Strong Polarization for Shortened and Punctured Polar Codes

Boaz Shuval, Ido Tal
The Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering,
Technion, Haifa 32000, Israel.
Email: {bshuval@, idotal@ee.}technion.ac.il

***Abstract*—Polar codes were originally specified for codelengths that are powers of two. In many applications, it is desired to have a code that is not restricted to such lengths. Two common strategies of modifying the length of a code are shortening and puncturing. Simple and explicit schemes for shortening and puncturing were introduced by Wang and Liu, and by Niu, Chen, and Lin, respectively. In this paper, we prove that both schemes yield polar codes that are capacity achieving. Moreover, the probability of error for both the shortened and the punctured polar codes decreases to zero at the same exponential rate as seminal polar codes. These claims hold for *all* codelengths large enough.**

I. Introduction

Polar codes [1] are based on a recursive transform, yielding codes whose codelengths are powers of two. They have been proven to achieve the capacity of many channel settings [2]–[21]. Often, it is desirable to transmit a message whose length is not limited to a power of 2. Shortening and puncturing [22, Problems 2.3 and 2.14], [23, Chapter 1§9] are two common methods of reducing the length of a given code. Such methods were extensively studied for polar codes, see [24]–[30] and the references therein. In this paper, we focus on the puncturing method of [24] and the shortening method of [25][1]. In the sequel, for brevity, we will refer to transforms based on these methods as the "shortening transform" and "puncturing transform," respectively. We show that these schemes achieve capacity, with probability of error decreasing at the same exponential rate as seminal polar codes. This holds for all codelengths large enough. For simplicity, we focus on the setting of a binary-input memoryless channel, which may be non-symmetric (BM channel).

The following theorem is a shortened version[2] of our main result. It will follow as a straightforward corollary of the more general Theorem 6. It assumes a fixed input distribution $p(x)$ and a fixed BM channel $W(y|x)$. We denote by $Z(X|Y)$ and $K(X|Y)$ the conditional Bhattacharyya parameter and the total variation distance, respectively (see [5, Definitions 2 and 3]). Furthermore for X and Y with joint distribution $W(x;y) \triangleq p(x)W(y|x)$, we denote by $H(X|Y)$ the conditional entropy of X given Y and by $H(X)$ the entropy of X.

Theorem 1. *Let $\mathbf{X}$ be a random vector of length M with i.i.d. entries, each sampled from an input distribution $p(x)$. Let $\mathbf{Y}$ be the result of passing $\mathbf{X}$ through a BM channel $W(y|x)$. Let $\mathbf{U}$ of length M be the result of transforming $\mathbf{X}$ via either the shortening transform or the puncturing transform. Fix $0 < \beta < 1/2$. Then,*

$$\lim_{M\to\infty} \frac{1}{M} \left|\left\{i : Z(U_i|U^{i-1}, \mathbf{Y}) < 2^{-M^\beta}\right\}\right| = 1 - H(X|Y), \quad (1)$$

$$\lim_{M\to\infty} \frac{1}{M} \left|\left\{i : K(U_i|U^{i-1}) < 2^{-M^\beta}\right\}\right| = H(X). \quad (2)$$

The above theorem implies that, similar to the power-of-two setting, we can use successive cancellation and the Honda-Yamamoto scheme [2] to define a code whose rate approaches $I(X;Y)$ and whose probability of error is upper bounded by 2^{-M^β} for $0 < \beta < 1/2$ fixed and *all* integer M large enough. Moreover, both encoding and decoding can be calculated in time $O(M \log M)$.

II. The shortening and Puncturing Transforms

In this section we define both the shortening and the puncturing transforms. To do so, for a given codelength M, we denote by N the smallest power of two greater or equal to M. That is,

$$N = 2^{\lceil \log_2 M \rceil}. \quad (3)$$

We also denote

$$n = \lceil \log_2 M \rceil = \log_2 N. \quad (4)$$

Since we will make heavy use of bit-reversals, it is natural to use zero-based indexing. That is, an index $0 \le i < N$ has binary representation $i = \sum_{j=0}^{n-1} b_j 2^j$. The corresponding vector is $\mathbf{b} = \begin{bmatrix} b_0 & b_1 & \cdots & b_{n-1} \end{bmatrix}$. The reversed vector is $\overleftarrow{\mathbf{b}} = \begin{bmatrix} b_{n-1} & b_{n-2} & \cdots & b_0 \end{bmatrix}$. The corresponding bit-reversed index is $\overleftarrow{i} = \sum_{j=0}^{n-1} b_j 2^{n-1-j}$.

A. Generalization of Key Polar Coding Concepts

Seminal polar codes revolve around three key concepts:

- The polar transform, an invertible transform that transforms a vector $\mathbf{x}$ of bits to a vector $\mathbf{u}$ of bits, both of the same length $N = 2^n$.
- The '−' and '+' operations, denoted $\boxminus$ and $\circledast$, respectively. They transform two joint distributions A and B into new joint distributions, $A \boxminus B$ and $A \circledast B$, respectively. This is a slight generalization of the seminal setting, in which A and B were the same distribution, in which case $A \boxminus A$ was denoted A^- and $A \circledast A$ was denoted A^+.
- The connection between the polar transform and the '−' and '+' operations.

We now briefly review these concepts and show how to generalize them to the shortening and puncturing setting.

[1] The title of [25] claims a puncturing method, but in fact describes a shortening method.

[2] Or is it a punctured version?

1) The Polar Transform: The seminal polar transform takes a vector $\mathbf{x}$ of length $N = 2^n$ and produces a transformed vector $\mathbf{u}$, also of length N. A simple way to define this transform is by two operations that take a vector of length N and produce a vector of length $N/2$. Namely,

$$\begin{aligned}&\begin{bmatrix}x_0 & x_1 & \cdots & x_{N-1}\end{bmatrix}^{[0]}\\ &\qquad = \begin{bmatrix}x_0 \oplus x_1 & x_2 \oplus x_3 & \cdots & x_{N-2} \oplus x_{N-1}\end{bmatrix}\end{aligned}\tag{5}$$

and

$$\begin{aligned}&\begin{bmatrix}x_0 & x_1 & \cdots & x_{N-1}\end{bmatrix}^{[1]}\\ &\qquad = \begin{bmatrix}x_0 \triangleright x_1 & x_2 \triangleright x_3 & \cdots & x_{N-2} \triangleright x_{N-1}\end{bmatrix},\end{aligned}\tag{6}$$

where[3] $\alpha \triangleright \beta = \beta$. We denote for $\mathbf{b} = \begin{bmatrix}b_0 & b_1 & \cdots & b_{\ell-1}\end{bmatrix}$,

$$\mathbf{x}^{[\mathbf{b}]} = \left(\cdots\left(\left(\mathbf{x}^{[b_0]}\right)^{[b_1]}\right)\cdots\right)^{[b_{\ell-1}]},\tag{7}$$

that is, the result of recursively applying $(\cdot)^{[0]}$ and $(\cdot)^{[1]}$ operations. Then, entry $i = \sum_{j=0}^{n-1} b_j 2^j$ of $\mathbf{u}$ is $\mathbf{x}^{[\overleftarrow{\mathbf{b}}]}$, where $\mathbf{b} = \begin{bmatrix}b_0 & b_1 & \cdots & b_{n-1}\end{bmatrix}$.

We now extend the definitions of operations $\oplus$ and $\triangleright$ to apply over the set $\{0, 1, \mathtt{s}, \mathtt{p}\}$. Here, s represents a shortened bit and p a punctured bit. Namely, the generalizations of both operations are given in the following tables, which are to be read as $\alpha \cdot \beta$ with α a row and β a column. E.g., $1 \triangleright 0 = 0$.

$$\begin{array}{c|cccc}\oplus & 0 & 1 & \mathtt{s} & \mathtt{p}\\ \hline 0 & 0 & 1 & 0 & \emptyset\\ 1 & 1 & 0 & 1 & \emptyset\\ \mathtt{s} & \emptyset & \emptyset & \mathtt{s} & \emptyset\\ \mathtt{p} & \mathtt{p} & \mathtt{p} & \mathtt{p} & \mathtt{p}\end{array}\quad ; \quad \begin{array}{c|cccc}\triangleright & 0 & 1 & \mathtt{s} & \mathtt{p}\\ \hline 0 & 0 & 1 & \mathtt{s} & \emptyset\\ 1 & 0 & 1 & \mathtt{s} & \emptyset\\ \mathtt{s} & \emptyset & \emptyset & \mathtt{s} & \emptyset\\ \mathtt{p} & 0 & 1 & \mathtt{s} & \mathtt{p}\end{array}\ .\tag{8}$$

In the above, $\emptyset$ denotes the "don't care" value. That is, s will never be the first argument, unless the second argument is s, and p will never be the second argument, unless the first argument is p. Also, although this is a setting we do not consider further in this paper, note that the above table implies that we can have both shortened and punctured bits in our codeword.

2) The '$-$' and '$+$' Operations: In the seminal setting, the '$-$' and '$+$' operations each transform two identical channels into a new channel. Here, they each transform two joint distributions into a new joint distribution. That is, let $A(x_0; y_0)$ be the joint distribution on the pair $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}_0$, where henceforth $\mathcal{X} = \{0, 1\}$. Further let $B(x_1; y_1)$ be the joint distribution on the pair $(x_1, y_1) \in \mathcal{X} \times \mathcal{Y}_1$. Then,

$$(A \boxplus B)(u_0; y_0, y_1) = \sum_{x_1 \in \mathcal{X}} A(u_0 \oplus x_1; y_0) B(x_1; y_1),\tag{9}$$

$$(A \circledast B)(u_1; u_0, y_0, y_1) = A(u_0 \oplus u_1; y_0) B(u_1; y_1).\tag{10}$$

We now define two special joint distributions, $\mathtt{S}$ and $\mathtt{P}$, corresponding to a "shortened" distribution and a "punctured" distribution, respectively. Both $\mathtt{S}$ and $\mathtt{P}$ are over $\mathcal{X} \times \{?\}$. They are given by

$$\mathtt{S}(x; y) = \begin{cases}1, & x = 0, y = ?,\\ 0, & \text{otherwise},\end{cases}\tag{11}$$

$$\mathtt{P}(x; y) = \left\{\tfrac{1}{2}, \quad x \in \mathcal{X}, y = ?.\right.\tag{12}$$

For reasons that will become clearer later, we call $\mathtt{S}$ the 'superb' distribution and $\mathtt{P}$ the 'pitiful' distribution.

