Capacity-Achieving Polar Codes for Arbitrarily-Permuted Parallel Channels

Eran Hof*   Igal Sason*   Shlomo Shamai*   Chao Tian[†]

* Department of Electrical Engineering

Technion – Israel Institute of Technology

Haifa 32000, Israel

E-mails: eran.hof@gmail.com, {sason@ee, sshlomo@ee}.technion.ac.il

[†]AT&T Labs-Research

180 Park Ave.

Florham Park, NJ 07932

Email: tian@research.att.com

Abstract

Channel coding over arbitrarily-permuted parallel channels was first studied by Willems et al. (2008). This paper introduces capacity-achieving polar coding schemes for arbitrarily-permuted parallel channels where the component channels are memoryless, binary-input and output-symmetric.

I. INTRODUCTION

Parallel channels are used to serve as a model for a time-varying communication channel. In this model, each one of the parallel channels corresponds to a possible state of the time-varying channel, and the communication takes place over one of these parallel channels according to the instantaneous state of the time-varying channel. The model of *arbitrarily-permuted parallel channels* was introduced in [1] where each message is encoded into a number (say S) of code-sequences with a common block length, each one of the S code-sequences is transmitted over a different parallel channel where the assignment of codewords to channels is known to the receiver, and it is modeled by an arbitrary permutation π of the set $\{1, \ldots, S\}$ where code-sequence no. $s \in \{1, \ldots, S\}$ is transmitted over the parallel channel no. $r = \pi(s)$. Finally, the receiver estimates the transmitted message based on the knowledge of this permutation and the received outputs from the S parallel channels. This model of parallel channels can be viewed as a special case of the classical compound channel setting [2].

Channel coding over arbitrarily-permuted parallel channels was studied in [1] and more recently in [3], where it was assumed that all these parallel channels have an identical input alphabet. In the case where all the parallel channels have the same capacity-achieving input distribution, it was proved in [1, Theorem 1] that the capacity of the system is equal to the sum of the capacities of the parallel channels. Furthermore, [1] also addresses the case where the parallel channels have different capacity-achieving input distributions, and it determines the capacity of the system also in this case (see [1, Theorem 2]).

Arbitrarily-permuted parallel channels may be of interest when analyzing, e.g., networking applications, OFDM and BICM systems. For example, the channel frequency bands or the bits may not be allocated at the transmitter level, and though this allocation is fixed, it takes the form of a random permutation that is selected once per transmission. In the setting of transmission of data through packets, these packets can be viewed as being transmitted over a set of parallel channels where each packet goes through one of the available parallel channels depending on the higher level of the communication protocol. The transmission in this case is done in an interleaved manner where consecutive bits are separated to different packets, the number of which is the cardinality S of the set of parallel channels. This, again, provides the model of arbitrarily-permuted parallel channels, though we do not deal in this work with data flow issues (assuming that the system is at equilibrium as far as the data/ packet rate is considered). It is also noted that these channels may be actually serial in time where a time frame of S consecutive symbols is interpreted as the time frame of a super-symbol. The mix in this case may result due to the random availability of the channels, which stays fixed for the whole codeword transmission.

The coding schemes suggested in [1] are based on random coding and decoding by joint typicality. One of the main contributions of [1] is the introduction of a concatenation of rate-matching codes with parallel copies of a fully random block code. A rate-matching code is a device that encodes a single message into a set of codewords, and it creates the required dependence between the codewords for the parallel channels. It was shown in [1] that under specific structural conditions on the rate-matching code, a sequential decoding procedure can achieve the capacity of the considered channel model. Moreover, it was shown that such rate-matching codes can be constructed from a set of maximum-distance separable (MDS) codes. In [3], space-time modulation was considered for the particular case of arbitrarily-permuted parallel Gaussian channels.

In this work, we consider the construction of polar codes as channel codes for arbitrarily-permuted parallel channels. Polar codes were recently proposed in [4], where it was demonstrated that this class of codes can achieve the capacity of a symmetric DMC with low encoding and decoding complexity. We propose two polar coding schemes in this work, and show that they achieve the capacity of arbitrarily-permuted parallel channels where each of these components is assumed to be a memoryless, binary-input and output-symmetric channel. Two simplifications of these schemes are also discussed in two special cases. The first simplification addresses the case where the communication is over two or three parallel channels, and the second simplification refers to the case of communication over parallel (stochastically) degraded channels.

The polar code framework is shown to suit well as a coding technique in the setting of arbitrarily-permuted parallel channels. The construction of the rate-matching codes in [1, Section 6] via the use of MDS codes suggests that they can also play an instrumental role when polar codes are used as channel codes for the considered setting of parallel channels. However, in order to use polar codes in the parallel channel setting, the concept of the fixed bits in the original polar codes [4] need to be slightly generalized. In [4], the values of these fixed bits can be chosen arbitrarily, independently of the transmitted message. In the proposed schemes for the arbitrarily-permuted parallel channels, some of the concerned bits need to incorporate an algebraic structure of the MDS codes, and they actually depend on the transmitted message in a manner similar to the rate-matching code in [1]. Another unique feature of the proposed scheme is that the successive cancellation techniques are applied in a parallel fashion on the channels.

This rest of the paper is structured as follows. Section II provides some preliminary material. The proposed parallel polar coding schemes are introduced and analyzed in Section III with some technicalities that are relegated to the appendix. Finally, Section IV concludes this work.
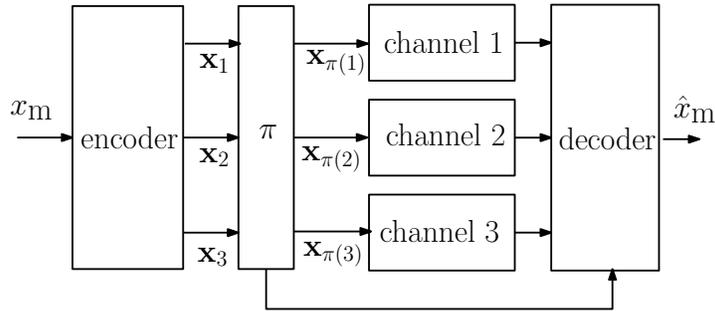
Fig. 1: Communication over an arbitrarily-permuted parallel channel with $S = 3$ in this example (taken from [1]).

II. PRELIMINARIES

A. Arbitrarily Permuted Parallel Channels

Consider the communication model depicted in Figure 1. A message x_{m} is transmitted over a set of S parallel memoryless channels. The notation $[S] \triangleq \{1, \ldots, S\}$ is used in this paper. All channels are assumed to have a common input alphabet $\mathcal{X}$, and possibly different output alphabets $\mathcal{Y}_s$, $s \in [S]$. The transition probability function of each channel is denoted by $P_s(y_s|x)$, where $y_s \in \mathcal{Y}_s$, $s \in [S]$, and $x \in \mathcal{X}$. The encoding operation maps the message x_{m} into a set of S codewords $\{\mathbf{x}_s \in \mathcal{X}^n\}_{s=1}^{S}$. Each of these codewords is of length n, and it is transmitted over a different channel. The assignment of codewords to channels is done by an arbitrary permutation $\pi : [S] \to [S]$ (note that π is fixed during the entire block transmission). The permutation π is a part of the communication channel model, the encoder has no control or information on the arbitrary permutation chosen during the codeword transmission. The set of possible S channels are known at both the encoder and decoder. In addition, the decoder knows the specific chosen permutation. Formally, the channel is defined by the following family of transition probabilities:

$$\left\{ P\big(\mathbf{Y}|\mathbf{X}; \pi\big) : \ \mathbf{Y} \in \{\mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_S\}^n, \ \mathbf{X} \in \mathcal{X}^{s \times n}, \ \pi : [S] \to [S] \right\}_{n=1}^{\infty}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_S)$ are the transmitted codewords, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_S)$ are the received vectors,

$$P\big(\mathbf{Y}|\mathbf{X}; \pi\big) = \prod_{s=1}^{S} P_s\big(\mathbf{y}_s|\mathbf{x}_{\pi(s)}\big) \tag{1}$$

is the probability law of the parallel channels, and $\pi : [S] \to [S]$ is the arbitrary permutation mapping of codewords to channels. The decoder produces the estimated message $\hat{x}_{\mathrm{m}}$ based on the received vectors $\mathbf{Y}$ and the permutation π. The case where the decoded message is different from the transmitted message, $\hat{x}_{\mathrm{m}} \neq x_{\mathrm{m}}$, is a block error event.

Definition 1 (Achievable rates and channel capacity). A rate $R > 0$ is achievable for communication over a set of S arbitrarily-permuted parallel channels if there exists a sequence of encoders and decoders such that for any $\delta > 0$ and sufficiently large block length n

$$\frac{1}{n} \log_2 M \geq R - \delta \tag{2}$$

$$P_{\mathrm{e}}^{(\pi)}(n) \leq \delta, \ \text{for all } S! \text{ permutations } \pi : [S] \to [S] \tag{3}$$

where M is the number of possible messages and $P_{\mathrm{e}}^{(\pi)}(n)$ is the average block error probability for a fixed permutation π and block length n. The capacity C_{Π} is the maximum of such achievable rates.

The capacity C_{Π} of this channel model can be derived as a particular case of the compound channel (see, e.g.,

[2] and reference therein). Specifically, if there exists an input distribution that achieves capacity for all the parallel channels, then the capacity C_Π is given by

$$C_\Pi = \sum_{s=1}^{S} C_s$$

where C_s is the capacity of the s-th channel, $s \in [S]$. Two capacity-achieving schemes were provided in [1]:

1) A random coding scheme with decoding by joint typicality over product channels. The notion of product channels is defined in (1) where each possible permutation π provides a different product channel. Consequently, there are $S!$ possible product channels. A properly chosen random code was shown to achieve the capacity C_Π with decoding by joint typicality for all possible permutations π.

