Polar Coding for Information Regular Processes

Gretchen L. Matthews
Department of Mathematics
Virginia Tech
Blacksburg, United States
gmatthews@vt.edu

Stephen Timmel
Department of Mathematics
Virginia Tech
Blacksburg, United States
stimmel@vt.edu

***Abstract*—Polar codes use an explicit channel model, so the precise nature of the communication channel determines if and when polarization occurs. In practice, communication channels typically have memory which affects the probability of error in nearby symbols. In this paper, we extend existing results to show polarization for the more general family of information regular processes and define a growth rate for promptness. We then show that information regular processes with linear promptness polarize and provide an example with practical applications for which the rate of polarization is very slow.**

***Index Terms*—channel memory, mixing processes, polar codes**

I. Introduction

Polar codes, first described in [1] by Arikan, achieve channel capacity using a process known as polarization. An invertible kernel matrix acts on communication channels to produce synthetic channels with modified capacities. The generator matrix formed by repeatedly tensoring the kernel matrix with itself produces synthetic channels with capacities converging to $\{0, 1\}$. The asymptotic rate of convergence as the length N of the codeword grows is $O(2^{-N^{\beta}})$ for all $\beta < \frac{1}{2}$ [2].

This polarization process acts directly on communication channels, offering considerable flexibility in design. Not only can polar codes be adjusted to specific communication channels, but this modification is essential to their construction. It is unsurprising that polar codes have expanded from binary discrete memoryless channels [1] to arbitrary finite fields [3], [4], [5], [6], generalized memoryless channels [7], [8], [9], and specific instances of channel memory [10], [11], [12]. It was only in 2018 that a general framework for channel memory was developed [13], [14]. These papers demonstrate that there is a periodic data source which does not polarize, a specialized family of Markov processes polarize quickly, and the slightly broader family of ψ^*-mixing processes polarize.

Many communication channels of practical interest exhibit some form of channel memory, so it is important to know what types of memory are compatible with polar codes. The literature provides some insight but falls short of a classification. We focus on the large space of relatively unstudied processes which are aperiodic but not ψ^*-mixing with the goal of clarifying what sorts of channel memory permit polarization. Specifically, we extend Şaşoğlu's polarization proofs to apply to the broader family of information regular processes and construct an information regular process which does not polarize quickly. A complication in this analysis is the long, adjacent blocks acted on by the Arikan kernel, since strong mixing processes are best suited for widely separated blocks. We partially overcome this limitation by defining a promptness growth rate. Our work has applications to communication in the presence of possibly prolonged transmission outages, but our primary goal is to better understand when channel memory is incompatible with the polarization process.

This work was supported in part by the Commonwealth Cyber Initiative.

Section II contains the necessary background. In Section III, we prove polarization of information regular processes with linear promptness and related results. Section IV explicitly constructs an information regular process which is not ψ^*-mixing and shows that the rate of polarization need not be fast. Closing remarks are in Section V.

II. Background

Let (X, Y) be processes with memory, where X represents a source and Y is the estimation of an encoding of X received from a communication channel. We use $X_i \in \{0, 1\}$, but the approach readily extends to prime alphabets using ideas from [14, Appendix B]. We assume that X_i may depend on X_1^{i-1} and that Y_i can depend both on Y_1^{i-1} and on X_1^i. This is a flexible model, encompassing common issues such as burst noise, intersymbol interference, and a broad family of data sources. Our discussion will focus on how flexible this model can be made without sacrificing polarization.

Since X is discrete, we use the complete Kolmogorov model described, among other places, in [15, Section 1.1]. This σ-algebra is generated by the cylinder sets

$$[a_m^n] = \{x \in X^{\infty} : x_i = a_i \ \forall i \in \{m, \ldots, n\}\}$$

where $m, n \in \mathbb{Z}$ and X^{∞} is the set of all infinite sequences on X. We use the notation $\sigma(T_b^c)$ to indicate the sub σ-algebra generated by $[a_m^n]$ where $b \leq m, n \leq c$. Except for a few temporary refinements, this σ-algebra will be used throughout.