3) The Connection between the Polar Transform and the '$-$' and '$+$' Operations: Consider the vector of joint distributions $\mathbf{A} = \begin{bmatrix}A_0 & A_1 & \cdots & A_{N-1}\end{bmatrix}$. We define $\mathbf{A}^{[0]}$, $\mathbf{A}^{[1]}$, and $\mathbf{A}^{[\mathbf{b}]}$ by adapting (5), (6), and (7), respectively. We adapt these by replacing x_i with A_i, $\oplus$ with $\boxplus$, and $\triangleright$ with $\circledast$.

Let $0 \le i < N$ with binary representation $i = \sum_{j=0}^{n-1} b_j 2^{n-1-j}$. Then, there exists an invertible function f such that

$$\mathbb{P}\left(U_i = u_i; U_0^{i-1} = u_0^{i-1}, \mathbf{Y} = \mathbf{y}\right) = \mathbf{A}^{[\mathbf{b}]}(u_i; f(u_0^{i-1}, \mathbf{y})).$$

B. The Shortening Transform

For a general (not necessarily polar) code $\mathcal{C}$ of length N, shortening is defined through an index set $\mathcal{S}$. Namely, to shorten $\mathcal{C}$, we first consider the subset of codewords $\mathbf{c} \in \mathcal{C}$ for which $c_i = 0$ for all $i \in \mathcal{S}$. For every such codeword, since we know the values at the indices $\mathcal{S}$, there is no point in transmitting them. Hence, the shortened code is the above subset, after removing the indices $\mathcal{S}$. Note that the shortened code has length $N - |\mathcal{S}|$.

In the Wang-Liu shortening scheme [25],

$$\mathcal{S} = \{\overleftarrow{N{-}1}, \overleftarrow{N{-}2}, \ldots, \overleftarrow{N{-}M}\}.\tag{13}$$

That is, the *last* $N - M$ bits of the codeword, before bit reversal, are constrained to be 0. This implies that the last $N - M$ entries of the corresponding transformed vector are frozen to 0. Successive-cancellation (SC) decoding is performed exactly as for seminal polar codes, save for setting a log-likelihood ratio (LLR) value of infinity to the shortened bits. See [25] for details.

We define the shortening transform of a vector $\mathbf{x}$ of M bits in two equivalent ways. In the first way, we define a vector $\bar{\mathbf{x}}$ of length $N = 2^{\lceil \log_2 M \rceil}$ with indices $\mathcal{S}$ set to s. We then copy $\mathbf{x}$ into $\bar{\mathbf{x}}$ in order. That is, removing from $\bar{\mathbf{x}}$ the indices in $\mathcal{S}$ recovers $\mathbf{x}$. Next, we compute $\bar{\mathbf{u}}$, as explained in Section II-A1. We note that by the special choice of $\mathcal{S}$, we will never encounter an '$\emptyset$' entry in (8). Lastly, we define $\mathbf{u}$ by the result of removing the last $N - M$ entries from $\bar{\mathbf{u}}$. We remark in passing that these removed entries were all equal to s.

Observe that had we replaced s with 0 in (8), no contradiction would have arisen. Thus, in the spirit of shortening, had we replaced s with 0 in the extension from $\mathbf{x}$ to $\bar{\mathbf{x}}$, then the last $N - M$ entries in $\bar{\mathbf{u}}$ would also have been 0, and $\mathbf{u}$ would have been the same as that from the previous paragraph. This is the second way of defining the shortening transform: replace all s in the above with 0.

Remark 1. Note that $\mathbf{u}$ equals the prefix of length M of $\bar{\mathbf{u}}$. That is, for $0 \le i < M$, $u_i = \bar{u}_i$.

C. The Puncturing Transform

Similar to shortening, for a general code $\mathcal{C}$ of length N, puncturing is defined through an index set $\mathcal{P}$. Namely, to puncture $\mathcal{C}$, we simply remove the indices $\mathcal{P}$ from the codeword. The punctured code has length $N - |\mathcal{P}|$.

In the Niu-Chen-Lin puncturing scheme [24],

$$\mathcal{P} = \{\overleftarrow{0}, \overleftarrow{1}, \ldots, \overleftarrow{N - M{-}1}\}.$$

That is, the *first* $N - M$ bits of the codeword, before bit reversal, are removed. This implies that the first $N - M$ entries of

[3] The notation $\triangleright$ is suggestive of an arrowhead pointing at the output of the operation.

the corresponding transformed vector are frozen. Successive-cancellation (SC) decoding is performed exactly as for seminal polar codes, save for setting a log-likelihood ratio (LLR) value of zero to the punctured bits. See [24] for details.

The puncturing transform of a vector $\mathbf{x}$ of M bits is also defined in two equivalent ways. In the first way, we define a vector $\tilde{\mathbf{x}}$ of length $N = 2^{\lceil \log_2 M \rceil}$ with indices $\mathcal{P}$ set to $\mathtt{p}$. We then copy $\mathbf{x}$ into $\tilde{\mathbf{x}}$ in order. That is, removing from $\tilde{\mathbf{x}}$ the indices in $\mathcal{P}$ recovers $\mathbf{x}$. Next, we compute $\tilde{\mathbf{u}}$, as explained in Section II-A1. We note that by the special choice of $\mathcal{P}$, we will never encounter a '$\emptyset$' entry in (8). Lastly, we define $\mathbf{u}$ as the result of removing the first $N - M$ entries from $\tilde{\mathbf{u}}$.

Observe that had we replaced the entries in $\mathcal{P}$ with arbitrary binary numbers, the last M entries of $\tilde{\mathbf{u}}$ would have been the same as the construction above. This is not surprising, since the generator matrix of the seminal polar codes is upper-triangular, after we apply bit reversal to the columns. This is the second way of defining the puncturing transform: replace every $\mathtt{p}$ with an arbitrary bit.

Remark 2. Note that $\mathbf{u}$ equals the suffix of length M of $\tilde{\mathbf{u}}$. That is, for $0 \le i < M$, $u_i = \tilde{u}_{i+|\mathcal{P}|} = \tilde{u}_{i+N-M}$.

III. The 'Inferior' and 'Improved' Relations

In this section, we define the 'inferior' and 'improved' relations between two joint distributions. Throughout, let $A(x_0; y_0)$ and $B(x_1; y_1)$ be joint distributions over $\mathcal{X} \times \mathcal{Y}_0$ and $\mathcal{X} \times \mathcal{Y}_1$, respectively. We denote that A is inferior to B by $A \sqsubseteq B$ and that A is improved from B by $A \sqsupseteq B$. In fact, we only need to specify when $A \sqsubseteq B$ holds, since $A \sqsubseteq B$ if and only if $B \sqsupseteq A$.

To define the 'inferior' relation, we define two auxiliary relations between joint distributions.

- **Degradation:** We say that A is (stochastically) degraded from B, denoted $A \overset{\mathrm{d}}{\sqsubseteq} B$, if there exists a conditional distribution $Q(y_0|y_1)$ over $\mathcal{Y}_0 \times \mathcal{Y}_1$ such that
$$A(x_0; y_0) = \sum_{y_1} B(x_0; y_1) Q(y_0|y_1). \tag{14}$$
- **Input Permutation:** We say that A has undergone an input permutation, resulting in A' if there exists a function $f : \mathcal{Y}_0 \to \mathcal{X}$ such that
$$A'(x_0; y_0) = A(x_0 \oplus f(y_0); y_0). \tag{15}$$
We denote this by $A' \overset{\mathrm{p}}{\sqsubseteq} A$. Note that, like A, A' is defined over $\mathcal{X} \times \mathcal{Y}_0$.

We now define that $A \sqsubseteq B$ if we can identify a finite sequence of 'degradation' and 'input permutation' relations that will lead to A from B. In other words, there exists $0 < t < \infty$, a sequence of joint distributions $C_1, C_2, \ldots, C_{t-1}$, and a sequence $\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_t \in \{\mathrm{d}, \mathrm{p}\}$ such that
$$A \overset{\mathrm{r}_1}{\sqsubseteq} C_1 \overset{\mathrm{r}_2}{\sqsubseteq} C_2 \overset{\mathrm{r}_3}{\sqsubseteq} \cdots \overset{\mathrm{r}_{t-1}}{\sqsubseteq} C_{t-1} \overset{\mathrm{r}_t}{\sqsubseteq} B. \tag{16}$$

Note that, essentially by definition, $\sqsubseteq$ is a transitive relation.

A. Order Preservation

For a joint distribution $A(x_0; y_0)$, we denote by $Z(A), K(A), H(A)$ the Bhattacharyya parameter $Z(X_0|Y_0)$, the total variation distance $K(X_0|Y_0)$, and the conditional entropy $H(X_0|Y_0)$, respectively, where (X_0, Y_0) are distributed according to A. It is well known that if $A \overset{\mathrm{d}}{\sqsubseteq} B$, then $Z(A) \ge Z(B)$, $K(A) \le K(B)$, and $H(A) \ge H(B)$. The following lemma asserts that these inequalities also hold for $\sqsubseteq$.

