V}_i|\mathsf{R}_i,\mathsf{S}_N,\mathsf{S}_{2N})\Big)\\
&\overset{(b)}{=} \frac{1}{2}\mathcal{H}(\mathsf{U}_i,\mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_N,\mathsf{S}_{2N})\\
&\overset{(c)}{=} \frac{1}{2}\mathcal{H}(\mathsf{U}_i+\mathsf{V}_i,\mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_N,\mathsf{S}_{2N})\\
&\overset{(d)}{=} \frac{1}{2}\Big(\mathcal{H}(\mathsf{U}_i+\mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_N,\mathsf{S}_{2N})\\
&\quad + \mathcal{H}(\mathsf{V}_i|\mathsf{U}_i+\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_N,\mathsf{S}_{2N})\Big)\\
&\overset{(e)}{\le} \frac{1}{2}\Big(\mathcal{H}(\mathsf{U}_i+\mathsf{V}_i|\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_{2N})\\
&\quad + \mathcal{H}(\mathsf{V}_i|\mathsf{U}_i+\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i,\mathsf{S}_0,\mathsf{S}_{2N})\Big)
\end{aligned}$$

where (a) is by stationarity, (b) is by (16), (c) is because the mapping $(\mathsf{U},\mathsf{V}) \mapsto (\mathsf{U}+\mathsf{V},\mathsf{V})$ is one-to-one and onto, (d) is by the chain rule for entropies, and (e) is by Lemma 2.

Thus, recalling (15) (applied to the BSI conditional entropy), $\hat{\mathsf{H}}_n$ is a submartingale sequence. It is also bounded, as $\hat{\mathsf{H}}_n \in [0,1]$ for any n. Hence, it converges almost surely to some random variable $\hat{\mathsf{H}}_\infty \in [0,1]$.

Consequently, $\Delta\mathsf{H}_n = \mathsf{H}_n - \hat{\mathsf{H}}_n$ converges almost surely to the random variable $\Delta\mathsf{H}_\infty = \mathsf{H}_\infty - \hat{\mathsf{H}}_\infty$. By Lemma 2, $\Delta\mathsf{H}_n \ge 0$ for any n, implying that $\Delta\mathsf{H}_\infty \ge 0$ almost surely.

The effect of knowing a block's initial and final states becomes negligible for sufficiently long blocks; thus, it can be shown that $\lim_{n\to\infty}\mathbb{E}[\Delta\mathsf{H}_n] = 0$. For details, see [11, Lemma 10]. Since $\Delta\mathsf{H}_n$ is a non-negative sequence that converges to $\Delta\mathsf{H}_\infty$ almost surely, by Fatou's lemma,

$$\begin{aligned}
0 \le \mathbb{E}[\Delta\mathsf{H}_\infty] &= \mathbb{E}\left[\liminf_{n\to\infty}\Delta\mathsf{H}_n\right]\\
&\le \liminf_{n\to\infty}\mathbb{E}[\Delta\mathsf{H}_n] = \lim_{n\to\infty}\mathbb{E}[\Delta\mathsf{H}_n] = 0.
\end{aligned}$$

In other words, $\mathbb{E}[\Delta\mathsf{H}_\infty] = 0$. By Markov's inequality, $\mathbb{P}(\Delta\mathsf{H}_\infty \ge \delta) \le \mathbb{E}[\Delta\mathsf{H}_\infty]/\delta = 0$ for any $\delta > 0$; consequently, $\mathbb{P}(\Delta\mathsf{H}_\infty = 0) = \mathbb{P}(\mathsf{H}_\infty = \hat{\mathsf{H}}_\infty) = 1$. That is, $\hat{\mathsf{H}}_\infty = \mathsf{H}_\infty$ almost surely. ∎

Corollary 8.