Even though there are a number of strong mixing conditions which satisfy a partial ordering [16] (see also [17]), most have not been applied to polar codes. We summarize existing results in the context of this framework.

Shuval showed polarization and fast polarization for finite-state, aperiodic, irreducible hidden Markov (FAIM) processes in [14]. FAIM processes represent channel memory using a hidden additional state sequence S, with the condition that

$(X_{j+1}, Y_{j+1}, S_{j+1})$ is allowed to depend only on (X_j, Y_j, S_j). The state S_j uses a finite alphabet, and for any $a, b \in S$ there is an N_0 such that $Pr_{S_j, S_{j+N}}(a, b) > 0$ for all $N > N_0$.

It was shown in [14] and [15, Section III.1.b] that FAIM processes are a subset of ψ^*-mixing processes. Şaşoğlu showed that these processes polarize in [13], but fast polarization has only been shown for the low entropy set. The ψ^*-mixing condition requires that any events $A \in \sigma(T_{-\infty}^0)$ and $B \in \sigma(T_n^\infty)$ satisfy

$$Pr(A \cap B) \leq \psi_n^* Pr(A) Pr(B)$$

where ψ_n^* is a nonincreasing sequence with $\psi_n^* \to 1$. We keep the notation as in [16] but note that [13] refers to ψ^*-mixing as ψ-mixing.

We are interested in the more general family of information regular processes, meaning those that satisfy

$$\sup \sum_{i=1}^{I} \sum_{j=1}^{J} Pr(A_i \cap B_j) \log \left(\frac{Pr(A_i \cap B_j)}{Pr(A_i) Pr(B_j)} \right) \leq I_n$$

where I_n is a nonincreasing sequence with $I_n \to 0$ and the supremum is taken over finite partitions $A_i \in \sigma(T_{-\infty}^0)$ and $B_j \in \sigma(T_n^\infty)$. Our goal will be to show that these processes polarize but that polarization need not be fast.

Our proof uses α-mixing processes [16], which are those that satisfy the following relation for any $A \in \sigma(T_{-\infty}^0)$ and $B \in \sigma(T_n^\infty)$:

$$|Pr(A \cap B) - Pr(A) Pr(B)| \leq \alpha_n$$

where α_n is nonincreasing and $\alpha_n \to 0$. We will not refer to α-mixing processes as strong mixing to avoid confusion with the broader family of strong mixing conditions. Note that ψ^*-mixing processes are information regular and information regular processes are α-mixing processes [16].

Finally, all strong mixing processes are a subset of ergodic processes, which are the strongest condition known to not guarantee polarization. Şaşoğlu provided an example in [13] of a periodic process which does not polarize.

The sequences described above are said to be prompt if α_0, I_0, ψ_0^* are finite. This notion is essential, but it is not particularly well adapted for polar codes. For a sequence c_n appearing in one of the mixing properties above, recall that we always use events chosen from $\sigma(T_{-\infty}^0), \sigma(T_n^\infty)$. Define the doubly indexed sequence $c_{n,k}$ using instead events from $\sigma(T_{-k}^0), \sigma(T_n^{n+k})$. It follows immediately that $c_{n,k}$ is finite, since both subalgebras are finite. Further, the traditional c_n can be thought of as $\lim_{k \to \infty} c_{n,k}$.

Definition 1. *A sequence $c_{n,k}$ is said to be $f(c)$-prompt if $\lim_{k \to \infty} f(c_{0,k})/k = 0$. The term prompt is used when $c_{0,k}$ is bounded.*

We will often refer to $f(c)$ by name instead of symbolically, so a process with logarithmic promptness is $\log(c)$-prompt and a process with linear promptness is c-prompt. Our analysis will focus on logarithmic promptness for ψ^*-mixing processes and linear promptness for information regular processes.