Lemma 2. *If $A \sqsubseteq B$, then $Z(A) \ge Z(B)$, $K(A) \le K(B)$, and $H(A) \ge H(B)$.*

Proof: By definition of $\sqsubseteq$, and since the assertion in the lemma holds when $\sqsubseteq$ is replaced by $\overset{\mathrm{d}}{\sqsubseteq}$, it suffices to show that it holds when $\sqsubseteq$ is replaced by $\overset{\mathrm{p}}{\sqsubseteq}$. This follows easily. ■

It is also well known that both $\boxast$ and $\circledast$ preserve $\overset{\mathrm{d}}{\sqsubseteq}$. The following lemma generalizes this to $\sqsubseteq$.

Lemma 3. *Let $A' \sqsubseteq A$ and $B' \sqsubseteq B$, then*
$$A' \boxast B' \sqsubseteq A \boxast B \quad \textit{and} \quad A' \circledast B' \sqsubseteq A \circledast B.$$

Proof: See Appendix. ■

The following lemma gives credence to names 'superb' and 'pitiful' for $\mathtt{S}$ and $\mathtt{P}$. Namely, it shows that $\mathtt{S}$ is 'improved' with respect to all other distributions while $\mathtt{P}$ is 'inferior' to all other distributions.

Lemma 4. *Let $A(x_0; y_0)$ be a joint distribution over $\mathcal{X} \times \mathcal{Y}_0$. Then,*
$$\mathtt{P} \sqsubseteq A \sqsubseteq \mathtt{S}.$$

Proof: See Appendix. ■

B. The Equivalence Relation and Resulting Simplifications

If $A \sqsubseteq B$ and $B \sqsubseteq A$, we denote $A \equiv B$ and call this the 'equivalence' relation.

Above, we defined the special distributions $\mathtt{S}$ and $\mathtt{P}$. In the shortened (punctured) transform, these distributions replace the distribution $W(x; y)$ in the indices $\mathcal{S}$ ($\mathcal{P}$). Hence, they will take part in '−' and '+' operations ('$\boxast$' and '$\circledast$'). The following lemma shows that the results of such transforms involving $\mathtt{S}$ and $\mathtt{P}$ can be simplified using the equivalence relation.

Lemma 5. *Let A and B be joint distributions. The following table summarizes the results of applying $\boxast$ and $\circledast$ operations to combinations of A, B, $\mathtt{S}$, and $\mathtt{P}$, up to equivalence.*

$$\begin{array}{c|ccc} \boxast & B & \mathtt{S} & \mathtt{P} \\ \hline A & A \boxast B & A & \mathtt{P} \\ \mathtt{S} & B & \mathtt{S} & \mathtt{P} \\ \mathtt{P} & \mathtt{P} & \mathtt{P} & \mathtt{P} \end{array} \; ; \quad \begin{array}{c|ccc} \circledast & B & \mathtt{S} & \mathtt{P} \\ \hline A & A \circledast B & \mathtt{S} & A \\ \mathtt{S} & \mathtt{S} & \mathtt{S} & \mathtt{S} \\ \mathtt{P} & B & \mathtt{S} & \mathtt{P} \end{array} \, . \tag{17}$$

Proof: See Appendix. ■

Remark 3. Tables (8) and (17) are connected by substitution. Namely, if in (17) we replace $\mathtt{S}$, $\mathtt{P}$, $\boxast$, $\circledast$ with $\mathtt{s}$, $\mathtt{p}$, $\oplus$, $\triangleright$, then it is consistent with (8), if we now think of A and B as bits.

Remark 4. The distribution $\mathtt{S}$ ($\mathtt{P}$) is consistent with the second way of defining the shortening (puncturing) transform. Namely, consider the pair of random vectors $\mathbf{X}, \mathbf{Y}$ of length M, drawn i.i.d. according to $W(x; y)$.

- *Shortening*: Let $\bar{\mathbf{X}}$ be the random vector defined in the second way of shortening. By definition, all entries $\bar{X}_i$ for $i \in \mathcal{S}$ are 0 with probability 1. Also, since for $i \in \mathcal{S}$ we do not transmit the corresponding symbol over the channel, $\bar{Y}_i = ?$. Thus, pairs $(\bar{X}_i, \bar{Y}_i)$ for $i \in \mathcal{S}$ are distributed according to $\mathtt{S}$. As a consequence of this and Remark 1, for $0 \le i < M$,
$$Z(U_i|U^{i-1}, \mathbf{Y}) = Z(\bar{U}_i|\bar{U}^{i-1}, \bar{\mathbf{Y}}), \tag{18}$$
$$K(U_i|U^{i-1}, \mathbf{Y}) = K(\bar{U}_i|\bar{U}^{i-1}, \bar{\mathbf{Y}}). \tag{19}$$
- *Puncturing*: Let $\tilde{\mathbf{X}}$ be the random vector defined in the second way of puncturing. By definition, we do not care about the value nor the distribution of any entry $\tilde{X}_i$ for $i \in \mathcal{P}$. However, we find it useful to set their distribution to be uniform and i.i.d. Also, since for $i \in \mathcal{P}$ we do not transmit the corresponding symbol over the channel, $\tilde{Y}_i = ?$. Thus, pairs $(\tilde{X}_i, \tilde{Y}_i)$ for $i \in \mathcal{P}$ are distributed according to $\mathtt{P}$. The reason for this choice is that now $\tilde{U}_0^{N-M-1}$ is independent of the triplet $\tilde{U}_{N-M}^{N} = \mathbf{U}$, $\mathbf{X}$, and $\mathbf{Y}$. This follows from the observation at the end of Section II-C. Thus, for $0 \le i < M$,
$$Z(U_i|U^{i-1}, \mathbf{Y}) = Z(\tilde{U}_{i+M}|\tilde{U}^{i+M-1}, \tilde{\mathbf{Y}}), \tag{20}$$
$$K(U_i|U^{i-1}, \mathbf{Y}) = K(\tilde{U}_{i+M}|\tilde{U}^{i+M-1}, \tilde{\mathbf{Y}}). \tag{21}$$

IV. Main Theorem

The following theorem is the more general form of Theorem 1. Indeed, Theorem 1 is a special case of Theorem 6, where we obtain (2) from (23) by defining $W(x; y)$ as being over $\mathcal{X} \times \{?\}$.

Theorem 6. *Let $W(x; y)$ be a joint distribution over $\mathcal{X} \times \mathcal{Y}$. Let $\mathbf{X}, \mathbf{Y}$ be a pair of random vectors of length M, with each (X_i, Y_i) sampled independently from W. Let $\mathbf{U}$ of length M be the result of transforming $\mathbf{X}$ via either the shortening transform or the puncturing transform. Fix $0 < \beta < 1/2$ and $\epsilon > 0$. Then, there exists M_0 such that for all $M \ge M_0$,*
$$\frac{1}{M}\left|\left\{i : Z(U_i|U^{i-1}, \mathbf{Y}) < 2^{-M^{\beta}}\right\}\right| > 1 - H(X|Y) - \epsilon, \tag{22}$$
$$\frac{1}{M}\left|\left\{i : K(U_i|U^{i-1}, \mathbf{Y}) < 2^{-M^{\beta}}\right\}\right| > H(X|Y) - \epsilon. \tag{23}$$

The proof will be divided into two conceptual stages. In the first, we limit M to be of a special form. That is, for some fixed t, $M = a \cdot 2^{n-t}$, where $a \in \{2^{t-1}+1, 2^{t-1}+2, \ldots, 2^t\}$. In the second stage, we show that such a restriction is not necessary.

The first stage is given in the following lemma.

Lemma 7. *Let $W(x; y)$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{U}$ be as in Theorem 6. Fix $0 < \beta' < 1/2$ and $\epsilon' > 0$. Fix integers $t > 0$ and $a \in \{2^{t-1}+1, 2^{t-1}+2, \ldots, 2^t\}$. There exists n_0 such that for all $n \ge n_0$, if $M = a \cdot 2^{n-t}$, then for $N = 2^n$,*
$$\frac{1}{M}\left|\left\{i : Z(U_i|U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}}\right\}\right| > 1 - H(X|Y) - \epsilon', \tag{24}$$
$$\frac{1}{M}\left|\left\{i : K(U_i|U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}}\right\}\right| > H(X|Y) - \epsilon'. \tag{25}$$

Observe that in (24) and (25), the inequality is given in terms of $N \ge M$ in the exponential, and thus is stronger than had it been given in terms of M, as is done in Theorem 6.

Proof: The proofs for the shortening case and the puncturing case are similar. We show here in detail the proof for the shortening case. First, note that n is consistent with the first equality in (4), and indeed $N = 2^n$ as in (3). For $0 \le i < N$, define the joint distribution A_i as
$$A_i = \begin{cases} W, & i \notin \mathcal{S}, \\ \mathtt{S}, & i \in \mathcal{S}. \end{cases}$$
Note that by our choice of $\mathcal{S}$ in (13) and the special structure $M = a \cdot 2^{n-t}$, the vector of joint distributions $\begin{bmatrix} A_0 & A_1 & \cdots & A_{N-1} \end{bmatrix}$ has period 2^t. Indeed, consider the subvector $\begin{bmatrix} A_{2^t \cdot k} & A_{2^t \cdot k+1} & \cdots & A_{2^t \cdot k + 2^t - 1} \end{bmatrix}$, for $0 \le k < 2^{n-t}$. When bit reversing its entries, we get
$$\begin{bmatrix} \underbrace{W \quad W \quad \cdots \quad W}_{a} & \underbrace{\mathtt{S} \quad \mathtt{S} \quad \cdots \quad \mathtt{S}}_{2^t - a} \end{bmatrix}. \tag{26}$$
As a consequence, for any $\mathbf{b}_{(t)} = \begin{bmatrix} b_0 & b_1 & \cdots & b_{t-1} \end{bmatrix} \in \{0,1\}^t$, all the entries of
$$\begin{bmatrix} A_0 & A_1 & \cdots & A_{N-1} \end{bmatrix}^{[\mathbf{b}_{(t)}]}$$
are equal, i.e., the same joint distribution. Denote this distribution by $\Omega_{\mathbf{b}_{(t)}}$. Observe from (26) that the mean conditional entropy of all such $\Omega_{\mathbf{b}_{(t)}}$ is $(a \cdot H(X|Y) + (2^t - a) \cdot 0)/2^t = a \cdot 2^{-t} \cdot H(X|Y) = M/N \cdot H(X|Y)$.