1) *The sequences Z_n and $\hat{\mathsf{Z}}_n$ polarize to random variables Z_∞ and $\hat{\mathsf{Z}}_\infty$, respectively. Moreover, $\mathsf{Z}_\infty = \hat{\mathsf{Z}}_\infty = \mathsf{H}_\infty$ almost surely.*
2) *The sequences K_n and $\hat{\mathsf{K}}_n$ polarize to random variables K_∞ and $\hat{\mathsf{K}}_\infty$, respectively. Moreover, $\mathsf{K}_\infty = \hat{\mathsf{K}}_\infty = 1 - \mathsf{H}_\infty$ almost surely.*

This follows from Lemma 1, Corollary 6, and Theorem 7.

C. Fast Polarization to the High Entropy Set

In this section, we establish fast polarization to the high entropy set. We do this by proving that the total variation process K_n polarizes fast to 0, which implies that the Bhattacharyya process Z_n polarizes fast to 1.

Proposition 9. *Let $(\mathsf{X}_j, \mathsf{Y}_j, \mathsf{S}_j)$, $j \in \mathbb{Z}$ be a FAIM process. Then, with $\psi(0)$ as in (11),*

$$\hat{\mathsf{K}}_{n+1} \le \begin{cases} \psi(0)\hat{\mathsf{K}}_n^2, & \text{if } \mathsf{B}_{n+1} = 0\\ 2\hat{\mathsf{K}}_n, & \text{if } \mathsf{B}_{n+1} = 1. \end{cases} \tag{17}$$

Proof of Proposition 9: Consider two adjacent blocks of length $N = 2^n$ and let $i = 1 + (\mathsf{B}_1\mathsf{B}_2\cdots\mathsf{B}_n)_2$, as in Figure 1. By stationarity,

$$\hat{\mathsf{K}}_n = \sum_{a,b\in\mathcal{S}} \pi_{N,0}(b,a)\mathcal{K}_a^b(\mathsf{U}_i|\mathsf{Q}_i) = \sum_{b,c\in\mathcal{S}} \pi_{2N,N}(c,b)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i). \tag{18}$$

As in (13), we denote $P_a^c(u,q) = P_{\mathsf{U}_i,\mathsf{Q}_i|\mathsf{S}_N,\mathsf{S}_0}(u,q|c,a) = P_{\mathsf{V}_i,\mathsf{R}_i|\mathsf{S}_{2N},\mathsf{S}_N}(u,q|c,a)$, and $P_a^c(s) = P_a^c(0,s) + P_a^c(1,s)$; in particular, $\sum_s P_a^c(s) = 1$. Further denote

$$\mu(b) = \pi_{2N|N}(c|b)\pi_{N|0}(b|a)\pi_0(a) = \frac{\pi_{2N,N}(c,b)\,\pi_{N,0}(b,a)}{\pi_N(b)}.$$

For brevity, we omit the dependence of $\mu(b)$ on a, c. By (11),

$$\mu(b) \le \psi(0)\cdot\pi_{2N,N}(c,b)\cdot\pi_{N,0}(b,a). \tag{19}$$

Also, since $\pi_N(b) = \sum_{a\in\mathcal{S}}\pi_{N,0}(b,a) = \sum_{c\in\mathcal{S}}\pi_{2N,N}(c,b)$,

$$\sum\nolimits_a \mu(b) = \pi_{2N,N}(c,b), \quad \sum\nolimits_c \mu(b) = \pi_{N,0}(b,a). \tag{20}$$

By (9) and (13),

$$\begin{aligned}
&\pi_{2N,0}(c,a)P_{\mathsf{U}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i|\mathsf{S}_{2N},\mathsf{S}_0}(u,v,q,r|c,a)\\
&\quad = \pi_0(a)\sum\nolimits_b P_{\mathsf{U}_i,\mathsf{Q}_i,\mathsf{S}_N|\mathsf{S}_0}(u,q,b|a)P_{\mathsf{V}_i,\mathsf{R}_i,\mathsf{S}_{2N}|\mathsf{S}_N}(v,r,c|b)\\
&\quad = \sum\nolimits_b \mu(b)P_a^b(u,q)P_b^c(v,r).
\end{aligned} \tag{21}$$