Next, we summarize some key properties relating to polarization. The main quantity of interest is the conditional entropy of a symbol U_i of the codeword given the previous symbols U_1^{i-1} and the received message Y_1^N. These variables are chosen to imitate the information available to an SC decoder. We are thus interested in $H(U_i | U_1^{i-1}, Y_1^N) =$

$$\sum_{u_i \in U_i} \sum_{u \in U_1^{i-1}} \sum_{y \in Y} Pr(u_i, u, y) \log \left(\frac{Pr(u_i, u, y)}{\sum_{u_i} Pr(u_i, u, y)} \right).$$

The goal is to relate this quantity to the average entropy per symbol in a very long codeword, defined as

$$H_{X|Y} = \lim_{n \to \infty} \frac{1}{N} H(X_1^N | Y_1^N).$$

We say that a process polarizes if the entropy of nearly all channels converges to $\{0, 1\}$, meaning that

$$\lim_{n \to \infty} \frac{1}{N} |\{i : H(U_i | U_1^{i-1}, Y_1^N) > 1 - \epsilon\}| = H_{X|Y}$$

and

$$\lim_{n \to \infty} \frac{1}{N} |\{i : H(U_i | U_1^{i-1}, Y_1^N) < \epsilon\}| = 1 - H_{X|Y}.$$

We say that the rate of polarization is fast if for all $\beta < c$, we can set $\epsilon = 2^{-N^\beta}$. For our purposes, $c = \frac{1}{2}$. Note that our definitions focus solely on the polarization of individual symbols. It is also common to consider the successful recovery of the entire low entropy set, which we leave to future work.

The inclusion of channel memory requires considerable caution regarding the relative locations of different channels, so we introduce some further notation to simplify our indexing. Note that $G_N = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{\otimes \log N}$ is the polar code generator matrix, and B_N is the bit reversal permutation matrix . Let

$$U_1^N = X_1^N B_N G_N, \qquad V_1^N = X_{N+1}^{2N} B_N G_N$$

represent the adjacent blocks combined in a polarization step, and let

$$Q_i = (U_1^{i-1}, Y_1^N), \qquad R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

represent the conditioning used in decoding.

III. Polarization

Our first goal is to better understand what sorts of channel memory support polarization. To do so, we first demonstrate that the results in [13, Lemma 8, Lemma 9] apply under weaker mixing and promptness conditions.

Lemma 1. *Suppose that X_i, Y_i are information regular processes with linear promptness. Then, as $N \to \infty$ and for any $\epsilon > 0$, the following conditions are met for a fraction of indices i which converges to 1:*

$$I(U_i; V_i | Q_i, R_i) < \epsilon,$$

$$I(U_i; R_i | Q_i) < \epsilon,$$

and

$$I(V_i; Q_i | R_i) < \epsilon.$$

Proof. Since X_i, Y_i have linear promptness,

$$I(U_1^N, Y_1^N; V_1^N, Y_{N+1}^{2N}) \leq I_{0,N}.$$

We then expand this expression as follows

$$\begin{array}{rcl} I_{0,N} & \geq & I(U_1^N, Y_1^N; V_1^N, Y_{N+1}^{2N}) \\ & \geq & I(U_1^N; V_1^N, Y_{N+1}^{2N} | Y_1^N) \\ & = & \sum_{i=1}^N I(U_i; V_1^N, Y_{N+1}^{2N} | Y_1^N, U_1^{i-1}) \\ & = & \sum_{i=1}^N I(U_i; R_i, V_i^N | Q_i) \\ & \geq & \sum_{i=1}^N I(U_i; R_i | Q_i) + I(U_i; V_i | Q_i, R_i). \end{array}$$

A sum of $2N$ positive terms is bounded by $I_{0,N}$, so the fraction of terms larger than $\sqrt{I_{0,N}/N}$ is bounded by $\sqrt{I_{0,N}/N}$. Both expressions converge to 0 by linear promptness of X_i, Y_i, so the first two inequalities hold. To show the third inequality, we repeat the above calculation with the block order reversed. □