We are now in the scenario of identical distributions, undergoing a seminal polar transform of depth $n - t$. Calling upon standard results in polar codes[4], there exists an n_0 such that for all $n > n_0$,
$$\frac{1}{N}\left|\left\{i : Z(\bar{U}_i|\bar{U}^{i-1}, \bar{\mathbf{Y}}) < 2^{-\left(\frac{N}{2^t}\right)^{\beta''}}\right\}\right| > 1 - \frac{M}{N}H(X|Y) - \epsilon'',$$
$$\frac{1}{N}\left|\left\{i : K(\bar{U}_i|\bar{U}^{i-1}, \bar{\mathbf{Y}}) < 2^{-\left(\frac{N}{2^t}\right)^{\beta''}}\right\}\right| > \frac{M}{N}H(X|Y) - \epsilon'',$$
where $\epsilon'' = \epsilon'/2$ and $\beta'' = \frac{\beta' + \frac{1}{2}}{2}$. Note that $\epsilon'' < \epsilon' \cdot M/N$.

Recall that in the first way of describing the shortening transform, the last $N - M$ entries of $\bar{\mathbf{u}}$ are all $\mathtt{S}$. Thus, the joint distributions $(\bar{U}_i; \bar{U}^{i-1}, \bar{\mathbf{Y}})$, where $M \le i < N$, are all equivalent to $\mathtt{S}$, by Remarks 3 and 4. Hence, for $M \le i < N$, $Z(\bar{U}_i|\bar{U}^{i-1}, \bar{\mathbf{Y}}) = 0$ and $K(\bar{U}_i|\bar{U}^{i-1}, \bar{\mathbf{Y}}) = 1$. Therefore, if we limit i in the braces to $0 \le i < M$, recall that $\bar{Y}_{N-M}^{N} = ?? \cdots ?$, and use Remark 1, we obtain
$$\frac{1}{N}\left|\left\{0 \le i < M : Z(U_i|U^{i-1}, \mathbf{Y}) < 2^{-\left(\frac{N}{2^t}\right)^{\beta''}}\right\}\right| > \frac{M}{N} - \frac{M}{N}H(X|Y) - \epsilon'',$$
$$\frac{1}{N}\left|\left\{0 \le i < M : K(U_i|U^{i-1}, \mathbf{Y}) < 2^{-\left(\frac{N}{2^t}\right)^{\beta''}}\right\}\right| > \frac{M}{N}H(X|Y) - \epsilon''.$$
Multiplying both sides by N/M and further requiring that n_0 be large enough so that $(N/2^t)^{\beta''} > N^{\beta'}$, which is possible as $\beta'' > \beta'$, completes the proof for the shortening case.

[4] The inequality on Z is given in [31], while for K we can, for example, combine [5, Prop. 4] with [32, Lemma 2].

In the puncturing case, we apply the above mechanics, extending $\mathbf{U}$ and $\mathbf{Y}$ to $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{Y}}$, respectively. We then need to consider only the suffix of $\tilde{\mathbf{U}}$, due to Remark 4. ■

The following corollary strengthens Lemma 7 by setting a single n_0 that holds for all $a \in \{2^{t-1}, 2^{t-1}+1, \ldots, 2^t\}$. Here the range of a is extended to also contain 2^{t-1}. Note that n and $N = 2^n$ are not consistent with (3) and (4) for $a = 2^{t-1}$.

Corollary 8. *Let $W(x;y)$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{U}$ be as in Theorem 6. Fix $0 < \beta' < 1/2$, $\epsilon' > 0$ and $t > 0$. There exists n_0 such that for $\underline{all}$ $n \geq n_0$, if $M = a \cdot 2^{n-t}$, where $a \in \{2^{t-1}, 2^{t-1}+1, \ldots, 2^t\}$, then for $N = 2^n$, (24) and (25) hold.*

Proof: For each $a \in \{2^{t-1}+1, 2^{t-1}+2, \ldots, 2^t\}$, Lemma 7 holds for some n_0. For the case $a = 2^{t-1}$, take $\beta'' = \frac{\beta'+\frac{1}{2}}{2}$ and use Lemma 7 with $t = 1$ to show that (24) and (25) hold with $N/2$ and β'' in place of N and β', respectively, for some n_0. Now take the largest n_0 and further require that it is large enough so that $(N/2)^{\beta''} > N^{\beta'}$. ■

Proof of Theorem 6: We focus here on the shortening case. Take $\epsilon' = \epsilon/2$, $\beta' = \beta$ and set t such that $2^{1-t} < \epsilon'$. Let n_0 be as in Corollary 8. We claim that $M_0 = 2^{n_0}$. Denote

$$\underset{\circ}{a} = \left\lfloor \frac{M}{2^{n-t}} \right\rfloor, \quad \underset{\circ}{M} = \underset{\circ}{a} \cdot 2^{n-t},$$

$$\underset{\bullet}{a} = \left\lceil \frac{M}{2^{n-t}} \right\rceil, \quad \underset{\bullet}{M} = \underset{\bullet}{a} \cdot 2^{n-t}.$$

Observe that both $\underset{\circ}{M}$ and $\underset{\bullet}{M}$ are of the form $a \cdot 2^{n-t}$ with $a \in \{2^{t-1}, 2^{t-1}+1, \ldots, 2^t\}$. These are the tightest choices of this form such that $N/2 \leq \underset{\circ}{M} \leq M \leq \underset{\bullet}{M}$. Moreover, $\underset{\bullet}{a} - \underset{\circ}{a} \leq 1$, yielding $\underset{\bullet}{M} - M \leq 2^{n-t}$ and $M - \underset{\circ}{M} \leq 2^{n-t}$.

We first prove (22). For this, we consider the random vectors $\mathbf{U}, \bar{\mathbf{U}}, \mathbf{X}, \bar{\mathbf{X}}, \mathbf{Y}, \bar{\mathbf{Y}}$ for our case of interest, i.e., shortening from length N to length M. We will also consider the corresponding vectors for the case of shortening the same N to length $\underset{\bullet}{M}$, denoted $\underset{\bullet}{\mathbf{U}}, \underset{\bullet}{\bar{\mathbf{U}}}, \underset{\bullet}{\mathbf{X}}, \underset{\bullet}{\bar{\mathbf{X}}}, \underset{\bullet}{\mathbf{Y}}, \underset{\bullet}{\bar{\mathbf{Y}}}$.

By Corollary 8, (24) holds for $\underset{\bullet}{M}$. Thus,

$$\begin{aligned}
&1 - H(X|Y) - \epsilon' \\
&\quad < \frac{1}{\underset{\bullet}{M}} \left| \left\{ 0 \leq i < \underset{\bullet}{M} : Z(\underset{\bullet}{U}_i | \underset{\bullet}{U}^{i-1}, \underset{\bullet}{\mathbf{Y}}) < 2^{-N^{\beta'}} \right\} \right| \\
&\quad \overset{(a)}{\leq} \frac{1}{\underset{\bullet}{M}} \left| \left\{ 0 \leq i < M : Z(\underset{\bullet}{U}_i | \underset{\bullet}{U}^{i-1}, \underset{\bullet}{\mathbf{Y}}) < 2^{-N^{\beta'}} \right\} \right| + \frac{\underset{\bullet}{M} - M}{\underset{\bullet}{M}} \\
&\quad \overset{(b)}{\leq} \frac{1}{\underset{\bullet}{M}} \left| \left\{ 0 \leq i < M : Z(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + \frac{\underset{\bullet}{M} - M}{\underset{\bullet}{M}} \\
&\quad \overset{(c)}{\leq} \frac{1}{M} \left| \left\{ 0 \leq i < M : Z(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + \frac{\underset{\bullet}{M} - M}{N/2} \\
&\quad \overset{(d)}{\leq} \frac{1}{M} \left| \left\{ 0 \leq i < M : Z(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + 2^{1-t} \\
&\quad \overset{(e)}{<} \frac{1}{M} \left| \left\{ 0 \leq i < M : Z(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + \epsilon'.
\end{aligned}$$

Rearranging and recalling that $\beta' = \beta$ yields (22). We now explain inequalities (a)–(e).

- (a): The index set on the right-hand-side is smaller, as $\underset{\bullet}{M} \geq M$. The contribution of the non-counted indices is at most $\underset{\bullet}{M} - M$.
- (b): We have for $0 \leq i < M$ that
$$\begin{aligned}
Z(\underset{\bullet}{U}_i | \underset{\bullet}{U}^{i-1}, \underset{\bullet}{\mathbf{Y}}) &= Z(\underset{\bullet}{\bar{U}}_i | \underset{\bullet}{\bar{U}}^{i-1}, \underset{\bullet}{\bar{\mathbf{Y}}}) \\
&\leq Z(\bar{U}_i | \bar{U}^{i-1}, \bar{\mathbf{Y}}) = Z(U_i | U^{i-1}, \mathbf{Y}),
\end{aligned}$$
where the equalities follow from (18) and the inequality follows from Lemma 2 as the joint distribution of $(\bar{U}_i; \bar{U}^{i-1}, \bar{\mathbf{Y}})$ is improved from $(\underset{\bullet}{\bar{U}}_i; \underset{\bullet}{\bar{U}}^{i-1}, \underset{\bullet}{\bar{\mathbf{Y}}})$. Indeed, this latter observation follows from Lemmas 3 and 4.
- (c): This follows from $M \leq \underset{\bullet}{M}$ and $N/2 \leq \underset{\bullet}{M}$.
- (d): This is due to $\underset{\bullet}{M} - M \leq 2^{n-t}$ and $N = 2^n$.
- (e): We defined t such that $2^{1-t} < \epsilon'$.