2) A rate-matching code together with random codebook generation and sequential decoding by joint typicality.

The construction technique for rate-matching codes in [1, Section 6C], based on MDS codes, provided an important intuition for the parallel polar schemes introduced in the next section.

For the binary coding schemes provided in this paper, it is assumed without any loss of generality that the message x_m is provided in terms of binary information (referred to as information bits or message bits). For the non-binary scheme, it is assumed that the message is provided in terms of information symbols, or message information symbols over a suitable non-binary finite field.

B. Polar Codes

The following basic definitions and results on polar codes (mainly extracted from [4] and [5]) are essential for the construction given in the next section. For a DMC, polar codes achieve the mutual information between an equiprobable input and the channel output.

Definition 2 (Symmetric binary-input channels). A DMC with a transition probability p, a binary-input alphabet $\mathcal{X} = \{0, 1\}$, and an output alphabet $\mathcal{Y}$ is said to be symmetric if there exists a permutation $\mathcal{T}$ over $\mathcal{Y}$ such that

1) The inverse permutation $\mathcal{T}^{-1}$ is equal to $\mathcal{T}$, i.e.,

$$\mathcal{T}^{-1}(y) = \mathcal{T}(y), \quad \forall y \in \mathcal{Y}.$$

2) The transition probability p satisfies

$$p(y|0) = p(\mathcal{T}(y)|1), \quad \forall y \in \mathcal{Y}.$$

Polar codes are defined in [4] using a recursive channel synthesizing operation which is referred to as channel combining. An alternative recursive algebraic construction is also provided in [4]. After $i \geq 1$ recursive steps, a $n \times n$ matrix G_n, where $n = 2^i$ is defined. The matrix G_n is refereed to as the polar generator matrix of size n.

Let $\mathcal{A}_n \subseteq [n]$, and denote by $\mathcal{A}_n^\mathrm{c}$ the complementary set of $\mathcal{A}_n$ (i.e., $\mathcal{A}_n^\mathrm{c} = [n] \setminus \mathcal{A}_n$). Given a set $\mathcal{A}_n$ and a polar generator matrix of size n, G_n, a class of block codes of block length n and code-rate $\frac{1}{n}|\mathcal{A}_n|$ are formed[1]. The set $\mathcal{A}_n$ is referred to as the information set. Polar codes are constructed by a specific choice of the information set $\mathcal{A}_n$.

The encoding of $|\mathcal{A}_n|$ information bits to a codeword $\mathbf{x} \in \{0, 1\}^n$ is carried in two steps. First, a binary length-n vector $\mathbf{w}$ is defined. Over the indices specified by $\mathcal{A}_n$, the components of $\mathbf{w}$ are set according to the information bits. The rest of the $|\mathcal{A}_n^\mathrm{c}|$ bits of $\mathbf{w}$ are predetermined and fixed according to a particular code design (these bits

---

[1]These codes can be shown to be coset codes.

are denoted as "frozen bits" in [4]). Next, a codeword is evaluated according to

$$\mathbf{x} = \mathbf{w}G_n \ . \tag{4}$$

Let p be a transition probability function of a binary-input DMC with an input-alphabet $\mathcal{X} = \{0, 1\}$ and an output-alphabet $\mathcal{Y}$. The equivalent synthesized channel construction, after $i \geq 1$ recursive steps, provides a channel denoted by p_n, $n = 2^i$, whose input is a binary vector in $\{0, 1\}^n$ and output in $\mathcal{Y}^n$. The channel p_n is noted as the combined channel in [4], and it can be shown to satisfy the equality

$$p_n(\mathbf{y}|\mathbf{w}) = p(\mathbf{y}|\, \mathbf{w}G_n) \quad \forall \ \mathbf{y} \in \mathcal{Y}^n \text{ and } \mathbf{w} \in \mathbf{X}^n \ . \tag{5}$$

Channel splitting is another important operation that is introduced in [4] for polar codes. The split channels $\{p_n^{(l)}\}_{l=1}^n$, all with a binary input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabets $\mathcal{Y}^n \times \mathcal{X}^{l-1}$, $l \in [n]$, are defined according to

$$p_n^{(l)}(\mathbf{y}, \mathbf{w}|x) \triangleq \frac{1}{|\mathcal{X}|^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} p_n(\mathbf{y}|(\mathbf{w}, x, \mathbf{c})) \tag{6}$$

where $\mathbf{y} \in \mathcal{Y}^n$, $\mathbf{w} \in \mathcal{X}^{l-1}$, and $x \in \mathcal{X}$. The importance of channel splitting is due to its role in the successive cancellation decoding procedure that is provided in [4]. Define

$$f_{\text{dec}}(p_n^{(l)}, \mathbf{y}, \mathbf{w}) \triangleq \arg\max_{x \in \mathcal{X}} p_n^{(l)}(\mathbf{y}, \mathbf{w}|x) \tag{7}$$

where $p_n^{(l)}$ is a split channel defined in (6), $\mathbf{y} \in \mathcal{Y}^n$, $\mathbf{w} \in \mathcal{X}^{l-1}$ and ties may be settled arbitrarily. For the particular case where $l = 1$, the parameter $\mathbf{w}$ is dropped from the notation. The decoding rule f defined in (7) may be interpreted as an optimal detection rule for a bit transmitted over the corresponding split channel. The decoding procedure for polar codes iterates over the index $l \in [n]$. If $l \in \mathcal{A}_n^c$, then the bit w_l is a predetermined and known bit. Otherwise, we decode the bit w_l according to $f_n(p, l, \mathbf{y}, (w_1, \ldots, w_{l-1}))$ where $\mathbf{y}$ is the received vector and $w_1, \ldots, w_{l-1}$ are the already decoded bits. It is shown in [4] that the described successive cancellation decoding procedure may be accomplished with a complexity of $\mathcal{O}(n \log n)$.

Lemma 1 (**Channel polarization properties [5]**). Let p be a binary-input symmetric DMC whose capacity is given by C and fix a rate $R < C$ and some $0 < \beta < \frac{1}{2}$. Then, there exists an information index set sequence $\mathcal{A}_n$ such that

1) Rate: $|\mathcal{A}_n| \geq nR$.
2) Performance: Assume that the information bits w_t, $t \in \mathcal{A}_n$, are chosen in a uniform manner over all possible options in $\{0, 1\}^{|\mathcal{A}_n|}$ and fix an arbitrary choice of the predetermined and fixed bits w_t, $t \in \mathcal{A}_n^c$. For every index $l \in \mathcal{A}_n$ the following upper bound is satisfied:

$$\Pr\left(\mathcal{E}_l(p)\right) \leq 2^{-n^\beta}$$

where

$$\mathcal{E}_l(p) \triangleq \left\{ p_n^{(l)}\left(\mathbf{y}, (w_1, w_2, \ldots, w_{l-1})\,|\,w_l\right) \leq p_n^{(l)}\left(\mathbf{y}, (w_1, w_2, \ldots, w_{l-1})\,|\,w_l + 1\right) \right\} \tag{8}$$

and the addition $w_l + 1$ on the right-hand side of (8) is carried modulo-2.

Remark 1 (**On the symmetry assumption in Lemma 1**). The symmetry of the channel in Lemma 1 is required in order to provide an arbitrary choice of the predetermined and fixed bits w_t, $t \in \mathcal{A}_n^c$. In the general case where the parallel channels are not necessarily output-symmetric, this vector can not be chosen arbitrarily (though the

results are satisfied for some choice).

The channel polarization phenomenon on q-ary channels has been considered in [8] and [9], where several sufficient conditions on the kernels have been derived for ensuring the occurrence of the channel polarization phenomenon. In the case where q is a power of 2, an explicit construction was provided in [8] in terms of an $n \times n$ generator polarization matrix G_n over $\mathrm{GF}(2^m)$ and an information index set sequence $\mathcal{A}_n$. Encoding of $|\mathcal{A}_n|$ message symbols to a codeword $\mathbf{x} \in \mathrm{GF}(2^m)$ is carried according to (4) where the operations are carried over the finite field $\mathrm{GF}(2^m)$, $\mathbf{w} = (w_1, \ldots, w_n) \in \mathrm{GF}(2^m)$, the symbol w_l is an information symbol for every $l \in \mathcal{A}_n$, and it is predetermined and fixed for every $l \notin \mathcal{A}_n$. Split channels and successive cancellation decoding procedures are defined similarly as in (6) and (7), except that the input alphabet $\mathcal{X}$ is no longer binary.

C. MDS codes

Some basic properties of MDS codes are provided. For complete details and proofs, the reader is referred, e.g., to [6] or [7].

Definition 3. An (n, k) linear block code $\mathcal{C}$ whose minimum distance is d is called a maximum distance separable (MDS) code if $d = n - k + 1$.

Since the minimum distance of an MDS code is $n - k + 1$, it follows that it can tolerate up to $n - k$ erasures, or in other words, any k symbols in a codeword completely determine the other symbols.

Example 1 (**MDS codes**). The $(n, 1)$ repetition code, $(n, n - 1)$ single parity-check (SPC) code, and the whole space of vectors over a finite field are all MDS codes.

In the following, we explain how to construct an MDS code of a block length S and a dimension $k \in [S]$. Let $S > 0$ be an integer number, and fix an integer $m > 0$ such that $2^m - 1 \geq S$. For every $k \in [2^m - 1]$, there exists a $(2^m - 1, k)$ Reed-Solomon (RS) code over the Galois field $\mathrm{GF}(2^m)$. Every RS code is an MDS code [7, Proposition 4.2]. To obtain an (S, k) MDS code, two alternatives are suggested:

1) **Punctured RS codes:** Consider a $(2^m - 1, k)$ RS code over the Galois field $\mathrm{GF}(2^m)$. Deleting $2^m - 1 - S$ columns from the generator matrix of the considered code results in an (S, k) linear block code over the same alphabet. The resulting code is an (S, k) MDS code over $\mathrm{GF}(2^m)$.