Lemma 2. *Any process X which has logarithmic promptness in the sense of ψ^*-mixing processes has linear promptness in the sense of information regularity.*

Proof. Let $A \in \sigma(T_{-\infty}^0)$ and $B \in \sigma(T_n^\infty)$ be arbitrary events. Since X has logarithmic promptness, we know that

$$Pr(A \cap B) \leq \psi_{0,k}^* Pr(A) Pr(B)$$

and

$$\log \psi_{0,k}^* \geq \log \frac{Pr(A \cap B)}{Pr(A) Pr(B)}.$$

Substituting into the definition of information regular processes, we have

$$I_{0,k} \leq \sup \sum_{i=1}^{I} \sum_{j=1}^{J} Pr(A_i \cap B_j) \log \psi_{0,k} = \log \psi_{0,k}^*.$$

Substituting into the promptness conditions gives

$$\lim_{k \to \infty} \frac{I_{0,k}}{k} \leq \lim_{k \to \infty} \frac{\log \psi_{0,k}^*}{k} = 0.$$

□

Lemma 3. *Suppose that X_i, Y_i are α-mixing processes. For all $\epsilon > 0$, there exists N_0 so that for $N > N_0$ and any $\{0,1\}$-valued function f with associated events $A_i = f(X_{(i-1)N+1}^{iN}, Y_{(i-1)N+1}^{iN})$ the following relation is satisfied*

$$Pr_{A_1}(1) \in (\epsilon, 1-\epsilon) \Rightarrow Pr_{A_1,A_2}(1,0) > \delta(\epsilon) = \frac{\epsilon^2}{4}.$$

Proof. Our proof generalizes [13, Lemma 10] to the weaker condition that

$$Pr(A \cap B) \leq Pr(A) Pr(B) + \alpha_i$$

for some sequence $\alpha_i \to 0$. Using stationarity and the memory model, we have

$$\begin{array}{rcl} 2Pr_{A_1,A_2}(1,0) & = & Pr_{A_1,A_2}(1,0) + Pr_{A_2,A_3}(1,0) \\ & \geq & Pr_{A_1,A_2,A_3}(1,0,0) \\ & & + Pr_{A_1,A_2,A_3}(1,1,0) \\ & = & Pr_{A_1,A_3}(1,0) \\ & = & Pr_{A_1}(1) - Pr_{A_1,A_3}(1,1) \\ & \geq & Pr_{A_1}(1)(1 - Pr_{A_3}(1)) - \alpha_N \\ & \geq & \epsilon^2 - \alpha_N. \end{array}$$

The proof is completed by choosing N_0 large enough that $\alpha_N < \frac{\epsilon^2}{2}$ for all $N > N_0$. □

We are now ready to state two polarization results which apply the above lemmas

Theorem 4. *Polarization is achieved when X_i, Y_i are information regular processes with linear promptness.*

Proof. The proof consists of substituting Lemmas 1 and 3 into Şaşoğlu's argument for ψ^*-mixing processes. For the sake of brevity, we summarize the key elements in the proof to show where these lemmas are used.

The approach in [13, Theorem 1] is to bound the change in entropy introduced by each source of channel memory. First, [13, Lemma 7] demonstrates that for any stationary process, the entropies $H_n = H(U_i | U_1^{i-1}, Y_1^N)$ converge to some limit. Channel memory might affect this limit through the joint distribution over adjacent blocks U_i, V_i or collusion via the conditioning Q_i, R_i. The first of these concerns is removed in [13, Lemma 8], which shows that when Lemma 1 holds, we may use marginal distributions for U, V with negligible loss. Collusion in the conditioning is then shown in [13, Lemma 10] to be always imperfect when Lemma 3 holds. Finally, [13, Lemma 12] combines these results to reduce the proof of convergence to a classic result for the memoryless case. □

Theorem 4 allows for an assignment of entropy to each part of the communication process and demonstrates that the total entropy polarizes. The next result may be easier to use in practice.