We now prove (23). For this, we again consider the random vectors $\mathbf{U}, \bar{\mathbf{U}}, \mathbf{X}, \bar{\mathbf{X}}, \mathbf{Y}, \bar{\mathbf{Y}}$ for our case of interest, i.e., shortening from length N to length M. We further consider the corresponding vectors for the case of shortening the same N to length $\underset{\circ}{M}$, denoted $\underset{\circ}{\mathbf{U}}, \underset{\circ}{\bar{\mathbf{U}}}, \underset{\circ}{\mathbf{X}}, \underset{\circ}{\bar{\mathbf{X}}}, \underset{\circ}{\mathbf{Y}}, \underset{\circ}{\bar{\mathbf{Y}}}$.

By Corollary 8, (25) holds for $\underset{\circ}{M}$. Thus,

$$\begin{aligned}
&H(X|Y) - \epsilon' \\
&\quad < \frac{1}{\underset{\circ}{M}} \left| \left\{ 0 \leq i < \underset{\circ}{M} : K(\underset{\circ}{U}_i | \underset{\circ}{U}^{i-1}, \underset{\circ}{\mathbf{Y}}) < 2^{-N^{\beta'}} \right\} \right| \\
&\quad \overset{(a)}{\leq} \frac{1}{\underset{\circ}{M}} \left| \left\{ 0 \leq i < \underset{\circ}{M} : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| \\
&\quad \overset{(b)}{\leq} \frac{1}{\underset{\circ}{M}} \left| \left\{ 0 \leq i < M : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| \\
&\quad = \left(\frac{1}{M} + \frac{1}{\underset{\circ}{M}} - \frac{1}{M} \right) \left| \left\{ 0 \leq i < M : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| \\
&\quad \overset{(c)}{\leq} \frac{1}{M} \left| \left\{ 0 \leq i < M : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + \frac{M - \underset{\circ}{M}}{\underset{\circ}{M}} \\
&\quad \overset{(d)}{\leq} \frac{1}{M} \left| \left\{ 0 \leq i < M : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + \frac{M - \underset{\circ}{M}}{N/2} \\
&\quad \overset{(e)}{\leq} \frac{1}{M} \left| \left\{ 0 \leq i < M : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + 2^{1-t} \\
&\quad \overset{(f)}{<} \frac{1}{M} \left| \left\{ 0 \leq i < M : K(U_i | U^{i-1}, \mathbf{Y}) < 2^{-N^{\beta'}} \right\} \right| + \epsilon'.
\end{aligned}$$

Rearranging yields (23). We now explain inequalities (a)–(f).

- (a): We have for $0 \leq i < \underset{\circ}{M}$ that
$$\begin{aligned}
K(\underset{\circ}{U}_i | \underset{\circ}{U}^{i-1}, \underset{\circ}{\mathbf{Y}}) &= K(\underset{\circ}{\bar{U}}_i | \underset{\circ}{\bar{U}}^{i-1}, \underset{\circ}{\bar{\mathbf{Y}}}) \\
&\leq K(\bar{U}_i | \bar{U}^{i-1}, \bar{\mathbf{Y}}) = K(U_i | U^{i-1}, \mathbf{Y}),
\end{aligned}$$
where the equalities follow from (19) and the inequality follows from Lemma 2 as the joint distribution of $(\bar{U}_i; \bar{U}^{i-1}, \bar{\mathbf{Y}})$ is inferior to $(\underset{\circ}{\bar{U}}_i; \underset{\circ}{\bar{U}}^{i-1}, \underset{\circ}{\bar{\mathbf{Y}}})$. Indeed, this latter observation follows from Lemmas 3 and 4.
- (b): As $M \geq \underset{\circ}{M}$, the right-hand-side is the size of a larger set than the left-hand-side.
- (c): The size of the set is at most M, and $M \cdot (1/\underset{\circ}{M} - 1/M) = (M - \underset{\circ}{M})/\underset{\circ}{M}$.
- (d): This is due to $\underset{\circ}{M} \leq N/2$.
- (d): This is due to $M - \underset{\circ}{M} \leq 2^{n-t}$ and $N = 2^n$.
- (f): We defined t such that $2^{1-t} < \epsilon'$.

This completes the proof for the shortening case.

The puncturing case uses similar mechanics. The proof of (22) for puncturing follows along the lines of the proof of (23) for shortening. The proof of (23) for puncturing follows along the lines of the proof of (22) for shortening. ■

Appendix

Proof of Lemma 3: Since $A' \sqsubseteq A$ and $B' \sqsubseteq B$, we recall (16) and denote

$$A' \overset{\mathrm{r}_1^A}{\sqsubseteq} C_1^A \overset{\mathrm{r}_2^A}{\sqsubseteq} C_2^A \overset{\mathrm{r}_3^A}{\sqsubseteq} \cdots \overset{\mathrm{r}_{t_A-1}^A}{\sqsubseteq} C_{t_A-1}^A \overset{\mathrm{r}_{t_A}^A}{\sqsubseteq} A$$

and

$$B' \overset{\mathrm{r}_1^B}{\sqsubseteq} C_1^B \overset{\mathrm{r}_2^B}{\sqsubseteq} C_2^B \overset{\mathrm{r}_3^B}{\sqsubseteq} \cdots \overset{\mathrm{r}_{t_B-1}^A}{\sqsubseteq} C_{t_B-1}^B \overset{\mathrm{r}_{t_B}^B}{\sqsubseteq} B,$$

where $\mathrm{r}_1^A, \mathrm{r}_2^A, \ldots, \mathrm{r}_{t_A}^A \in \{\mathrm{d}, \mathrm{p}\}$ and also $\mathrm{r}_1^B, \mathrm{r}_2^B, \ldots, \mathrm{r}_{t_B}^B \in \{\mathrm{d}, \mathrm{p}\}$. The proof is by induction on $t_A + t_B$.

For the base case, take $t_A + t_B = 0$. That is, $t_A = t_B = 0$, which implies that $A' = A$ and $B' = B$, and there is nothing to prove.

For the induction step, assume the claim holds when $t_A + t_B = \ell$, and consider a case where $t_A + t_B = \ell + 1$. Thus, either $t_A > 0$ or $t_B > 0$ (or both). If $t_A > 0$, it suffices to prove that

$$A' \boxplus B' \sqsubseteq C_1^A \boxplus B' \quad \text{and} \quad A' \circledast B' \sqsubseteq C_1^A \circledast B',$$

since we have by the induction hypothesis that

$$C_1^A \boxplus B' \sqsubseteq A \boxplus B \quad \text{and} \quad C_1^A \circledast B' \sqsubseteq A \circledast B,$$

and the claim follows by the transitivity of the $\sqsubseteq$ relation. Similarly, if $t_B > 0$ it suffices to prove that

$$A' \boxplus B' \sqsubseteq A' \boxplus C_1^B \quad \text{and} \quad A' \circledast B' \sqsubseteq A' \circledast C_1^B.$$

There are 8 cases to consider, since there are two options for the transform, $\boxplus$ and $\circledast$; two options for the gateway joint distribution, C_1^A and C_1^B; and two options of getting to the gateway joint distribution, $\overset{\mathrm{d}}{\sqsubseteq}$ and $\overset{\mathrm{p}}{\sqsubseteq}$. The first 4 cases will deal with C_1^A and the last 4 with C_1^B. In the interest of keeping the notation light, in the first 4 cases we rename C_1^A to A and B' to B and in the last 4 cases we rename C_1^B to B and A' to A. The cases in which we consider $\overset{\mathrm{d}}{\sqsubseteq}$ are brought here completeness, as they have already been proven in [33, Lemma 4.7].

1) We show that

$$A' \overset{\mathrm{d}}{\sqsubseteq} A \Longrightarrow A' \boxplus B \overset{\mathrm{d}}{\sqsubseteq} A \boxplus B.$$

If $A' \overset{\mathrm{d}}{\sqsubseteq} A$ then by (14), for some $Q(y_0'|y_0)$ we have

$$A'(x_0; y_0') = \sum_{y_0} A(x_0; y_0) Q(y_0'|y_0). \tag{27}$$

Thus, by (9),

$$\begin{aligned}
&(A' \boxplus B)(u_0; y_0', y_1') \\
&\quad = \sum_{x_1} A'(u_0 \oplus x_1; y_0') B(x_1; y_1') \\
&\quad = \sum_{x_1} \sum_{y_0} A(u_0 \oplus x_1; y_0) Q(y_0'|y_0) B(x_1; y_1') \\
&\quad = \sum_{y_0} \sum_{x_1} A(u_0 \oplus x_1; y_0) B(x_1; y_1') Q(y_0'|y_0) \\
&\quad = \sum_{y_0} (A \boxplus B)(u_0; y_0, y_1') Q(y_0'|y_0).
\end{aligned}$$

We now define

$$Q'(y_0', y_1'|y_0, y_1) = \begin{cases} Q(y_0'|y_0), & y_1' = y_1 \\ 0, & \text{otherwise,} \end{cases}$$

and continue the above derivation as

$$\begin{aligned}
&\sum_{y_0} (A \boxplus B)(u_0; y_0, y_1') Q(y_0'|y_0) \\
&\qquad = \sum_{y_0, y_1} (A \boxplus B)(u_0; y_0, y_1) Q'(y_0', y_1'|y_0, y_1).
\end{aligned}$$

The claim follows by (14).