2) **Generalized RS (GRS) codes:** GRS codes are MDS codes which can be constructed over $\mathrm{GF}(2^m)$ for every block length S and dimension k (as long as $2^m - 1 \geq S$).

III. THE PROPOSED CODING SCHEMES

We first provide a simplified version of the proposed scheme that is suitable for $S = 3$ parallel channels, relying on binary polar codes and binary MDS codes (note that the scheme for $S = 2$ can be directly obtained from the studied case where $S = 3$). For $S > 3$, this scheme must be generalized to utilize non-binary MDS codes. Two alternative schemes are therefore proposed: a scheme based on non-binary polar codes and a scheme based on binary interleaved polar codes. For the special case where the channels are stochastically degraded, a simplification is possible based on non-binary MDS codes and binary (non-interleaved) polar codes.

A. A Simplified Coding Scheme for $S = 3$

Let $\mathcal{A}_n^{(1)}$, $\mathcal{A}_n^{(2)}$ and $\mathcal{A}_n^{(3)}$ be three information bit sets, and let $k_n \triangleq |\mathcal{A}_n^{(1)}| + |\mathcal{A}_n^{(2)}| + |\mathcal{A}_n^{(3)}|$. The polar encoding is preceded by mapping k_n information bits to three length-n binary vectors $\mathbf{w}_s = (w_{s,1}, w_{s,2}, \ldots, w_{s,n}) \in \{0,1\}^n$, for $s = 1, 2, 3$, as follows:

1) The k_n bits of $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$, referring to the set union $\mathcal{A}_n^{(1)} \cup \mathcal{A}_n^{(2)} \cup \mathcal{A}_n^{(3)}$, are set to the values of the k_n information bits.

2) For every $l \in [n]$, consider the binary triple $(w_{1,l}, w_{2,l}, w_{3,l})$ and fill the remaining bits as follows:

   a) If none of the bits in $(w_{1,l}, w_{2,l}, w_{3,l})$ are information bits, they are set to some arbitrarily fixed values, whose values are made known to both the encoder and the decoder.

   b) If one (and only one) bit in $(w_{1,l}, w_{2,l}, w_{3,l})$ is an information bit, the remaining two bits are set to the same value as this information bit.

   c) If two (and only two) of the bits in $(w_{1,l}, w_{2,l}, w_{3,l})$ are information bits, the remaining bit is set to the exclusive-or value of the two information bits.

Finally, the codewords $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are calculated via the equality

$$\mathbf{x}_s \triangleq \mathbf{w}_s G_n, \quad s \in [3]$$

where G_n is the generator matrix of the polar code.

The codeword $\mathbf{x}_{\pi(s)}$ is then transmitted over the symmetric channel P_s (see Definition 2), $s \in [3]$, as depicted in Figure 1. The split channels defined in (6) are therefore evaluated with respect to the permuted indices of the transmitted vectors as well. Specifically, let $\mathbf{y}_s$ denotes the length-n observation vector received at the output of the channel P_s, $s \in [3]$, and the corresponding split channels are evaluated with respect to the binary vector $\mathbf{w}_{\pi(s)}$, $s \in [3]$. Given previously decoded bits $w_{\pi(s),1}, w_{\pi(s),2}, \ldots, w_{\pi(s),l-1}$ for some $s \in [3]$ and $l \in [n]$, the bit $w_{\pi(s),l}$ is decoded based on the split channel

$$P_{s,n}^{(l)}\big(\mathbf{y}_s, w_{\pi(s),1}, \ldots, w_{\pi(s),l-1}|w\big) = \frac{1}{2^{n-1}} \sum_{\mathbf{c} \in \{0,1\}^{n-l}} P_s\big(\mathbf{y}_s|(w_{\pi(s),1}, \ldots, w_{\pi(s),l-1}, w, \mathbf{c})G_n\big) \tag{9}$$

where $w \in \{0,1\}$ is the binary input to the considered split channel.

The l-th symbol (for $l = 1, 2, \ldots, n$) in each codeword is decoded sequentially as follows:

1) If $l \in \mathcal{A}_n^{(s)}$ for every $s \in [3]$, then decode

$$w_{\pi(s),l} = f_{\text{dec}}\big(P_{s,n}^{(l)}, \mathbf{y}, (w_{\pi(s),1}, w_{\pi(s),2}, \ldots, w_{\pi(s),l-1})\big), \quad s \in [3] \tag{10}$$

where f_{dec} is the decoding rule in (7), and $P_{s,n}^{(l)}$, $s \in [3]$ are the split channels in (9).

2) Otherwise, if $l \in \mathcal{A}_n^{(s)}$ and $l \in \mathcal{A}_n^{(s')}$ for some $1 \leq s < s' \leq 3$, then decode $w_{\pi(s),l}$ as in (10) and $w_{\pi(s'),l}$ as in (10) with s replaced by s'. Furthermore, set the remaining bit $w_{\pi(s^*),l}$ (where $s^* \neq s, s'$ and $s^* \in [3]$) to $w_{\pi(s),l} + w_{\pi(s'),l}$.

3) Otherwise, if $l \in \mathcal{A}_n^{(s)}$ for a single $s \in [3]$, decode $w_{\pi(s),l}$ as in (10). Then, set the remaining two bits $w_{\pi(s'),l}$ and $w_{\pi(s''),l}$ (where $s' \neq s$ and $s'' \neq s$) to $w_{\pi(s),l}$.

Note that at each decoding stage, all the triples that precede the current stage are already determined, matching the evaluation requirement of the corresponding split channels as given in (9).

Proposition 1. The parallel binary polar coding scheme for $S = 3$ achieves the capacity C_Π of the arbitrarily-permuted parallel channels where these three channels are memoryless, binary-input and output-symmetric.

Proof: Fix an arbitrary rate triple (R_1, R_2, R_3) satisfying $R_s < C_s$ for $s = 1, 2, 3$, and some $0 < \beta < \frac{1}{2}$. The error probability P_e of the provided decoding procedure is upper bounded, via the union bound, by

$$P_e \leq \sum_{s \in [3]} \sum_{l \in \mathcal{A}_n^{(s)}} \Pr\big(\mathcal{E}_l(P_s)\big) \tag{11}$$

where $\mathcal{E}_l(P_s)$ in (8) is the error event that a decision on a bit in the split channel is incorrect. According to Lemma 1, there exists index set sequences $\mathcal{A}_n^{(s)}$, $s \in [3]$, such that the number of information bits k_n satisfies

$$k_n \geq n(R_1 + R_2 + R_3)$$

while assuring that for symmetric channels (see Remark 1) the decoding error probability P_e in (11) is upper bounded by

$$P_\text{e} \leq n2^{-n^\beta}.$$

Taking the block length n large enough concludes the proof.   ∎

It is clear that the repetition and exclusive-or operation are essentially the encoding operations for the binary $(3,1)$ and $(3,2)$ MDS codes, respectively. The information bits and the fixed bits in the polar code framework naturally lead to the application of symbol-level MDS codes. For $S > 3$, because the appropriate MDS codes only exist for larger alphabets, the coding operations are not exactly performed on the single bit level. However, as we shall discuss next, this difficulty can be solved by using non-binary polar codes or an interleaving technique.

B. Coding for $S > 3$ Using Non-Binary Polar Codes

For $S > 3$, the binary MDS codes applied in Section III-A must be replaced by MDS codes of block length S. The only binary MDS codes are the trivial codes (repetition, single parity-check and the whole space). As MDS codes of additional dimensions are required (for $S > 3$), we must turn to larger alphabets. For each $k \in [S]$, an (S, k) MDS codes over the Galois field GF(2^m) is chosen, which is denoted by $\mathcal{C}_k$ (see Section II-C for possible constructions based on RS and GRS codes). A singleton set, whose sole member is an arbitrary and fixed length-S binary vector is also chosen. This singleton set is denoted by the codebook $\mathcal{C}_0$.

In order to apply the non-binary polarization coding scheme, a new set of parallel channels $\{W_s\}_{s=1}^{S}$ is defined according to

$$W_s(\mathbf{y}|x) \triangleq \prod_{i=1}^{m} P_s(y_i|b_i)$$

where $\mathbf{y} = (y_1, \ldots, y_m) \in \mathcal{Y}_s$, $x \in$ GF(2^m), $s \in [S]$, $\big(b_1(x), \ldots, b_m(x)\big)$ is the binary m-length vector representation of the symbol $x \in$ GF(2^m) and P_s, $s \in [S]$ are the binary-input symmetric parallel DMC over which the communication takes place. The corresponding split channels are denoted by $W_{s,n}^{(l)}$, $l \in [n]$. A coding scheme for the parallel channels W_s, $s \in [S]$ is equivalent to a coding scheme for the original binary parallel channels where the transmission of a symbol x over a channel W_s is replaced with m transmissions over the channel P_s, $s \in [S]$. With some abuse of notations, the information index set sequence for each of the non-binary channels W_s, $s \in [S]$, is also denoted by $\mathcal{A}_n^{(s)}$. For every $l \in [n]$ define

$$k_l \triangleq |\{s : \ l \in A_n^{(s)}\}|. \tag{12}$$

The encoding of the parallel non-binary polarization scheme is carried as follows:

1) For every channel index $s \in [S]$ and every information index $l \in \mathcal{A}_n^{(s)}$, denote by $a_s^{(l)}$ the symbol in GF(2^m) corresponding to m information bits.
2) For every $l \in [n]$, choose the unique codeword $\mathbf{c}^{(l)} = (c_1^{(l)}, c_2^{(l)}, \ldots, c_S^{(l)}) \in \mathcal{C}_{k_l}$, satisfying $c_{s'}^{(l)} = a_{s'}^{(l)}$ for every $s' \in \{s : \ l \in A_n^{(s)}\}$.
3) Compute S polar codewords $\mathbf{x}_s$, for $s \in [S]$, according to

$$\mathbf{x}_s = \big(c_s^{(1)}, c_s^{(2)}, \ldots, c_s^{(n)}\big) \cdot G_n$$
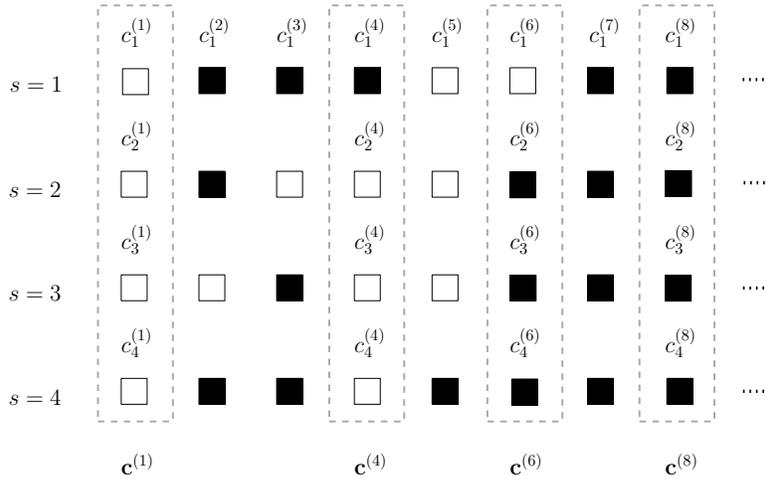
Fig. 2: Illustration of the non-binary parallel polar encoding procedure in the particular case of $S = 4$. The grid of rectangles illustrates the symbols $c_s^{(l)}$, for $1 \le l \le 8$ and $s \in [4]$, as are defined in the encoding procedure. Each row of squares represents the vector $(c_s^{(1)}, c_s^{(2)}, ...c_s^{(8)})$, $s \in [4]$, where each of the squares represents a symbol. A filled square represents a symbol $c_s^{(l)}$ for which $l \in \mathcal{A}_n^{(s)}$, $s \in [4]$. For the depicted grid, $A_n^{(1)} \cap [8] = \{2, 3, 4, 7, 8\}$, $A_n^{(2)} \cap [8] = \{2, 6, 7, 8\}$, $A_n^{(3)} \cap [8] = \{3, 6, 7, 8\}$ and $A_n^{(4)} \cap [8] = \{2, 3, 5, 6, 7, 8\}$. According to the decoding procedure, the symbols represented by the filled squares are set to the message symbols. An empty square represents the opposite case where $l \notin \mathcal{A}_n^{(s)}$, $s \in [4]$. The symbols represented by the empty squares are determined such that each column forms an MDS codeword. The 4 vertical rectangles mark 4 of these codewords: $\mathbf{c}^{(1)}$, $\mathbf{c}^{(4)}$, $\mathbf{c}^{(6)}$ and $\mathbf{c}^{(8)}$. Codewords $\mathbf{c}^{(1)}$ and $\mathbf{c}^{(8)}$ belong to codes of dimensions 0 and 4, respectively (a constant vector and the whole space). Accordingly, in $\mathbf{c}^{(1)}$ all 4 squares are empty to represent 4 predetermined and fixed symbols while in $\mathbf{c}^{(8)}$ all 4 squares are filled squares, representing 4 arbitrary information symbols (an arbitrary vector in the whole space). The codeword $\mathbf{c}^{(4)}$ belongs to a code of dimension 1, accordingly $c_2^{(1)} = c_3^{(1)} = c_4^{(1)} = c_1^{(1)}$ (the empty squares equal to the value of the single filled square). The codeword $\mathbf{c}^{(6)}$ belongs to a code of dimension 3 where the 3 filled squares completely determine the value of the single empty square according to $c_1^{(6)} = -c_2^{(6)} - c_3^{(6)} - c_4^{(6)}$.

where G_n is the polar generator matrix, and arithmetic is carried over $\mathrm{GF}(2^m)$.

The encoding procedure is further detailed in Figure 2 via an illustrative example. The codeword $\mathbf{x}_s$, where $s \in [S]$, is transmitted over the channel W_s, and let $\mathbf{y}_s$ denote the vector received at the output of the channel W_s. The l-th symbol (for $l = 1, 2, \ldots, n$) of each codeword is decoded sequentially as follows:

1) For every $s \in [S]$ such that $l \in A_n^{(s)}$, let $c_{\pi(s)}^{(l)} = f_{\mathrm{dec}}(W_{s,n}^{(l)}, \mathbf{y}_s, c_{\pi(s)}^{(1)}, c_{\pi(s)}^{(2)}, \ldots, c_{\pi(s)}^{(l-1)})$.

2) Find the unique codeword $\mathbf{c} = (c_1, c_2, \ldots, c_S)$ in $\mathcal{C}_{k_l}$ satisfying $c_{\pi(s')}^{(l)} = c_{\pi(s')}$ for every $s' \in \{s : l \in A_n^{(s)}\}$.

The decoding procedure is further detailed in Figure 3. via an illustrative example.

Proposition 2. The parallel non-binary polar coding scheme achieves the capacity of the considered model.

Proof: The ability to choose a unique codeword in $\mathcal{C}_{k_l}$, $l \in [n]$, follows directly from the fact that an (S, k_l) MDS code can correct up to $S - k_l$ erasures. For a DMC W with an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, define the events

$$\mathcal{E}_l^d(W) \triangleq \left\{ (\mathbf{w}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : W_n^{(l)}(\mathbf{y}, (w_1, \ldots, w_{l-1})|w_l) \le W_n^{(l)}(\mathbf{y}, (w_1, \ldots, w_{l-1})|w_l + d) \right\}, \quad l \in [n], d \in \mathcal{X} \tag{13}$$

where $W_n^{(l)}$, $l \in [n]$ are the split channels of W. The error probability P_e for the non-binary decoding procedure is upper bounded by

$$P_e \le \sum_{s \in [S]} \sum_{l \in \mathcal{A}_n^{(s)}} \sum_{d \in \mathcal{X} \setminus \{0\}} \Pr\left(\mathcal{E}_l^d(W_s)\right). \tag{14}$$

It follows from [8] and [9] that the probability of the event $\mathcal{E}_l^d(W_s)$ can be made exponentially low as the block length increase while having the cardinality of the information sets arbitrarily close to the capacity of the corresponding DMC. Hence, the error probability in (14) can be made arbitrarily low. Detailed inspection of the results in [8]
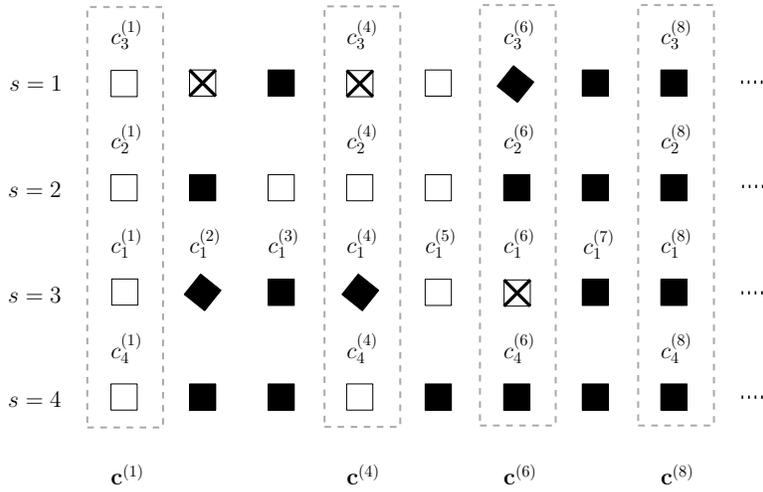
Fig. 3: Illustration of the non-binary parallel polar decoding procedure in the particular case of $n = 8$ and $S = 4$. The grid of rectangles refers to Figure 2 with the difference that, due to the transmission permutation π, codeword $\mathbf{x}_1$ is transmitted over channel P_3, codeword $\mathbf{x}_3$ is transmitted over channel P_1, and codewords $\mathbf{x}_2$ and $\mathbf{x}_4$ are transmitted over channels P_2 and P_4, respectively. All the symbols in $\mathbf{c}^{(1)}$ (empty squares) are predetermined and fixed, so the first decoding stage is redundant. Due to the channel permutation, some fixed symbols may be decoded via (7). Such fixed symbols are represented by empty squares filled with an x-mark (e.g., $c_3^{(4)}$, note that $4 \in \mathcal{A}_n^{(1)}$). As another consequence of the channel permutation, some information symbols cannot by decoded via (7). Such information symbols are represented by filled and rotated squares (e.g., $c_1^{(4)}$, note that $4 \notin \mathcal{A}_n^{(3)}$). Consequently, at the forth decoding stage, even though the message symbol is $c_1^{(4)}$ (represented by a filled rotated square), due to the transmission permutation only the symbol $c_3^{(4)}$ that is represented by an empty x-marked square can be decoded via (7). Nevertheless, due to the MDS structure in the columns, the two symbols are equal. At the sixth stage of the decoding, the message symbol $c_3^{(6)}$ is the rotated square and $c_1^{(6)}$ is now x-marked (as $6 \notin \mathcal{A}_n^{(1)}$ and $6 \in \mathcal{A}_n^{(3)}$). The message symbols (filled squares) $c_2^{(6)}$ and $c_4^{(6)}$ can be decoded via (7) but $c_3^{(6)}$ cannot. Nevertheless, the non-message symbol (filled x-marked square) $c_1^{(6)}$ is decoded via (7) and due to the MDS structure of columns, the message symbol (rotated square) $c_3^{(6)} = -c_1^{(6)} - c_2^{(6)} - c_4^{(6)}$. All the symbols (filled squares) in $\mathbf{c}^{(8)}$ are decoded via (7) as $8 \in \mathcal{A}_n^{(s)}$ for every $s \in [4]$.

and [9] reveal the lack of the symmetry property in Remark 1 which is crucial for the provided scheme. This property is therefore provided for non-binary polar codes in Appendix A. The symmetry of W_s, $s \in [S]$ according to Definition 4 for non-binary channels is imposed directly from the symmetry of the binary-input channels P_s, $s \in [S]$ according to Definition 2. ∎

Remark 2 (Coding for non-binary parallel symmetric channels). The coding scheme provided in this section can be easily adapted to parallel, output-symmetric and memoryless channels where the cardinality of the input alphabet is a power of a prime (the symmetry condition in the non-binary case is stated in Definition 4 of Appendix A).