Corollary 5. *Polarization is achieved when X_i, Y_i are α-mixing and logarithmically prompt in the sense of ψ^*-mixing processes.*

IV. Intermittent Outages

In the last section, we showed that information regular processes with linear promptness polarize. We now demonstrate that the rate of polarization might be arbitrarily slow using the following construction, inspired by a renewal process in [15, § III.1.c] whose k-blocks converge slowly in frequency.

Definition 2. *Let $w = \{w_i\}$ be a sequence satisfying $\hat{w} = \sum_{i=1}^\infty w_i \leq \frac{1}{8}$. Let $\beta_i = i w_i$ and further suppose that $\sum_{i=1}^\infty \beta_i = \frac{1}{4}$. An outage is a process X_w which independently combines the following blocks: the 0 symbol occurring with probability $\frac{1}{4}$, the 1 symbol occurring with probability $\frac{1}{2} - 2\hat{w}$, and two 1 symbols separated by a block of i 0 symbols occurring with probability w_i for each $i \geq 1$.*

We start by constructing outages explicitly using a technique from probability theory known as cutting and stacking, described in [18] and [15, § I.10]. Consider a unit interval partitioned into the subinterval $\left[0, \frac{1}{2}\right)$ labelled 1 and the subinterval $\left[\frac{1}{2}, 1\right]$ labelled 0. Recall that w_i is the probability that an arbitrary X_k is the first in a block of exactly i zero symbols and the measure of each block is β_i. To simplify the construction, let $\mathcal{C} = [0, 2\hat{w}) \cup (\frac{3}{4}, 1]$ and $\mathcal{R} = [2\hat{w}, \frac{3}{4}]$. We will apply cutting and stacking to these two intervals separately.

The role of $\mathcal{C}$ is to provide a source of 0 blocks to be used in the stacking of $\mathcal{R}$. In the ith round of cutting and stacking, we form a subinterval of measure β_i into a column of zeros with height i, capped with two blocks of ones of width w_i. The resulting column is transferred to $\mathcal{R}$ and the remainder of $\mathcal{C}$ is left for future steps.

The role of $\mathcal{R}$ is to apply independent cutting and stacking to 0, 1, and the blocks constructed from $\mathcal{C}$. We start by defining a virtual column $C_* \in \mathcal{R}$ of measure $\hat{w}$ to represent future output from $\mathcal{C}$. In each round of cutting and stacking, we apply independent cutting and stacking to $\mathcal{R}$, ignoring any step involving C_*. We then add the column C_i of length $i+2$ from $\mathcal{C}$, replace part of the virtual C_* with C_i, and reapply the operations from earlier rounds which pertain to C_i. The indirect approach to $\mathcal{R}$ is necessary since we cannot construct all of the C_i at once, so we need to independently combine columns which have not been constructed yet.

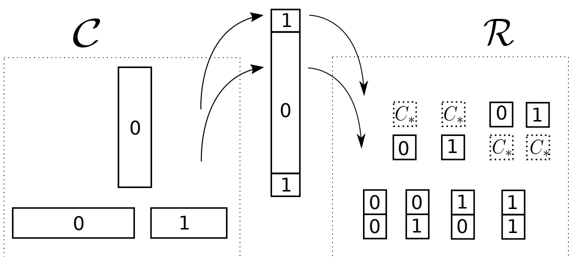

Fig. 1. An illustration of the cutting and stacking process

Theorem 6. *An outage is a stationary, prompt, information regular process.*

Proof. Stationarity follows from the use of cutting and stacking. The remaining conditions can be checked in three cases. Two nonintersecting blocks X_{-k}^0, X_n^{n+k} of length k may both lie completely in the same zero block, at least one may intersect the boundary, or there may not be a zero block intersecting both intervals. These events, which we label C_1, C_2, C_3, are not included in the σ-algebra since they use intermediate states of $\mathcal{C}$ and $\mathcal{R}$ from the construction of X. There are several ways to proceed, but the simplest is to temporarily extend our σ-algebra so C_1, C_2, C_3 are basis elements.