Set $\mathsf{T}_i = \mathsf{U}_i + \mathsf{V}_i$. Using (7), a single-step polarization from $\hat{\mathsf{K}}_n$ to $\hat{\mathsf{K}}_{n+1}$ becomes

$$\hat{\mathsf{K}}_{n+1} = \begin{cases} \sum_{a,c}\pi_{2N,0}(c,a)\mathcal{K}_a^c(\mathsf{T}_i|\mathsf{Q}_i,\mathsf{R}_i), & \text{if } \mathsf{B}_{n+1} = 0\\ \sum_{a,c}\pi_{2N,0}(c,a)\mathcal{K}_a^c(\mathsf{V}_i|\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i), & \text{if } \mathsf{B}_{n+1} = 1. \end{cases}$$

Here, $\mathcal{K}_a^c(\mathsf{T}_i|\mathsf{Q}_i,\mathsf{R}_i)$ and $\mathcal{K}_a^c(\mathsf{V}_i|\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i)$ are computed for a length-$2N$ block with initial state $\mathsf{S}_0 = a$ and final state $\mathsf{S}_{2N} = c$. The state at the middle of the block is $\mathsf{S}_N = b$. Denote by (6),

$$\begin{aligned}\bar{P}_a^c(t,v,q,r) &= P_{\mathsf{T}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i|\mathsf{S}_{2N},\mathsf{S}_0}(t,v,q,r|c,a)\\ &= P_{\mathsf{U}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i|\mathsf{S}_{2N},\mathsf{S}_0}(t+v,v,q,r|c,a)\end{aligned}$$

and $\bar{P}_a^c(t,q,r) = \sum_{v=0}^1 \bar{P}_a^c(t,v,q,r)$.

Consider first the case $\mathsf{B}_{n+1} = 0$:

$$\begin{aligned}&\pi_{2N,0}(c,a)\mathcal{K}_a^c(\mathsf{T}_i|\mathsf{Q}_i,\mathsf{R}_i)\\ &= \pi_{2N,0}(c,a)\sum\nolimits_{q,r}\left|\bar{P}_a^c(0,q,r) - \bar{P}_a^c(1,q,r)\right|\\ &\overset{(a)}{=} \sum_{q,r}\left|\sum_b \mu(b)\sum_{v=0}^1 P_b^c(v,r)(P_a^b(v,q) - P_a^b(v+1,q))\right|\\ &\overset{(b)}{\le} \sum_{q,r,b}\mu(b)\left|\sum_{v=0}^1 P_b^c(v,r)(P_a^b(v,q) - P_a^b(v+1,q))\right|\\ &= \sum\nolimits_{q,r,b}\mu(b)\left|P_a^b(0,q) - P_a^b(1,q)\right|\cdot\left|P_b^c(0,r) - P_b^c(1,r)\right|\\ &= \sum\nolimits_b \mu(b)\mathcal{K}_a^b(\mathsf{U}_i|\mathsf{Q}_i)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i)\\ &\overset{(c)}{\le} \psi(0)\sum\nolimits_b\Big(\pi_{2N,N}(c,b)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i)\Big)\Big(\pi_{N,0}(b,a)\mathcal{K}_a^b(\mathsf{U}_i|\mathsf{Q}_i)\Big)\\ &\overset{(d)}{\le} \psi(0)\sum_b \pi_{2N,N}(c,b)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i)\sum_{b'}\pi_{N,0}(b',a)\mathcal{K}_a^{b'}(\mathsf{U}_i|\mathsf{Q}_i),\end{aligned}$$

where (a) is by (21), (b) is by the triangle inequality, (c) is by (19), and (d) is by the inequality $\sum_j a_j b_j \le \sum_j a_j \sum_{j'} b_{j'}$, which holds for $a_j, b_j \ge 0$. By (18), the sum over $a, c \in \mathcal{S}$ yields

$$\sum\nolimits_{a,c}\pi_{2N,0}(c,a)\mathcal{K}_a^c(\mathsf{T}_i|\mathsf{Q}_i,\mathsf{R}_i) \le \psi(0)\hat{\mathsf{K}}_n^2.$$