C. A Binary Interleaved Polar Coding Scheme

The scheme provided in this section is based on $m > 1$ binary interleaved polar codes for every binary-input symmetric DMC P_s, $s \in [S]$. The m interleaved polar codes for each channel P_s, $s \in [S]$, are defined based on the same information set sequence $\mathcal{A}_n^{(s)}$. As in Section III-B, let $\mathcal{C}_k$ denote an MDS code over $\mathrm{GF}(2^m)$ of dimension k, and let k_l be defined as in (12). The encoding process is carried as follows:

1) For every information index $l \in \mathcal{A}_n^{(s)}$, and every channel index $s \in [S]$: Pick m information bits, denoted by $u_{(l-1)m+g}^{(s)}$, $1 \le g \le m$.

2) For every $l \in [n]$, choose the unique codeword $\mathbf{c}^{(l)} = (c_1^{(l)}, c_2^{(l)}, \dots, c_S^{(l)}) \in \mathcal{C}_{k_l}$ for which the binary representation of $c_{s'}^{(l)} \in \mathrm{GF}(2^m)$ is equal to

$$\left(u_{(l-1)m+1}^{(s)}, \dots, u_{(l-1)m+m}^{(s)} \right)$$

for every $s' \in \{s : l \in A_n^{(s)}\}$.

3) For every $s \in [S]$ and index $l \notin \mathcal{A}_n^{(s)}$, define the length-$m$ binary vector

$$\left(u_{(l-1)m+1}^{(s)}, u_{(l-1)m+2}^{(s)}, \ldots, u_{(l-1)m+m}^{(s)}\right) \in \{0,1\}^m$$

as the binary representation of the symbol $c_s^{(k)}$.

4) Compute the $m \cdot S$ polar codewords $\mathbf{x}_{g,s} \in \{0,1\}^n$, $g \in [m]$, $s \in [S]$ where

$$\mathbf{x}_{g,s} = \left(u_g^{(s)}, u_{m+g}^{(s)}, \ldots, u_{(n-1)m+g}^{(s)}\right) \cdot G_n$$

and G_n is the binary polar generator matrix.

5) For every channel index $s \in [S]$, construct a codeword $\mathbf{x}^{(s)}$ based on the concatenation

$$\mathbf{x}^{(s)} = \left(\mathbf{x}_{1,s}, \mathbf{x}_{2,s}, \ldots, \mathbf{x}_{m,s}\right).$$

The concatenated codeword $\mathbf{x}^{(\pi(s))}$ is transmitted over the channel P_s, $s \in [S]$, and let $\mathbf{y}^{(s)} = (y_1^{(s)}, ..., y_{mn}^{(s)})$ denote the received vector at the output of this channel.

Assuming that the bits $u_{m(l'-1)+g}^{(s)}$ ($s \in [S]$, $g \in [m]$ and $l' \le l-1$) were already decoded, the bits $u_{m(l-1)+g}^{(s)}$ are decoded sequentially at the l-th stage (for $l = 1, \ldots, n$) as follows:

1) For every $s \in [S]$ such that $l \in A_n^{(s)}$, decode

$$u_{(l-1)m+g}^{(\pi(s))} = f_{\text{dec}}\left(P_{s,n}^{(l)}, \left(y_{1+(g-1)n}^{(s)}, y_{2+(g-1)n}^{(s)}, \ldots, y_{gn}^{(s)}\right), \left(u_g^{(\pi(s))}, u_{m+g}^{(\pi(s))}, \ldots, u_{(l-1)m+g}^{(\pi(s))}\right)\right), \quad g \in [m]$$

where f_{dec} and $P_{s,n}^{(l)}$ are defined in (7) and (9), respectively.

2) Find the unique codeword $\mathbf{c} = (c_1, c_2, \ldots, c_S) \in \mathcal{C}_{k_l}$ for which the symbol $c_{\pi(s')}$ is equal to

$$\left(u_{(l-1)m+1}^{(\pi(s'))}, u_{(l-1)m+2}^{(\pi(s'))}, \ldots, u_{lm}^{(\pi(s'))}\right)$$

for every $s' \in \{s : l \in A_n^{(s)}\}$.

3) For every $s \in [S]$ for which $l \notin A_n^{(s)}$, the bits $\left(u_{(l-1)m+1}^{(\pi(s'))}, u_{(l-1)m+2}^{(\pi(s'))}, \ldots, u_{lm}^{(\pi(s'))}\right)$ are set according to the binary representation of the symbol $c_{\pi(s')} \in \text{GF}(2^m)$.

Proposition 3. The parallel binary-interleaved parallel polar coding scheme achieves the capacity of the considered model of parallel channels.

Proof: The ability to choose unique codewords in $\mathcal{C}_{k_l}$, $l \in [n]$, follows directly from the fact that an (S, k_l) MDS code can correct up to $S - k_l$ erasures. For every channel $s \in [S]$, m interleaved polar codes of block length n are applied. Hence, the code rate R_n of the parallel binary-interleaved polar scheme is given by

$$R_n = \sum_{s=1}^{S} \frac{m \, |\mathcal{A}_n^{(s)}|}{m \, n} = \frac{1}{n} \sum_{s=1}^{S} |\mathcal{A}_n^{(s)}|.$$

For n sufficiently large, it follows from Lemma 1 that R_n can be made arbitrarily close to $\sum_{s=1}^{S} C_s$. The part of the proof that refers the reliability of the provided decoding procedure is omitted as it follows from Lemma 1, and it goes along similar steps as the proof of Proposition 1. ∎

D. Coding for Stochastically Degraded Channels

The non-binary scheme provided in Section III-B can be simplified to include only binary polar codes if the parallel channels are assumed to be stochastically degraded (without relying on binary interleavers as required in Section III-C). The scheme is further simplified in terms of the decoding procedure. Instead of performing

successive cancellation in parallel for all the received sequences simultaneously, the decoding procedure is performed sequentially channel by channel. The simplification follows from the following technical property:

Corollary 1 (**On monotonic information sets for stochastically degraded parallel channels**). Consider a set of S memoryless, binary-input and output-symmetric parallel channels $\{P_s\}_{s=1}^S$. Assume that the channels are stochastically degraded, such that $P_{s'}$ is a degraded version of P_s for every $s' > s \in [S]$. Let C_s be the capacity of the channel P_s, $s \in [S]$. Fix $0 < \beta < \frac{1}{2}$ and a set of rates $R_1, \ldots, R_{[s]}$ such that $0 \le R_s \le C_s$ for every $s \in [S]$. Then, there exists a sequence of information sets $\mathcal{A}_n^{(s)} \subseteq [n]$, $s \in [S]$ and $n = 2^i$ where $i \in \mathbb{N}$, satisfying the following properties:

1) Rate:
$$|\mathcal{A}_n^{(s)}| \ge nR_s, \quad \forall s \in [S]. \tag{15}$$

2) Monotonicity:
$$\mathcal{A}_n^{(S)} \subseteq \mathcal{A}_n^{(S-1)} \subseteq \cdots \subseteq \mathcal{A}_n^{(1)}. \tag{16}$$

3) Performance:
$$\Pr\big(\mathcal{E}_l(P_s)\big) \le 2^{-n^\beta} \tag{17}$$

for all $l \in \mathcal{A}_n^{(s)}$ and $s \in [S]$, and

$$\mathcal{E}_l(p) \triangleq \Big\{ p_n^{(l)}(\mathbf{y}, \mathbf{w}^{(l-1)}|w_l) \le p_n^{(l)}(\mathbf{y}, \mathbf{w}^{(l-1)}|w_l + 1) \Big\}, \quad l \in [n]$$

Proof: See Appendix B. ∎

Let $\{P_s\}_{s=1}^S$ be a set of parallel channels as in Corollary 1, and let $\mathcal{C}_k$, $1 \le k \le S-1$ denote an MDS code over $\text{GF}(2^m)$ of block length S and dimension k. Define, $k_n^{(s)} \triangleq |\mathcal{A}_n^{(s)}|$, $s \in [S]$, where $\mathcal{A}_n^{(s)}$ is the information index set sequence of the channel P_s, $s \in [S]$, satisfying the properties in (15)-(17). In addition, define $K_{s-1} \triangleq (k_n^{(s-1)} - k_n^{(s)})/m$, $s \in [S]$ (for the purpose of simplicity, it is assumed that $k_n^{(s)}$ are integral multiples of m). Prior to the stage of polar encoding, $k_n = \sum_{s \in [S]} k_n^{(s)}$, information bits are mapped into a set of binary row vectors $\{\mathbf{u}_{s,l}\}$, $s, l \in [S]$ where the vector $\mathbf{u}_{s,l}$ is of length $k_n^{(S-l+1)} - k_n^{(S-l+2)}$ bits. The vectors $\mathbf{u}_{s,1}$, $s \in [S]$ and $\mathbf{u}_{s,2} = \big(u_{s,2}(1), u_{s,2}(2), \ldots, u_{s,2}(k_{S-1} - k_S)\big)$, $s \in [S-1]$, are set to information bits. Next, the vector $\mathbf{u}_{S,2}$ is determined (the following steps are accompanied with the illustration in Figure 4):

1) Construct the $(S-1) \times K_{S-1}$ matrix over $\text{GF}(2^m)$, $C^{(2)}$, from the row vectors $\mathbf{u}_{s,2}$ ($s \in [S-1]$) where the (i, j) element in this matrix is defined by the m bits

$$\Big(\mathbf{u}_{i,2}\big((j-1)m+1\big), \mathbf{u}_{i,2}\big((j-1)m+2\big), \ldots, \mathbf{u}_{i,2}\big(jm\big)\Big), \quad i \in [S-1], \ j \in [K_{S-1}]$$

(see Figure 4 where each vector is represented with a horizontal rectangle).