2) We show that

$$A' \overset{\mathrm{d}}{\sqsubseteq} A \Longrightarrow A' \circledast B \overset{\mathrm{d}}{\sqsubseteq} A \circledast B.$$

As in the previous case, there exists $Q(y_0'|y_0)$ such that (27) holds. Thus, by (10),

$$\begin{aligned}
&(A' \circledast B)(u_1; u_0', y_0', y_1') \\
&\quad = A'(u_0' \oplus u_1; y_0') B(u_1; y_1') \\
&\quad = \sum_{y_0} A(u_0' \oplus u_1; y_0) Q(y_0'|y_0) B(u_1; y_1') \\
&\quad = \sum_{y_0} A(u_0' \oplus u_1; y_0) B(u_1; y_1') Q(y_0'|y_0) \\
&\quad = \sum_{y_0} (A \circledast B)(u_1; u_0', y_0, y_1') Q(y_0'|y_0).
\end{aligned}$$

We now define

$$Q'(u_0', y_0', y_1'|u_0, y_0, y_1) = \begin{cases} Q(y_0'|y_0), & y_1' = y_1, u_0' = u_0 \\ 0, & \text{otherwise,} \end{cases}$$

and continue the above derivation as

$$\begin{aligned}
&\sum_{y_0} (A \circledast B)(u_1; u_0', y_0, y_1') Q(y_0'|y_0) \\
&\quad = \sum_{u_0, y_0, y_1} (A \circledast B)(u_1; u_0, y_0, y_1) Q'(u_0', y_0', y_1'|u_0, y_0, y_1).
\end{aligned}$$

The claim follows by (14).

3) We show that

$$A' \overset{\mathrm{p}}{\sqsubseteq} A \Longrightarrow A' \boxplus B \overset{\mathrm{p}}{\sqsubseteq} A \boxplus B.$$

If $A' \overset{\mathrm{p}}{\sqsubseteq} A$ then by (15), for some $f(y_0)$ we have

$$A'(x_0; y_0) = A(x_0 \oplus f(y_0); y_0). \tag{28}$$

Thus, by (9),

$$\begin{aligned}
&(A' \boxplus B)(u_0; y_0, y_1) \\
&\quad = \sum_{x_1} A'(u_0 \oplus x_1; y_0) B(x_1; y_1) \\
&\quad = \sum_{x_1} A(u_0 \oplus x_1 \oplus f(y_0); y_0) B(x_1; y_1) \\
&\quad = (A \boxplus B)(u_0 \oplus f(y_0); y_0, y_1).
\end{aligned}$$

We now define

$$g(y_0, y_1) = f(y_0)$$

and continue the above derivation as

$$\begin{aligned}
&(A \boxplus B)(u_0 \oplus f(y_0); y_0, y_1) \\
&\qquad = (A \boxplus B)(u_0 \oplus g(y_0, y_1); y_0, y_1).
\end{aligned}$$

The claim follows by (15).

4) We show that

$$A' \overset{\mathrm{p}}{\sqsubseteq} A \Longrightarrow A' \circledast B \sqsubseteq A \circledast B.$$

Specifically, we show that

$$A' \overset{\mathrm{p}}{\sqsubseteq} A \Longrightarrow A' \circledast B \overset{\mathrm{d}}{\sqsubseteq} A \circledast B.$$

As in the previous case, there exists $f(y_0)$ such that (28) holds. Thus, by (10),

$$\begin{aligned}
&(A' \circledast B)(u_1; u_0', y_0', y_1')\\
&\quad = A'(u_0' \oplus u_1; y_0')B(u_1; y_1')\\
&\quad = A(u_0' \oplus u_1 \oplus f(y_0'); y_0')B(u_1; y_1')\\
&\quad = (A \circledast B)(u_1; u_0' \oplus f(y_0'), y_0', y_1').
\end{aligned}$$

We now define

$$\begin{aligned}
&Q'(u_0', y_0', y_1' | u_0, y_0, y_1)\\
&\qquad = \begin{cases} 1, & (u_0' \oplus f(y_0'), y_0', y_1') = (u_0, y_0, y_1)\\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

and continue the above derivation as

$$\begin{aligned}
&(A \circledast B)(u_1; u_0' \oplus f(y_0'), y_0, y_1')\\
&\quad = \sum_{u_0, y_0, y_1} (A \circledast B)(u_1; u_0, y_0, y_1)Q'(u_0', y_0', y_1' | u_0, y_0, y_1).
\end{aligned}$$

The claim follows by (14).

5) We show that

$$B' \overset{\mathrm{d}}{\sqsubseteq} B \Longrightarrow A \boxast B' \overset{\mathrm{d}}{\sqsubseteq} A \boxast B.$$

If $B' \overset{\mathrm{d}}{\sqsubseteq} B$ then by (14), for some $Q(y_1'|y_1)$ we have

$$B'(x_1; y_1') = \sum_{y_1} B(x_1; y_1)Q(y_1'|y_1). \tag{29}$$

Thus, by (9),

$$\begin{aligned}
&(A \boxast B')(u_0; y_0', y_1')\\
&\quad = \sum_{x_1} A(u_0 \oplus x_1; y_0')B'(x_1; y_1')\\
&\quad = \sum_{x_1} A(u_0 \oplus x_1; y_0') \sum_{y_1} B(x_1; y_1)Q(y_1'|y_1)\\
&\quad = \sum_{y_1} \sum_{x_1} A(u_0 \oplus x_1; y_0')B(x_1; y_1)Q(y_1'|y_1)\\
&\quad = \sum_{y_1} (A \boxast B)(u_0; y_0', y_1)Q(y_1'|y_1).
\end{aligned}$$

We now define

$$Q'(y_0', y_1' | y_0, y_1) = \begin{cases} Q(y_1'|y_1), & y_0' = y_0\\ 0, & \text{otherwise,} \end{cases}$$

and continue the above derivation as

$$\begin{aligned}
&\sum_{y_1} (A \boxast B)(u_0; y_0', y_1)Q(y_1'|y_1)\\
&\quad = \sum_{y_0, y_1} (A \boxast B)(u_0; y_0, y_1)Q'(y_0', y_1' | y_0, y_1).
\end{aligned}$$

The claim follows by (14).

6) We show that

$$B' \overset{\mathrm{d}}{\sqsubseteq} B \Longrightarrow A \circledast B' \overset{\mathrm{d}}{\sqsubseteq} A \circledast B.$$

As in the previous case, there exists $Q(y_1'|y_1)$ such that (29) holds. Thus, by (10),

$$\begin{aligned}
&(A \circledast B')(u_1; u_0', y_0', y_1')\\
&\quad = A(u_0' \oplus u_1; y_0')B'(u_1; y_1')\\
&\quad = A(u_0' \oplus u_1; y_0') \sum_{y_1} B(u_1; y_1)Q(y_1'|y_1)\\
&\quad = \sum_{y_1} A(u_0' \oplus u_1; y_0')B(u_1; y_1)Q(y_1'|y_1)\\
&\quad = \sum_{y_1} (A \circledast B)(u_1; u_0', y_0', y_1)Q(y_1'|y_1).
\end{aligned}$$

We now define

$$Q'(u_0', y_0', y_1' | u_0, y_0, y_1) = \begin{cases} Q(y_1'|y_1), & y_0' = y_0, u_0' = u_0\\ 0, & \text{otherwise,} \end{cases}$$

and continue the above derivation as

$$\begin{aligned}
&\sum_{y_1} (A \circledast B)(u_1; u_0', y_0', y_1)Q(y_1'|y_1)\\
&\quad = \sum_{u_0, y_0, y_1} (A \circledast B)(u_1; u_0, y_0, y_1)Q'(u_0', y_0', y_1' | u_0, y_0, y_1).
\end{aligned}$$

The claim follows by (14).

7) We show that

$$B' \overset{\mathrm{p}}{\sqsubseteq} B \Longrightarrow A \boxast B' \overset{\mathrm{p}}{\sqsubseteq} A \boxast B.$$

If $B' \overset{\mathrm{p}}{\sqsubseteq} B$ then by (15), for some $f(y_1)$ we have

$$B'(x_1; y_1) = B(x_1 \oplus f(y_1); y_1). \tag{30}$$

Thus, by (9),

$$\begin{aligned}
&(A \boxast B')(u_0; y_0, y_1)\\
&\quad = \sum_{x_1} A(u_0 \oplus x_1; y_0)B'(x_1; y_1)\\
&\quad = \sum_{x_1} A(u_0 \oplus x_1; y_0)B(x_1 \oplus f(y_1); y_1)\\
&\quad \overset{(a)}{=} \sum_{x_1} A(u_0 \oplus x_1 \oplus f(y_1); y_0)B(x_1; y_1)\\
&\quad = (A \boxast B)(u_0 \oplus f(y_1); y_0, y_1).
\end{aligned}$$

Note that (a) holds both when $f(y_1) = 0$ and $f(y_1) = 1$. In the former this is trivial and in the latter we're simply changing the order of summation. We now define

$$g(y_0, y_1) = f(y_1)$$

and continue the above derivation as

$$\begin{aligned}
&(A \boxast B)(u_0 \oplus f(y_1); y_0, y_1)\\
&\quad = (A \boxast B)(u_0 \oplus g(y_0, y_1); y_0, y_1).
\end{aligned}$$

The claim follows by (15).