The last case, C_3, is the easiest to check. If there is no zero block intersecting both X_{-k}^0, X_n^{n+k}, then the two blocks are independent. Therefore, for any events $A \in \sigma(T_1^k)$ and $B \in \sigma(T_n^{n+k})$ we have that

$$\begin{aligned} &Pr(A \cap B \cap C_3) \log\left(\tfrac{Pr(A \cap B \cap C_3)}{Pr(A \cap C_3) Pr(B \cap C_3)}\right) \\ &= Pr(A \cap B \cap C_3) \log\left(\tfrac{Pr(A \cap B | C_3) Pr(C_3)}{Pr(A|C_3) Pr(B|C_3) Pr(C_3)^2}\right) \\ &= Pr(A \cap B \cap C_3) \log\left(\tfrac{1}{Pr(C_3)}\right). \end{aligned}$$

Note that this function is bounded and approaches 0 as $Pr(C_3)$ approaches $\{0, 1\}$.

In the first case, C_1, X_{-k}^0 and X_n^{n+k} have fixed values as subsets of a zero block, so $Pr(A), Pr(B) \in \{0, 1\}$. This means that $Pr(A \cap C_1), Pr(B \cap C_1), Pr(A \cap B \cap C_1) \in \{0, Pr(C_1)\}$. We exclude events with zero probability for now, and note that the others satisfy

$$\begin{aligned} &Pr(A \cap B \cap C_1) \log\left(\tfrac{Pr(A \cap B \cap C_1)}{Pr(A \cap C_1) Pr(B \cap C_1)}\right) \\ &= Pr(A \cap B \cap C_1) \log\left(\tfrac{1}{Pr(C_1)}\right). \end{aligned}$$

To show the second case, C_2, suppose wlog that the zero block starts at index i and ends at index $n+\ell$. We allow the cases $i < 0$ or $\ell > k$ so the block need not terminate inside both X_{-k}^0 and X_n^{n+k}. We know that any events $A \in \sigma(T_0^k)$ and $B \in \sigma(T_n^{n+k})$ can be represented as a union of cylinder sets $A = \cup_r A_r$ and $B = \cup_r B_r$. Each of these sets can be split at the indices i, ℓ to form $A_r = A_{1,r} \cap A_{2,r}$ and $B_r = B_{1,r} \cap B_{2,r}$ where $A_{1,r} \in \sigma(T_0^{i-1})$, $A_{2,r} \in \sigma(T_i^k)$, $B_{1,r} \in \sigma(T_n^{n+\ell})$, $B_{2,r} \in \sigma(T_{n+\ell+1}^{n+k})$. Let $A_1 = \cup_r A_{1,r}$ and define A_2, B_1, B_2 similarly. By construction, $A_{2,r}, B_{1,r}$ are events defined over a zero block, so $A_{2,r} \cap C_2, B_{1,r} \cap C_2 \in \{\emptyset, C_2\}$. Therefore,

$$\begin{aligned} A \cap C_2 = \cup_r (A_{1,r} \cap A_{2,r} \cap C_2) \in \{\emptyset, A_1 \cap C_2\} \\ B \cap C_2 = \cup_r (B_{1,r} \cap B_{2,r} \cap C_2) \in \{\emptyset, B_2 \cap C_2\} \\ A \cap B \cap C_2 \in \{\emptyset, A_1 \cap B_2 \cap C_2\}. \end{aligned}$$

Applying these substitutions, and recalling that A_1, B_2 are independent, we have reduced this case to C_3.