2) Find the unique codewords $\{\mathbf{c}_j : \ j \in [K_{S-1}]\}$ in $\mathcal{C}_{S-1}$, whose first $S-1$ symbols are the columns of $C^{(2)}$ (represented by the dashed vertical rectangles in Figure 4).

3) A K_{S-1}–length vector $\tilde{\mathbf{u}}_{S,2}$ over $\text{GF}(2^m)$ is defined using the last symbol of each of the codewords $\mathbf{c}_j$, $j \in [K_{S-1}]$ (the symbols are represented by filled black squares in Figure 4).

4) The vector $\mathbf{u}_{S,2}$ is defined by the binary representation of the vector $\tilde{\mathbf{u}}_{S,2}$.

Let $2 < l \le S$, and assume that the vectors $\mathbf{u}_{s,l'}$, $s \in [S]$, $l' < l$, are already defined. The vectors $\mathbf{u}_{s,l}$, $s \in [S]$ are defined as follows:

1) The binary row vectors $\mathbf{u}_{s,l}$, $1 \le s \le S - (l-1)$, are set to information bits.

2) Construct the $(S - (l-1)) \times K_{S-(l-1)}$ matrix $C^{(l)}$ over $\text{GF}(2^m)$ from the row vectors in step 1, where the
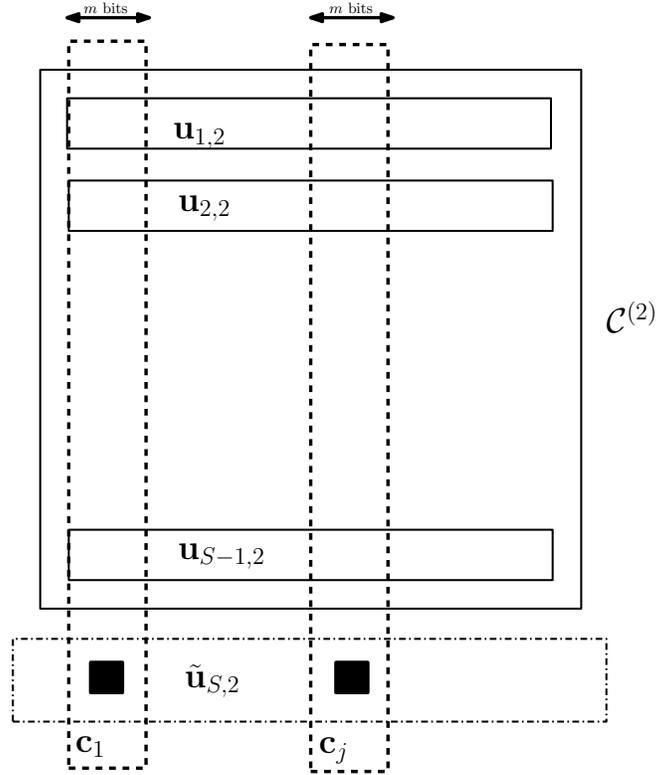
Fig. 4: Illustration of the construction of the vector $\tilde{\mathbf{u}}_{S,2}$. The vectors $\mathbf{u}_{k,2}$, $k \in [S-1]$ defining the rows of the matrix $C^{(2)}$ are shown, along with the columns defining the codewords $\mathbf{c}_j$, $j \in [K_{S-1,S}]$ in $\mathcal{C}_{\mathrm{MDS}}^{(S-1)}$.

(i, j) element of $C^{(l)}$ is defined by the m bits

$$\Big(\mathbf{u}_{i,l}\big((j-1)m+1\big), \mathbf{u}_{i,l}\big((j-1)m+2\big), \ldots, \mathbf{u}_{i,2}\big(jm\big) \Big).$$

3) Find the unique codewords $\mathbf{c}_j = (c_{j,1}, c_{j,2}, \ldots, c_{j,S}) \in \mathcal{C}_{S-(l-1)}$, $j \in [K_{S-(l-1)}]$, whose first $S - (l-1)$ symbols are the column of $C^{(l)}$.

4) The vectors $\mathbf{u}_{s,l}$, $s > S - (l-1)$ are set to the binary representation of $\tilde{\mathbf{u}}_{s,l}$.

Finally, the codewords $\mathbf{x}_{\pi(s)}$ is transmitted over the channel P_s, $s \in [S]$, where

$$\mathbf{x}_s = \sum_{l=1}^{S} \mathbf{u}_{s,l} G_n \left(\mathcal{A}_n^{(S-(l-1))} \setminus \mathcal{A}_n^{(S-(l-2))} \right) + \mathbf{b} G_n \left([n] \setminus \mathcal{A}_n^{(1)} \right), \quad s \in [S].$$

Here $\mathcal{A}_n^{(S+1)} \triangleq \emptyset$, $\mathbf{b}$ is a binary predetermined and fixed vector, and G_n is the polar generator matrix.

The decoding process starts with the observations received from the channel P_1. A polar successive cancellation decoding, with respect to the information index set $\mathcal{A}_n^{(1)}$, is applied to the received vector. This allows the decoding of the vectors $\mathbf{u}_{\pi(1),l}$, $l \in [S]$. Next, the decoding proceeds to successive cancellation decoding procedure for the vector received at the output of the channel P_2 (i.e., the channel with the second largest capacity). This decoding procedure is capable of decoding $|\mathcal{A}_n^{(2)}|$ bits based on $n - |\mathcal{A}_n^{(2)}|$ predetermined and fixed bits. For the current decoding procedure, $n - |\mathcal{A}_n^{(1)}|$ of these bits are the predetermined and fixed bits in $\mathbf{b}$. The rest of $|\mathcal{A}_n^{(1)}| - |\mathcal{A}_n^{(2)}|$ bits are based on the bits decoded at the previous decoding stage. Specifically, the bit vector $\mathbf{u}_{\pi(2),S}$ can be evaluated using the bit vector $\mathbf{u}_{\pi(1),S}$ due to incorporated MDS codes. After the second decoding stage, all the S binary vectors $\mathbf{u}_{\pi(2),s}$, $s \in [S]$, are fully determined. Moreover, based on the codewords $\mathbf{c}_j$, $j \in [K_{1,2}]$, the vectors $\mathbf{u}_{\pi(s),S}$, are fully determined for all $s \geq 2$ as well.

TABLE I: The order of decoding the information bits for all possible assignments of codewords over a set of three parallel and degraded channels.

Channel P_1		Channel P_2		Channel P_3	
Transmitted Codeword	**Decoded Information**	**Transmitted Codeword**	**Decoded Information**	**Transmitted Codeword**	**Decoded Information**
$\mathbf{x}_1$	$\mathbf{u}_{1,1}$, $\mathbf{u}_{1,2}$, $\mathbf{u}_r$	$\mathbf{x}_2$	$\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$	$\mathbf{x}_3$	$\mathbf{u}_3$
		$\mathbf{x}_3$	$\mathbf{u}_3$, $\mathbf{u}_{1,2} + \mathbf{u}_{2,2}$	$\mathbf{x}_2$	$\mathbf{u}_{2,1}$
$\mathbf{x}_2$	$\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$, $\mathbf{u}_r$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$, $\mathbf{u}_{1,2}$	$\mathbf{x}_3$	$\mathbf{u}_3$
		$\mathbf{x}_3$	$\mathbf{u}_3$, $\mathbf{u}_{1,2} + \mathbf{u}_{2,2}$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$
$\mathbf{x}_3$	$\mathbf{u}_3$, $\mathbf{u}_{1,2} + \mathbf{u}_{2,2}$, $\mathbf{u}_r$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$, $\mathbf{u}_{1,2}$	$\mathbf{x}_2$	$\mathbf{u}_{2,1}$
		$\mathbf{x}_2$	$\mathbf{u}_{2,1}$, $\mathbf{u}_{2,2}$	$\mathbf{x}_1$	$\mathbf{u}_{1,1}$

Next, the remaining $S - 2$ decoding stages are followed. Note that after the $(s-1)$-th decoding stage, where $2 < s < S$, the vectors $\mathbf{u}_{\pi(s'),l}$ for either $1 \leq s' < s$ and $l \in [S]$, or $s' \geq s$ and $S - s + 3 \leq l \leq S$, were decoded at previous stages. At the s-th stage, the decoding is extended for the vectors $\mathbf{u}_{\pi(s),l}$ for all $l \in [S]$ and the vectors $\mathbf{u}_{\pi(s'),S-s+2}$ for all $s' \in [S]$. In order to apply the polar successive cancellation decoding procedure to the vector received over the channel P_s, the bits in $\mathbf{b}$ and $\{\mathbf{u}_{\pi(s),l}\}_{l \geq S-(s-2)}$ must be known. The vector $\mathbf{b}$ is clearly known. In addition, the bits in $\{\mathbf{u}_{\pi(s),l}\}_{l \geq S-(s-3)}$ are already decoded in previous stages. It is left to determine the bits in $\mathbf{u}_{\pi(s),S-(s-2)}$. Nevertheless, these bits are fully determined due to the algebraic constraints imposed by the MDS codes (the determination of $\mathbf{u}_{\pi^{-1}(s),S-(s-2)}$ is also established along with the determination of $\mathbf{u}_{\pi^{-1}(s'),S-(s-2)}$ for all $s' \geq s$).