Next, let $\mathsf{B}_{n+1} = 1$. We have

$$\begin{aligned}&\pi_{2N,0}(c,a)\mathcal{K}_a^c(\mathsf{V}_i|\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i)\\ &= \pi_{2N,0}(c,a)\sum\nolimits_{t,q,r}\left|\bar{P}_a^c(t,0,q,r) - \bar{P}_a^c(t,1,q,r)\right|\\ &\overset{(a)}{=} \sum_{t,q,r}\Big|\sum_b \mu(b)(P_a^b(t,q)P_b^c(0,r) - P_a^b(t+1,q)P_b^c(1,r))\Big|\\ &\overset{(b)}{=} \frac{1}{2}\sum\nolimits_{t,q,r}\Big|\sum\nolimits_b \mu(b)P_a^b(q)(P_b^c(0,r) - P_b^c(1,r))\\ &\qquad + \sum\nolimits_b \mu(b)P_b^c(r)(P_a^b(t,q) - P_a^b(t+1,q))\Big|\\ &\overset{(c)}{\le} \sum\nolimits_{q,b}\mu(b)P_a^b(q)\Big(\sum\nolimits_r\left|P_b^c(0,r) - P_b^c(1,r)\right|\Big)\\ &\quad + \sum\nolimits_{r,b}\mu(b)P_b^c(r)\Big(\sum\nolimits_q\left|P_a^b(0,q) - P_a^b(1,q)\right|\Big)\\ &= \sum\nolimits_b \mu(b)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i) + \sum\nolimits_b \mu(b)\mathcal{K}_a^b(\mathsf{U}_i|\mathsf{Q}_i),\end{aligned}$$

where (a) is by (21), (b) is by the equality

$$(\alpha\beta - \gamma\delta) = \big((\alpha+\gamma)(\beta-\delta) + (\beta+\delta)(\alpha-\gamma)\big)/2,$$

which holds for any four numbers $\alpha, \beta, \gamma, \delta$, and (c) is by the triangle inequality. By (20),

$$\sum\nolimits_{a,b,c}\mu(b)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i) = \sum\nolimits_{b,c}\pi_{2N,N}(c,b)\mathcal{K}_b^c(\mathsf{V}_i|\mathsf{R}_i) = \hat{\mathsf{K}}_n,$$

$$\sum\nolimits_{a,b,c}\mu(b)\mathcal{K}_a^b(\mathsf{U}_i|\mathsf{Q}_i) = \sum\nolimits_{a,b}\pi_{N,0}(b,a)\mathcal{K}_a^b(\mathsf{U}_i|\mathsf{Q}_i) = \hat{\mathsf{K}}_n.$$

Thus,

$$\sum\nolimits_{a,c}\pi_{2N,0}(c,a)\mathcal{K}_a^c(\mathsf{V}_i|\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i) \le 2\hat{\mathsf{K}}_n. \qquad \blacksquare$$

Theorem 10. *Let $(\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j)$, $j \in \mathbb{Z}$ be a FAIM process. Then K_n polarizes fast to 0 and Z_n polarizes fast to 1 for any $\beta < 1/2$.*

Proof: Fix $\beta < 1/2$. By Corollary 8 and (17), we can invoke Lemma 3 for $\hat{\mathsf{K}}_n$ with $E = 1/2$. Consequently, $\hat{\mathsf{K}}_n$ polarizes fast to 0, i.e.,