8) We show that

$$B' \overset{\mathrm{p}}{\sqsubseteq} B \Longrightarrow A \circledast B' \sqsubseteq A \circledast B.$$

Specifically, we show that

$$B' \overset{\mathrm{p}}{\sqsubseteq} B \Longrightarrow A \circledast B' \overset{\mathrm{d}}{\sqsubseteq} C \overset{\mathrm{p}}{\sqsubseteq} A \circledast B$$

for a joint distribution C we will shortly define. As in the previous case, there exists $f(y_1)$ such that (30) holds. Further, let $f'(u_0, y_0, y_1) = f(y_1)$ and define

$$\begin{aligned}
C(u_1; u_0, y_0, y_1) &= (A \circledast B)(u_1 \oplus f'(u_0, y_0, y_1); u_0, y_0, y_1)\\
&= (A \circledast B)(u_1 \oplus f(y_1); u_0, y_0, y_1).
\end{aligned}$$

Clearly, by (15), $C \overset{\mathrm{p}}{\sqsubseteq} A \circledast B$. Next, by (10),

$$\begin{aligned}&(A \circledast B')(u_1; u_0', y_0', y_1')\\&\quad= A(u_0' \oplus u_1; y_0')B'(u_1; y_1')\\&\quad= A(u_0' \oplus u_1; y_0')B(u_1 \oplus f(y_1'); y_1')\\&\quad= A(u_0' \oplus f(y_1') \oplus u_1 \oplus f(y_1'); y_0')B(u_1 \oplus f(y_1'); y_1')\\&\quad= (A \circledast B)(u_1 \oplus f(y_1'); u_0' \oplus f(y_1'), y_0', y_1')\\&\quad= C(u_1; u_0' \oplus f(y_1'), y_0', y_1').\end{aligned}$$

We now define

$$\begin{aligned}&Q'(u_0', y_0', y_1' | u_0, y_0, y_1)\\&\qquad\qquad= \begin{cases}1, & (u_0' \oplus f(y_1'), y_0', y_1') = (u_0, y_0, y_1)\\0, & \text{otherwise,}\end{cases}\end{aligned}$$

and continue the above derivation as

$$\begin{aligned}&C(u_1; u_0' \oplus f(y_1'), y_0', y_1')\\&\quad= \sum_{u_0, y_0, y_1} C(u_1; u_0, y_0, y_1)Q'(u_0', y_0', y_1' | u_0, y_0, y_1).\end{aligned}$$

The claim follows by (14). ■

Proof of Lemma 4: We first show that $\mathtt{P} \sqsubseteq A$. That is, we show that

$$\mathtt{P} \overset{\mathrm{d}}{\sqsubseteq} C_1 \overset{\mathrm{p}}{\sqsubseteq} C_2 \overset{\mathrm{d}}{\sqsubseteq} A.$$

To this end, let $Q(y_0, x_1 | y_0') = 1/2$ if $y_0' = y_0$ and 0 otherwise. Then, by (14), indeed $C_2 \overset{\mathrm{d}}{\sqsubseteq} A$, with

$$C_2(x_0; y_0, x_1) = \sum_{y_0' \in \mathcal{Y}_0} A(x_0; y_0')Q(y_0, x_1 | y_0').$$

Next, define $f(y_0, x_1) = x_1$. Then, by (15), $C_1 \overset{\mathrm{p}}{\sqsubseteq} C_2$ with

$$C_1(x_0; y_0, x_1) = C_2(x_0 \oplus f(y_0, x_1); y_0, x_1).$$

Observe by marginalization that $C_1(x_0) = 1/2$ for $x_0 \in \mathcal{X}$. Finally, again by (14), $\mathtt{P} \overset{\mathrm{d}}{\sqsubseteq} C_1$ with

$$\mathtt{P}(x_0; y_0') = \sum_{y_0, x_1} C_1(x_0; y_0, x_1)Q'(y_0' | y_0, x_1),$$

where $Q'(y_0' | y_0, x_1) = 1$ whenever $y_0' =?$ and 0 otherwise.

Next, we show that $A \sqsubseteq \mathtt{S}$. That is, we show that

$$A \overset{\mathrm{d}}{\sqsubseteq} D_1 \overset{\mathrm{p}}{\sqsubseteq} D_2 \overset{\mathrm{d}}{\sqsubseteq} \mathtt{S}.$$

First, let $R(y_0, x_0 | y) = A(x_0; y_0)$. Then, by (14), $D_2 \overset{\mathrm{d}}{\sqsubseteq} \mathtt{S}$ with

$$D_2(x; x_0, y_0) = \sum_{y} \mathtt{S}(x; y)R(y_0, x_0 | y).$$

Observe that $D_2(x; x_0, y_0) = A(x_0; y_0)$ when $x = 0$ and 0 otherwise. Next, define $g(x_0, y_0) = x_0$. By (15), $D_1 \overset{\mathrm{p}}{\sqsubseteq} D_2$ with

$$D_1(x'; x_0, y_0) = D_2(x' \oplus g(x_0, y_0); x_0, y_0).$$

Observe that

$$D_1(x'; x_0, y_0) = \begin{cases}A(x_0; y_0), & x' = x_0,\\0, & \text{otherwise.}\end{cases}$$

For the final step, take $R'(y' | x_0, y_0) = 1$ if $y' = y_0$ and 0 otherwise. Observe that

$$A(x'; y') = \sum_{x_0, y_0} D_1(x'; x_0, y_0)R'(y' | x_0, y_0).$$

Indeed, the sum is nonzero only when $x' = x_0$ and $y' = y_0$, in which case it equals $A(x'; y')$. Thus, by (14), $A \overset{\mathrm{d}}{\sqsubseteq} D_1$. ■

Proof of Lemma 5: Throughout the proof we will use (11) and (12).

Step 1: $A \boxast \mathtt{S} \equiv A$. First, note by (9) that $(A \boxast \mathtt{S})(u_0; y_0, ?) = A(u_0; y_0)$. To see the equivalence, we show that $A \sqsubseteq A \boxast \mathtt{S} \sqsubseteq A$. Indeed, $A \boxast \mathtt{S} \overset{\mathrm{d}}{\sqsubseteq} A$, since in (14) we can take

$$Q(y_0, ? | y') = \begin{cases}1, & y' = y_0,\\0, & \text{otherwise,}\end{cases}$$

and $A \overset{\mathrm{d}}{\sqsubseteq} A \boxast \mathtt{S}$, by (14) with

$$Q(y' | y_0, ?) = \begin{cases}1, & y' = y_0,\\0, & \text{otherwise.}\end{cases}$$

Step 2: $A \boxast \mathtt{P} \equiv \mathtt{P}$. By Lemma 4, $\mathtt{P} \sqsubseteq A \boxast \mathtt{P}$. It remains to show $A \boxast \mathtt{P} \sqsubseteq \mathtt{P}$. By (9), $(A \boxast \mathtt{P})(u_0; y_0, ?) = A(y_0)/2$, where we denote $A(y_0) = A(0; y_0) + A(1; y_0)$. Next, $A \boxast \mathtt{P} \overset{\mathrm{d}}{\sqsubseteq} \mathtt{P}$ by (14) with $Q(y_0, ? | ?) = A(y_0)$.

Step 3: $\mathtt{S} \boxast B \equiv B$. By (9), we have $(\mathtt{S} \boxast B)(u_0; ?, y_1) = B(u_0; y_1)$. Next, $\mathtt{S} \boxast B \equiv B$ by the same arguments as in Step 1.

Step 4: $\mathtt{P} \boxast B \equiv \mathtt{P}$. By (9), we have $(\mathtt{P} \boxast B)(u_0; ?, y_1) = B(y_1)/2$, where $B(y_1) = B(0; y_1) + B(y_1)$. Next, $\mathtt{P} \boxast B \equiv \mathtt{P}$ by the same arguments as in Step 2.

Step 5: $\mathtt{S} \boxast \mathtt{S} \equiv \mathtt{S}$*,* $\mathtt{S} \boxast \mathtt{P} \equiv \mathtt{P}$*,* $\mathtt{P} \boxast \mathtt{S} \equiv \mathtt{P}$*, and* $\mathtt{P} \boxast \mathtt{P} \equiv \mathtt{P}$. These are special cases of Steps 1–4.

Step 6: $A \circledast \mathtt{S} \equiv \mathtt{S}$. By Lemma 4, $A \circledast \mathtt{S} \sqsubseteq \mathtt{S}$. It remains to show that $\mathtt{S} \sqsubseteq A \circledast \mathtt{S}$. By (10),

$$(A \circledast \mathtt{S})(u_1; u_0, y_0, ?) = \begin{cases}A(u_0; y_0), & u_1 = 0,\\0, & \text{otherwise.}\end{cases}$$

Next, $\mathtt{S} \overset{\mathrm{d}}{\sqsubseteq} A \circledast \mathtt{S}$ by (14) with $Q(? | u_0, y_0, ?) = 1$.

Step 7: $A \circledast \mathtt{P} \equiv A$. First, note by (10) that

$$(A \circledast \mathtt{P})(u_1; u_0, y_0, ?) = \frac{1}{2}A(u_0 \oplus u_1; y_0). \tag{31}$$

To see the equivalence, we show that $A \sqsubseteq A \circledast \mathtt{P} \sqsubseteq A$.