The proof of the following proposition goes along similar steps as in Proposition 1, and it is therefore omitted.

Proposition 4. The provided parallel coding scheme achieves the capacity of the considered model of parallel and degraded channels.

Example 2 (**Coding for 3 stochastically degraded channels**). The coding scheme described in this section is exemplified for the particular case of three parallel degraded channels P_1, P_2 and P_3. It is assumed that P_3 is a degraded version of P_2, and P_2 is a degraded version of P_1. We first describe the encoding:

- The k_1 information bits that are used to encode $\mathbf{x}_1$ are (arbitrarily) partitioned into three subsets: $\mathbf{u}_{1,1} \in \mathcal{X}^{k_3}$, $\mathbf{u}_{1,2} \in \mathcal{X}^{k_2-k_3}$ and $\mathbf{u}_r \in \mathcal{X}^{k_1-k_2}$.
- The k_2 information bits used to encode $\mathbf{x}_2$ are (arbitrarily) partitioned into two subsets: $\mathbf{u}_{2,1} \in \mathcal{X}^{k_3}$ and $\mathbf{u}_{2,2} \in \mathcal{X}^{k_2-k_3}$. In addition, $\mathbf{u}_r$ (used for encoding $\mathbf{x}_1$) is also involved in the encoding of $\mathbf{x}_2$.
- The codewords $\mathbf{x}_1$ and $\mathbf{x}_2$ are defined as follows:

$$\mathbf{x}_1 = \mathbf{u}_{1,1}G_n\left(\mathcal{A}_n^{(3)}\right) + \mathbf{u}_{1,2}G_n\left(\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}\right) + \mathbf{u}_r G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b}G_b\left([n] \setminus \mathcal{A}_n^{(1)}\right)$$

$$\mathbf{x}_2 = \mathbf{u}_{2,1}G_n\left(\mathcal{A}_n^{(3)}\right) + \mathbf{u}_{2,2}G_n\left(\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}\right) + \mathbf{u}_r G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b}G_b\left([n] \setminus \mathcal{A}_n^{(1)}\right)$$

where $\mathbf{b} \in \mathcal{X}^{n-k}$ is a predetermined and fixed vector.
- The encoding of the codeword $\mathbf{x}_3$ is based on the remaining k_3 information bits, denoted by $\mathbf{u}_3 \in \mathcal{X}^{k_3}$:

$$\mathbf{x}_3 = \mathbf{u}_3 G_n\left(\mathcal{A}_n^{(3)}\right) + (\mathbf{u}_{1,2} + \mathbf{u}_{2,2})G_n\left(\mathcal{A}_n^{(2)} \setminus \mathcal{A}_n^{(3)}\right) + \mathbf{u}_r G_n\left(\mathcal{A}_n^{(1)} \setminus \mathcal{A}_n^{(2)}\right) + \mathbf{b}G_n\left([n] \setminus \mathcal{A}_n^{(1)}\right).$$

The order of decoding the information bits for all possible assignments of codewords over a set of three parallel channels is provided in Table I.

Remark 3 (On the order of successive cancellation). The parallel coding scheme provided in this section is capable to decode sequentially the information bits from each channel, due to the monotonic sequence of index sets $\{\mathcal{A}_n^{(s)}\}_{s \in [S]}$ satisfying the conditions in Corollary 1. It is noted that the index sets in $\mathcal{A}_n^{(s)}$, $s \in [S]$, are 'good' for all the channels $P_{s'}$ where $s' \geq s$. The problem of finding an index set which is 'good' for a set of channels is much harder if the channels are not degraded. This problem is studied in [12] in the context of the compound capacity of polar codes. Upper and lower bounds on the compound capacity of polar codes under successive cancellation decoding are provided in [12]. Although the study in [12] concerns two channels, the techniques are suitable for the case at hand. Specifically, it can be shown that if successive cancellation decoding is performed sequentially as in this section (channel by channel), then the achievable rates are bounded below the channel capacity of the general model (where the parallel channels are not ordered by stochastic degradation). Hence, the parallel progress of the successive cancellation decoders applied in Sections III-A–III-C) is inevitable for the general case.

IV. SUMMARY AND CONCLUSIONS

Capacity-achieving parallel polar coding schemes are provided in this paper for reliable communications over a set of arbitrarily-permuted parallel channels that are binary-input, output-symmetric and memoryless. These schemes are based on the channel polarization method [4], combined with MDS codes of various dimensions. Two coding alternatives are suggested in this paper, one is based on non-binary polar codes (see [8], [9]), and the second is based on binary-interleaved polar codes.

The definition of polar codes includes a set of predetermined and fixed bits, which are crucial to the decoding process. In the original polarization scheme in [4], these predetermined and fixed bits may be chosen arbitrarily (in the case of symmetric channels). For the proposed parallel coding schemes, on the other hand, the predetermined and fixed bits are determined based on some algebraic coding constraints. The MDS coding, suggested in this paper is similar to the rate-matching scheme in [1].

Successive cancellation decoding is applied in both the non-binary and the binary interleaved schemes. The decoding must process in parallel the received observations from all the parallel channels. It is characterized as parallel operations of the successive cancellation decoding procedures provided by a single channel in [4], while exchanging information due to the algebraic constraints imposed by the incorporated MDS codes.

For the particular case of two or three parallel channels, binary channel polarization codes are suitable without relying on interleavers. The same simplification is shown for the particular case of stochastically degraded parallel channels. For the degraded parallel channel model, the decoding may progress in a serial manner, that is the successive cancellation can be carried sequentially channel by channel.

The following topics are suggested for further research:

1) Symmetry condition: For symmetric channels, the predetermined and fixed bits may be chosen arbitrarily. For non-symmetric channels, good predetermined and fixed bits (called also frozen bits in [4]) are shown to exist, but their choice may not be arbitrary. It is an open question if there is a more general construction that does not require the symmetry property of the parallel channels.

2) Generalized parallel polar coding such as in [13]-[15].

3) Generalized channel models: Arbitrarily-permuted parallel channels form just one particularization of the compound setting. It is of interest to enlarge the family of parallel channels for which the studied coding scheme may be applicable. Of specific interest is the case of parallel channels where a sum-rate constraint is provided by the channel model characterization.

4) Studying the impact of improved list based decoding strategies [16] of polar codes on permuted channels.

APPENDIX

A. On the symmetry property of non-binary polarization

Definition 4 (Non-binary symmetry). A DMC which is characterized by a transition probability p, an input-output alphabet $\mathcal{X}$ and a discrete output alphabet $\mathcal{Y}$ is *symmetric* if there exists a function $\mathcal{T} : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$ which satisfies the following properties:

1) For every $x \in \mathcal{X}$, the function $\mathcal{T}(\cdot, x) : \mathcal{Y} \to \mathcal{Y}$ is bijective.
2) For every $x_1, x_2 \in \mathcal{X}$ and $y \in \mathcal{Y}$, the following equality holds:

$$p\big(y|x_1\big) = p\big(\mathcal{T}(y, x_2 - x_1)|x_2\big).$$

Lemma 2 (Message independence property for non-binary symmetric-channel polarization). Let p be a symmetric DMC and $\mathbf{0}$ be the n-length all-zero vector over $\mathcal{X}$. Denote by $P_{\mathrm{e}}(\mathcal{E}_l^d|\mathbf{u})$ the probability of the event $\mathcal{E}_l^d$ in (13), assuming that $\mathbf{w} = \mathbf{u}$ in (5). Then,

$$P_{\mathrm{e}}(\mathcal{E}_l^d(p)|\mathbf{u}) = P_{\mathrm{e}}(\mathcal{E}_l^d(p)|\mathbf{0})$$

for every $\mathbf{u} \in \mathcal{X}^n$ and $l \in [n]$.

Proof: Let $\mathcal{T}$ be the corresponding function in Definition 4. With abuse of notation, the operation of $\mathcal{T}$ on vectors $\mathbf{y} \in \mathcal{Y}^n$ and $\mathbf{x} \in \mathcal{X}^n$ is defined by

$$\mathcal{T}(\mathbf{y}, \mathbf{x}) \triangleq \big(\mathcal{T}(y_1, x_1),\ \mathcal{T}(y_2, x_2),\ \ldots, \mathcal{T}(y_n, x_n)\big).$$

Subtraction of a vector is also defined item-wise, that is $-(x_1, \ldots, x_n) = (-x_1, \ldots x_n)$. Based on the symmetry property of that channel, for every $l \in [n]$, $\mathbf{y} \in \mathcal{Y}^n$, $\mathbf{w} \in \mathcal{X}^{l-1}$, $w \in \mathcal{X}$ and $\mathbf{a} \in \mathcal{X}^n$, we have

$$p_n^{(l)}\big(\mathbf{y}, (w_1, \ldots, w_{l-1})|w_l\big)$$

$$\overset{(a)}{=} \frac{1}{|\mathcal{X}|^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} \prod_{t=1}^{n} p\Big(y_t|\big((w_1, \ldots, w_l, \mathbf{c})G_n\big)_t\Big)$$

$$\overset{(b)}{=} \frac{1}{|\mathcal{X}|^{n-1}} \sum_{\mathbf{c} \in \mathcal{X}^{n-l}} \prod_{t=1}^{n} p\Big(\mathcal{T}\big(y_t, (\mathbf{a}G_n)_t\big)|\big((\mathbf{w}, w_l, \mathbf{c})G_n\big)_t + \big(\mathbf{a}G_n\big)_t\Big)$$

where $(\mathbf{x})_t$ denotes the t-th element of a vector $\mathbf{x} = (x_1, \ldots, x_n)$, (a) follows for memoryless channels from (5) and (6) and (b) follows from the symmetry property of the channel. Consequently, it follows that

$$p_n^{(l)}\big(\mathbf{y}, (w_1, \ldots, w_{l-1})|w_l\big) = p_n^{(l)}\Big(\mathcal{T}\big(\mathbf{y}, \mathbf{a}G_n\big), (w_1, \ldots, w_{l-1}) + (a_1, \ldots, a_{l-1})\big)|w_l + a_l\Big). \tag{18}$$