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}(\hat{\mathsf{K}}_n < 2^{-N^\beta}) &= \mathbb{P}(\hat{\mathsf{K}}_\infty = 0)\\ &= \mathbb{P}\left(\mathsf{H}_\infty = 1\right) = \mathcal{H}_\star(\mathsf{X}|\mathsf{Y}).\end{aligned}$$

For any n, by Lemma 1 and Lemma 2,

$$1 - \mathsf{Z}_n \le \sqrt{1-\mathsf{Z}_n} \le \mathsf{K}_n \le \hat{\mathsf{K}}_n.$$

Thus, K_n polarizes fast to 0. Moreover, $\mathbb{P}(\mathsf{Z}_n > 1 - 2^{-N^\beta}) \ge \mathbb{P}(\hat{\mathsf{K}}_n < 2^{-N^\beta})$. Taking limits, we obtain that $\lim_{n\to\infty}\mathbb{P}(\mathsf{Z}_n > 1 - 2^{-N^\beta}) \ge \mathcal{H}_\star(\mathsf{X}|\mathsf{Y})$.

On the other hand, by Corollary 6,

$$\lim_{n\to\infty}\mathbb{P}(\mathsf{Z}_n < 2^{-N^\beta}) = 1 - \mathcal{H}_\star(\mathsf{X}|\mathsf{Y}).$$

Since $\mathbb{P}(\mathsf{Z}_n < 2^{-N^\beta}) + \mathbb{P}(\mathsf{Z}_n > 1 - 2^{-N^\beta}) \le 1$ for any n, we must have

$$\lim_{n\to\infty}\mathbb{P}(\mathsf{Z}_n > 1 - 2^{-N^\beta}) = \mathcal{H}_\star(\mathsf{X}|\mathsf{Y}). \qquad \blacksquare$$

Corollary 11. *Let $(\mathsf{X}_j,\mathsf{Y}_j,\mathsf{S}_j)$, $j \in \mathbb{Z}$ be a FAIM process. Then H_n, $\hat{\mathsf{H}}_n$, Z_n, $\hat{\mathsf{Z}}_n$, K_n, and $\hat{\mathsf{K}}_n$ polarize fast both to 0 and to 1 with any $\beta < 1/2$.*

References

[1] E. Arıkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. 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 *Proc. IEEE Int. Sym. on Information Theory*, June 2009, pp. 1493–1495.

[3] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *2009 IEEE Information Theory Workshop*, October 2009, pp. 144–148.

[4] E. Şaşoğlu, "Polar Coding Theorems for Discrete Systems," Ph.D. dissertation, IC, Lausanne, 2011.

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

[6] E. Arıkan, "Source polarization," in *2010 IEEE Int. Sym. on Information Theory*, June 2010, pp. 899–903.

[7] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, December 2013.

[8] M. Alsan, "Re-proving Channel Polarization Theorems," Ph.D. dissertation, IC, Lausanne, 2015. [Online]. Available: https://infoscience.epfl.ch/record/203886/files/EPFL_TH6403.pdf

[9] E. Şaşoğlu and I. Tal, "Polar coding for processes with memory," in *2016 IEEE Int. Sym. on Information Theory (ISIT)*. IEEE, 2016, pp. 225–229.

[10] R. Wang, J. Honda, H. Yamamoto, R. Liu, and Y. Hou, "Construction of polar codes for channels with memory," in *2015 IEEE Information Theory Workshop*, October 2015, pp. 187–191.

[11] B. Shuval and I. Tal, "Fast polarization for processes with memory," 2017. [Online]. Available: arXiv:1710.02849

[12] I. Tal, "A simple proof of fast polarization," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7617–7619, Dec 2017.

[13] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

V. Addendum

There is a subtle error in the statement of Lemma 1. However, the consequence (4) and thus the rest of the paper are correct. For the correct statement of Lemma 1, see our journal paper.