This technique directly extends from events over k-blocks to arbitrary events in $\sigma(T_{-\infty}^0), \sigma(T_n^\infty)$ if we extend the unions in the second case to include countable intersections of cylinder sets. We therefore need only combine the joint probabilities in C_1, C_2, C_3 to demonstrate information regularity. In the calculations that follow, the events with zero probability which were excluded earlier are inserted in each sum over C_i immediately prior to its elimination. By the log-sum inequality,

$$\begin{aligned} &\textstyle\sum_{C_i} Pr(C_i) \log\left(\tfrac{1}{Pr(C_i)}\right) \\ &= \textstyle\sum_{A_i, B_i, C_i} Pr(A_i \cap B_i \cap C_i) \log\left(\tfrac{1}{Pr(C_i)}\right) \\ &\geq \textstyle\sum_{A_i, B_i, C_i} Pr(A_i \cap B_i \cap C_i) \log\left(\tfrac{Pr(A_i \cap B_i \cap C_i)}{Pr(A_i \cap C_i) Pr(A_i \cap C_i)}\right) \\ &\geq \textstyle\sum_{A_i, B_i} Pr(A_i \cap B_i) \log\left(\tfrac{Pr(A_i \cap B_i)}{\sum_{C_i} Pr(A_i \cap C_i) Pr(B_i \cap C_i)}\right) \\ &\geq \textstyle\sum_{A_i, B_i} Pr(A_i \cap B_i) \log\left(\tfrac{Pr(A_i \cap B_i)}{Pr(A_i) Pr(B_i)}\right). \end{aligned}$$

Since this bound is finite for $n = 0$, the process is prompt. To show that the bound converges to 0, recall that the width of very long zero blocks converges to 0. So, when n is large $Pr(C_3) \to 1$ and $Pr(C_1), Pr(C_2) \to 0$. Since each term in the bound approaches 0 when $Pr(C_i)$ is close to $\{0, 1\}$, $I_n \to 0$ and the process is information regular. □

Proposition 7. *An outage X_w is not ψ^*-mixing unless there is an n_0 for which $w_n \leq c2^{-an}$ for all $n > n_0$ and fixed a, c.*

Proof. Let the events $A \in \sigma(T_{-k}^0)$ and $B \in \sigma(T_n^k)$ represent blocks of k zeros. We have three cases to examine. Each symbol in the intervals can be zero independently, the intervals can be contained in separate zero blocks, or one zero block might contain both intervals.

Suppose that a zero block of length n starting at index 1 contains the interval X_i^{i+k}. Containment requires $1 \leq i \leq$

$n-k$, and each of these values occurs with probability w_n. Therefore, the probability that a block of length k is contained in any zero block is given by

$$c_k = \sum_{i=k}^{\infty} (i-k) w_i.$$

Recalling from the definition of X_w that k independent zeros occur with probability 2^{-2k}, substituting this expression gives

$$Pr(A) = Pr(B) = c_k + 2^{-2k}$$

$$Pr(A \cap B) = (c_k - c_{n+2k} + 2^{-2k})^2 + c_{n+2k} \geq c_{n+2k}$$

We can now bound ψ_n as

$$\psi_n \geq \frac{Pr(A \cap B)}{Pr(A)Pr(B)} \geq \frac{c_{n+2k}}{(c_k + 2^{-2k})^2}.$$

If we assume that $c_k \geq w_{k+1}$ is not bounded by $a2^{-bk}$ for large k, then we can find a large k satisfying $c_k > 2^{-2k}$ and so the fraction is bounded as a function of n when $c_{n+2k} \leq ac_k^2$ for some constant a. This recurrence relation describes an exponential function, so $w_{n+1} \leq c_n \leq c2^{-an}$. □

Proposition 8. *Suppose the process X_w is used to construct a binary erasure channel in the following way: if $(X_w)_i = 0$ then insert an erasure, otherwise transmit the input symbol. A memoryless binary source X is then transmitted over this channel. By appropriate choice of (X_w) and for any sequence $\alpha_i \to 0$, we can ensure*

$$\lim_{N\to\infty} \frac{1}{N} |\{i : H(U_i|U_1^{i-1}, Y_1^N) < \alpha_N\}| < 1 - H(X|Y).$$