From (13) and (18) it follows for every pair $(\mathbf{w}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ and every $\mathbf{a} \in \mathcal{X}^n$ that

$$\big(\mathbf{w}, \mathbf{y}\big) \in \mathcal{E}_l^d(p) \iff \big(\mathbf{a} + \mathbf{w}, \mathcal{T}(\mathbf{y}, \mathbf{a} \cdot G_n)\big) \in \mathcal{E}_l^d(p). \tag{19}$$

Next, let $1_{\mathcal{E}_l^d(p)}(\mathbf{u}, \mathbf{y})$ denote the indicator of the event $\mathcal{E}_l^d(p)$. For every $\mathbf{u} \in \mathcal{X}^n$ it follows that

$$
\begin{aligned}
& P_{\mathrm{e}}(\mathcal{E}_l^d(p)|\mathbf{u}) \\
&= \sum_{\mathbf{y} \in \mathcal{Y}^n} p_n(\mathbf{y}|\mathbf{u}) 1_{\mathcal{E}_l^d(p)}(\mathbf{u}, \mathbf{y}) \\
&\stackrel{(a)}{=} \sum_{\mathbf{y} \in \mathcal{Y}^n} p(\mathbf{y}|\mathbf{u}G_n) 1_{\mathcal{E}_l^d(p)}(\mathbf{u}, \mathbf{y}) \\
&\stackrel{(b)}{=} \sum_{\mathbf{y} \in \mathcal{Y}^n} p(\mathcal{T}(\mathbf{y}, -\mathbf{u}G_n)|\mathbf{0}) 1_{\mathcal{E}_l^d(p)}(\mathbf{0}, \mathcal{T}(\mathbf{y}, -\mathbf{u}G_n)) \\
&= \sum_{\mathbf{y} \in \mathcal{Y}^n} p_n(\mathbf{y}|\mathbf{0}) 1_{\mathcal{E}_l^d(p)}(\mathbf{0}, \mathbf{y}) \\
&= P_{\mathrm{e}}(\mathcal{E}_l^d(p)|\mathbf{0})
\end{aligned}
$$

where (a) follows from (5), (b) follows from (19) by plugging $\mathbf{a} = \mathbf{u}$, and (c) follows since $\mathcal{T}(y, x)$ is a bijective function of $y \in \mathcal{Y}$ for every fixed symbol $x \in \mathcal{X}$. ∎

B. Stochastically degraded parallel channels

Definition 5 (Stochastically degraded channels). Consider two memoryless channels with a common input alphabet $\mathcal{X}$, transition probability functions P_1 and P_2, and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively. The channel P_2 is a stochastically degraded version of channel P_1 if there exists a channel D with an input alphabet $\mathcal{Y}_1$ and an output alphabet $\mathcal{Y}_2$ such that

$$
P_2(y_2|x) = \sum_{y_1 \in \mathcal{Y}_1} P_1(y_1|x) D(y_2|y_1), \quad \forall x \in \mathcal{X}, y_2 \in \mathcal{Y}_2.
$$

Lemma 3 (On the degradation of split channels). Let P_1 and P_2 be two transition probability functions with a common binary input alphabet $\mathcal{X} = \{0, 1\}$ and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively. For a block length n, the split channels of P_1 and P_2 are denoted by $P_{1,n}^{(l)}$ and $P_{2,n}^{(l)}$, respectively, for all $l \in [n]$. Assume that the channel P_2 is a stochastically degraded version of channel P_1. Then, for every $l \in [n]$ the split channel $P_{2,n}^{(l)}$ is a stochastically degraded version of the split channel $P_{1,n}^{(l)}$.

Proof: The proof follows by induction (see [10], [11]). ∎

Definition 6 (Stochastically degraded parallel channels). Let $\{P_s\}_{s=1}^S$ be a set of S parallel memoryless channels. The channels $\{P_s\}_{s=1}^S$ are stochastically degraded if there exists a sequence of unique indices $s_1, s_2, \ldots s_S$, $s_i \in [S]$ for every $i \in [S]$, such that the channel $P_{s_{i+1}}$ is a stochastically degraded version of P_{s_i} for every $i \in [S-1]$.

Proof of Corollary 1: From [5], it follows that there exists a sequence of sets $\{\mathcal{A}_n^{(S)}\}$ satisfying $|\mathcal{A}_n^{(s)}| \geq nR_s$ and $\Pr(\mathcal{E}_l(P_s)) \leq 2^{-n^\beta}$. These are the rate and performance properties in (15) and (17) for the particular case of $s = S$. Fix an $s' \in [S-1]$ and assume that the set sequences $\{\mathcal{A}_n^{(s)}\}$, $s > s'$, can be chosen such the following properties are met:

1) The rate and performance properties in (15) and (17) are satisfied for every $s > s'$.
2) For every block length n, the sets $\mathcal{A}_n^{(s)}$, $s > s'$, are monotonic:

$$
\mathcal{A}_n^{(S)} \subseteq \mathcal{A}_n^{(S-1)} \subseteq \cdots \subseteq \mathcal{A}_n^{(s'+1)}. \tag{20}
$$

An information index set $\{\mathcal{A}_n^{(s')}\}$ is next constructed. From [5], it follows that there exists a sequence of sets $\{\mathcal{A}_n\}$ satisfying $|\mathcal{A}_n| \geq nR_{s'}$ and $\Pr(\mathcal{E}_l(P_{s'})) \leq 2^{-n^\beta}$. These are the rate and performance properties in (15) and (17)

for the particular case of $s = s'$. Choose an arbitrary index $l \in \mathcal{A}_n^{(s'+1)}$. Since $P_{s'+1}$ is a degraded version of $P_{s'}$, then according to Lemma 3, the split channel $P_{s'+1,n}^{(l)}$ is a degraded version of the split channel $P_{s',n}^{(l)}$. It is clearly suboptimal to first degrade the observation vector $\mathbf{y} \in \mathcal{Y}_{s'}$ to create a vector $\tilde{\mathbf{y}} \in \mathcal{Y}_{s'+1}$, and only then decode the corresponding information bit. Consequently, $\mathcal{E}_l(p) \subseteq \mathcal{E}_{l+1}(p)$ which implies that if $P_{s'+1,n}^{(l)}$ satisfies (17), so does $P_{s',n}^{(l)}$. It follows that l is a valid index for $\mathcal{A}_n^{(s')}$ in terms of the performance property. That is, for every $l \in \mathcal{A}_n^{(s'+1)}$ it follows that $\Pr\big(\mathcal{E}_l(P_{s'})\big) \leq 2^{-n^\beta}$. Therefor we set $\mathcal{A}_n^{(s')} = \mathcal{A}_n^{(s'+1)} \cup \mathcal{A}_n$. As a results, the set sequences $\{A_n^s\}$, $s \geq s'$ satisfy the following properties:

1) The rate and performance properties in (15) and (17) are satisfied for every $s \geq s'$
2) Monotonicity property: $\mathcal{A}_n^{(S)} \subseteq \mathcal{A}_n^{(S-1)} \subseteq \cdots \subseteq \mathcal{A}_n^{(s'+1)} \subseteq \mathcal{A}_n^{(s')}$.

The proof follows by induction. ∎

REFERENCES

[1] F. M. J. Willems and A. Gorokhov, "Signaling over arbitrarily permuted parallel channels," *IEEE Trans. on Information Theory*, vol. 54, no. 3, pp. 1374–1382, March 2008.

[2] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2148–2177, October 1998.

[3] A. Hitron, A. Khina and U. Erez, "Transmission over arbitrarily permuted parallel Gaussian channels," *Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT 2012)*, pp. 2671–2675, Boston, MA, USA, July 2012.

[4] E. Arikan, "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.

[5] E. Arikan and E. Telatar, "On the rate of channel polarization," *Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, pp. 1493–1495, Seoul, South Korea, June 2009.

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

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

[8] R. Mori and T. Tanaka, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernels," *Proceedings of the 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, pp. 894–898, Austin, Texas, June 2010.

[9] E. Sasoglu, E. Telatar and E. Arikan, "Polarization for arbitrary discrete memoryless channels," *Proceedings of the 2009 IEEE Information Theory Workshop (ITW 2009)*, pp. 144–148, Taormina, Sicily, October 2009.

[10] S. B. Korada, *Polar Codes for Channel and Source Coding*, Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2009.

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

[12] S. H. Hassani, S. B. Korada and R. Urbanke, "The compound capacity of polar codes," *Proceedings of the Forty-Seventh Annual Allerton Conference on Communication, Control and Computing*, pp. 16–21, Allerton, Monticello, Illinois, USA, September 2009.

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

[14] S. B. Korada and E. Sasoglu, "A class of transformations that polarize binary-input memoryless channels ," *Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT 2009)*, pp. 1478–1482, Seoul, South Korea, June 2009.

[15] E. Arikan and G. Markarian, "Two-dimensional polar coding," *Proceedings of the 10th International Symposium of Communication Theory and Applications (ISCTA 2009)*, Ambleside, UK, July 2009.

[16] I. Tal and A. Vardy, "List decoding of polar codes," May 2012. [Online]. Available: http://arxiv.org/abs/1206.0050.