To show that $A \circledast \mathtt{P} \sqsubseteq A$ we show that

$$A \circledast \mathtt{P} \overset{\mathrm{p}}{\sqsubseteq} C_1 \overset{\mathrm{d}}{\sqsubseteq} A.$$

To this end, for $u_0 \in \mathcal{X}$, let $Q(u_0, y_0, ? | y_1) = 1/2$ if $y_1 = y_0$ and 0 otherwise. Then, by (14), indeed $C_1 \overset{\mathrm{d}}{\sqsubseteq} A$, with

$$C_1(u_1; u_0, y_0, ?) = \sum_{y_1} A(u_1; y_1)Q(u_0, y_0, ? | y_1) = \frac{1}{2}A(u_1; y_0).$$

Next, define $f(u_0, y_0, ?) = u_0$. Then, by (15) and (31), $A \circledast \mathtt{P} \overset{\mathrm{p}}{\sqsubseteq} C_1$ as

$$\begin{aligned}C_1(u_1 \oplus f(u_0, y_0, ?); u_0, y_0, ?) &= \frac{1}{2}A(u_1 \oplus u_0; y_0)\\&= \frac{1}{2}A(u_0 \oplus u_1; y_0)\\&= (A \circledast \mathtt{P})(u_1; u_0, y_0, ?).\end{aligned}$$

To show that $A \sqsubseteq A \circledast \mathtt{P}$ we show that

$$A \overset{\mathrm{d}}{\sqsubseteq} C_2 \overset{\mathrm{p}}{\sqsubseteq} A \circledast \mathtt{P}.$$

First, define $g(u_0, y_0, ?) = u_0$. Then, by (15) and (31),

$$\begin{aligned} C_2(u_1; u_0, y_0, ?) &= (A \circledast \mathtt{P})(u_1 \oplus g(u_0, y_0, ?); u_0, y_0, ?) \\ &= \frac{1}{2} A(u_1; y_0). \end{aligned}$$

Let $Q(y_1|u_0, y_0, ?) = 1$ if $y_1 = y_0$ and 0 otherwise. Then, by (14), indeed $A \overset{\mathrm{d}}{\sqsubseteq} C_2$, as

$$\begin{aligned} \sum_{u_0, y_0} C_2(u_1; u_0, y_0, ?) Q(y_1|u_0, y_0, ?) &= \sum_{u_0} \frac{1}{2} A(u_1; y_1) \\ &= A(u_1; y_1). \end{aligned}$$

Step 8: $\mathtt{S} \circledast B \equiv \mathtt{S}$. By Lemma 4, $\mathtt{S} \circledast B \sqsubseteq \mathtt{S}$. It remains to show $\mathtt{S} \sqsubseteq \mathtt{S} \circledast B$. By (10), we have

$$\begin{aligned} (\mathtt{S} \circledast B)(u_1; u_0, ?, y_1) &= \mathtt{S}(u_0 \oplus u_1; ?) B(u_1; y_1) \\ &= \begin{cases} B(u_1; y_1), & u_1 = u_0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

We now show that $\mathtt{S} \overset{\mathrm{d}}{\sqsubseteq} C_3 \overset{\mathrm{p}}{\sqsubseteq} \mathtt{S} \circledast B$. Let $h(u_0, ?, y_1) = u_0$. Then, by (15), $C_3 \overset{\mathrm{p}}{\sqsubseteq} \mathtt{S} \circledast B$ with

$$\begin{aligned} C_3(u_1; u_0, ?, y_1) &= (\mathtt{S} \circledast B)(u_1 \oplus h(u_0, ?, y_1); u_0, ?, y_1) \\ &= (\mathtt{S} \circledast B)(u_1 \oplus u_0; u_0, ?, y_1) \\ &= \begin{cases} B(u_1 \oplus u_0; y_1), & u_1 \oplus u_0 = u_0, \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} B(u_1 \oplus u_0; y_1), & u_1 = 0, \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} B(u_0; y_1), & u_1 = 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Next, by (14), $\mathtt{S} \overset{\mathrm{d}}{\sqsubseteq} C_3$ with $Q(?|u_0, ?, y_1) = 1$.

Step 9: $\mathtt{P} \circledast B \equiv B$. By (10), we have $(\mathtt{P} \circledast B)(u_1; u_0, ?, y_1) = B(u_1; y_1)/2$. We now show that $B \sqsubseteq \mathtt{P} \circledast B \sqsubseteq B$, which will prove the equivalence. Indeed, $\mathtt{P} \circledast B \overset{\mathrm{d}}{\sqsubseteq} B$, since in (14) we can take

$$Q(u_0, ?, y_1|y') = \begin{cases} \frac{1}{2}, & y' = y_1, \\ 0, & \text{otherwise,} \end{cases}$$

and $B \overset{\mathrm{d}}{\sqsubseteq} \mathtt{P} \circledast B$, by (14) with

$$Q(y'|u_0, ?, y_1) = \begin{cases} 1, & y' = y_1, \\ 0, & \text{otherwise.} \end{cases}$$

Step 10: $\mathtt{S} \circledast \mathtt{S} \equiv \mathtt{S}$*,* $\mathtt{S} \circledast \mathtt{P} \equiv \mathtt{S}$*,* $\mathtt{P} \circledast \mathtt{S} \equiv \mathtt{S}$*, and* $\mathtt{P} \circledast \mathtt{P} \equiv \mathtt{P}$. These are special cases of Steps 6–9. ■

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] J. Honda and H. Yamamoto, “Polar coding without alphabet extension for asymmetric channels,” *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7829–7838, December 2012.

[3] E. Şaşoğlu, “Polar codes for discrete alphabets,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2012)*, Cambridge, Massachusetts, 2012, pp. 2137–21 141.

[4] E. Şaşoğlu and I. Tal, “Polar coding for processes with memory,” *IEEE Trans. Inform. Theory*, vol. 65, no. 4, pp. 1994–2003, April 2019.

[5] B. Shuval and I. Tal, “Fast polarization for processes with memory,” *IEEE Trans. Inform. Theory*, vol. 65, no. 4, pp. 2004–2020, April 2019.

[6] I. Tal, H. D. Pfister, A. Fazeli, and A. Vardy, “Polar codes for the deletion channel: weak and strong polarization,” *IEEE Trans. Inform. Theory*, vol. 68, no. 4, pp. 2239–2265, April 2022.

[7] H. D. Pfister and I. Tal, “Polar codes for channels with insertions, deletions, and substitutions,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2021)*, Melbourne, Victoria, Australia, 2021, pp. 2554–2559.

[8] E. Hof, I. Sason, S. S. (Shitz), and C. Tian, “Capacity-achieving polar codes for arbitrarily permuted parallel channels,” *IEEE Trans. Inform. Theory*, vol. 59, March 2013.

[9] E. Hof and S. Shamai, “Secrecy-achieving polar-coding for binary-input memoryless symmetric wire-tap channels,” `arXiv:1005.2759v2`, 2010.

[10] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, “Nested polar codes for wiretap and relay channels,” *IEEE Commmun. Lett.*, vol. 14, pp. 752–754, 2010.

[11] H. Mahdavifar and A. Vardy, “Achieving the secrecy capacity of wiretap channels using polar codes,” *IEEE Trans. Inform. Theory*, vol. 57, pp. 6428–6443, 2011.

[12] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, “Achieving marton’s region for broadcast channels using polar codes,” *IEEE Trans. Inform. Theory*, vol. 61, pp. 783–800, 2015.

[13] M. Mondelli, S. H. Hassani, and R. Urbanke, “How to achieve the capacity of asymmetric channels,” *IEEE Trans. Inform. Theory*, vol. 64, no. 5, pp. 3371–3393, May 2018.

[14] K. Tian, A. Fazeli, and A. Vardy, “Polar coding for channels with deletions,” *IEEE Trans. Inform. Theory*, vol. 67, no. 11, pp. 7081–7095, November 2021.

[15] E. Şaşoğlu and L. Wang, “Universal polarization,” *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 2937–2946, June 2016.

[16] B. Shuval and I. Tal, “Universal polarization for processes with memory,” `arXiv:1811.05727v1`, 2018.

[17] E. Abbe and E. Telatar, “Polar codes for the m-user multiple access channel,” *IEEE Trans. Inform. Theory*, vol. 58, pp. 5437–5448, 2012.

[18] E. Şaşoğlu, E. Telatar, and E. Yeh, “Polar codes for the two-user multiple-access channel,” *IEEE Trans. Inform. Theory*, vol. 59, pp. 6583–6592, 2013.

[19] N. Goela, E. Abbe, and M. Gastpar, “Polar codes for broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 61, pp. 758–782, 2015.

[20] H. Mahdavifar, “Polar coding for non-stationary channels,” *IEEE Trans. Inform. Theory*, vol. 66, pp. 6920–6938, 2020.

[21] D. Arava and I. Tal, “Stronger polarization in the deletion channel,” in *Proc. IEEE Int’l Symp. Inform. Theory (ISIT’2023)*, Taipei, Taiwan, 2023, pp. 1711–1716.

[22] R. M. Roth, *Introduction to Coding Theory*. Cambridge, UK: Cambridge University Press, 2006.

[23] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam: North-Holland, 1977.

[24] K. Niu, K. Chen, and J.-R. Lin, “Beyond turbo codes: Rate-compatible punctured polar codes,” in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 3423–3427.

[25] R. Wang and R. Liu, “A novel puncturing scheme for polar codes,” *IEEE Communications Letters*, vol. 18, no. 12, pp. 2081–2084, 2014.

[26] V. Bioglio, F. Gabry, and I. Land, “Low-complexity puncturing and shortening of polar codes,” in *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2017, pp. 1–6.

[27] R. M. Oliveira and R. C. de Lamare, “Rate-compatible polar codes based on polarization-driven shortening,” *IEEE Communications Letters*, vol. 22, no. 10, pp. 1984–1987, 2018.

[28] ——, “Puncturing based on polarization for polar codes in 5g networks,” in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–5.

[29] T. Tonnellier, A. Cavatassi, and W. J. Gross, “Length-compatible polar codes: A survey : (invited paper),” in *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, 2019, pp. 1–6.

[30] X. Yao and X. Ma, “A balanced tree approach to construction of length-flexible polar codes,” *IEEE Trans. Commun. Early Access*, 2023.

[31] 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.

[32] I. Tal, “A simple proof of fast polarization,” *IEEE Trans. Inform. Theory*, vol. 63, no. 12, pp. 7617–7619, December 2017.

[33] S. B. Korada, “Polar codes for channel and source coding,” Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2009.