Proof. We start by examining the conditions under which a codeword of length N is entirely erased. In the first case, we might have N independent erasures, which occurs with probability 2^{-2N}. In the second case, a block of length N is entirely contained in a zero block. For this to happen, the start of the code must be at least N symbols before the end of the zero block, which occurs with probability $\sum_{i=N}^{\infty}(i-N)w_i$. Substituting these probabilities, we can bound the entropy as

$$\begin{aligned} H(U_i|U_1^{i-1}, Y_1^N) &\geq Pr_{Y_1^N}(e^N) H(U_i|U_1^{i-1}, Y_1^N = e^N) \\ &= H(X)\left(2^{-2N} + \textstyle\sum_{i=N}^{\infty}(i-N)w_i\right). \end{aligned}$$

$H(X)$ is fixed and w_i can be chosen so that $\sum_{i=N}^{\infty}(i-N)w_i$ decays arbitrarily slowly, so an outage X_w exists with

$$\lim_{N\to\infty} \frac{1}{N} |\{i : H(U_i|U_1^{i-1}, Y_1^N) < \alpha_N\}| = 0 < 1 - H(X|Y).$$

□

The example in Proposition 8 has a useful physical interpretation, which motivates the name "outage." Suppose the source X is encoded and then transmitted over a wire. In addition to the usual memoryless transmission errors, the wire is occasionally damaged and some time elapses before it can be repaired. Then w_i is the probability that communication is disrupted starting at time t and $\sum \beta_i = \frac{1}{4}$ is the probability that a particular symbol is disrupted in this manner. We have demonstrated that polarization occurs when both transmitter and receiver are oblivious to arbitrarily long outages, but that the asymptotic rate depends heavily on the distribution of outages with different lengths.

V. Conclusion

We have demonstrated that information regular processes with linear promptness polarize, allowing the use of entropy to describe every part of the communication process. In addition, we constructed an information regular process which models transmission over a communication channel suffering from occasional lengthy outages. Finally, we showed that this example is not ψ^*-mixing and that the distribution of outages affects the rate of polarization.

References

[1] E. Arikan, "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, 2009.

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

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

[4] R. Mori and T. Tanaka, "Source and channel polarization over finite fields and reed–solomon matrices," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2720–2736, 2014.

[5] R. Nasser, "An ergodic theory of binary operations—part i: Key properties," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6931–6952, 2016.

[6] ——, "An ergodic theory of binary operations—part ii: Applications to polarization," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1063–1083, 2017.

[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, 2013.

[8] R. Nasser, "On the convergence of the polarization process in the noisiness/weak-∗ topology," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 2099–2103.

[9] N. Goela and M. Raginsky, "Channel polarization through the lens of blackwell measures," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6222–6241, 2020.

[10] A. Bhatia, V. Taranalli, P. H. Siegel, S. Dahandeh, A. R. Krishnan, P. Lee, D. Qin, M. Sharma, and T. Yeo, "Polar codes for magnetic recording channels," in *2015 IEEE Information Theory Workshop (ITW)*. IEEE, 2015, pp. 1–5.

[11] E. Sasoglu, "Polarization in the presence of memory." in *ISIT*, 2011, pp. 189–193.

[12] 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-Fall (ITW)*. IEEE, 2015, pp. 187–191.

[13] E. Şaşoğlu and I. Tal, "Polar coding for processes with memory," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 1994–2003, 2019.

[14] B. Shuval and I. Tal, "Fast polarization for processes with memory," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2004–2020, 2018.

[15] P. C. Shields, *The ergodic theory of discrete sample paths*. American Mathematical Soc., 1996, vol. 13.

[16] R. C. Bradley, "Basic properties of strong mixing conditions. a survey and some open questions," *arXiv preprint math/0511078*, 2005.

[17] ——, "Basic properties of strong mixing conditions," in *Dependence in probability and statistics*. Springer, 1986, pp. 165–192.

[18] P. C. Shields, "Cutting and stacking: a method for constructing stationary processes," *IEEE transactions on information theory*, vol. 37, no. 6, pp. 1605–1617, 1991.