Capacity-Achieving Polar Codes for Arbitrarily Permutable Parallel Channels

Eran Hof, Student Member, IEEE, Igal Sason, Senior Member, IEEE, Shlomo Shamai (Shitz), Fellow, IEEE, and Chao Tian, Senior Member, IEEE

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

Index Terms—Capacity-achieving codes, maximum distance separable (MDS) codes, parallel channels, polar codes, stochastically degraded channels.

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 blocklength, 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 \( \pi \) of the set \( \{1, \ldots, S\} \) where code-sequence number \( s \in \{1, \ldots, S\} \) is transmitted over the parallel channel number \( r = \pi(s) \). 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 in the review paper [2] and references therein.

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, Th. 1] that the capacity of the system is equal to the sum of the capacities of the parallel channels. Furthermore, Willems and Gorokhov [1] also address 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, Th. 2]).

Arbitrarily permuted parallel channels may be of interest when analyzing, e.g., networking applications, orthogonal frequency division multiplexing, and bit-interleaved coded modulation 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 supersymbol. 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 could 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.

Manuscript received July 29, 2010; revised August 14, 2012; accepted November 11, 2012. Date of publication January 04, 2013; date of current version February 12, 2013. This work was supported in part by the Israel Science Foundation under Grant 1070/07 and in part by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++. This paper was presented in part at the 2010 IEEE Information Theory Workshop, Dublin, Ireland, and in part at the IEEE 26th Convention of Electrical and Electronics Engineers, Eliat, Israel.

E. Hof was with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. He is now with Samsung Semiconductor Israel R&D Center, Ramat Gan 52522, Israel (e-mail: eran.hof@gmail.com).

I. Sason and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. He is now with Samsung Semiconductor Israel R&D Center, Ramat Gan 52522, Israel (e-mail: sason@ee.technion.ac.il; shlomo@ee.technion.ac.il).

C. Tian is with AT&T Labs Research, Florham Park, NJ 07932 USA (e-mail: tian@research.att.com).

Communicated by U. Erez, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2012.2236971
In this paper, we consider the construction of polar codes as channel codes for arbitrarily permutated 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 permutated 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 permutated parallel channels. The construction of the rate-matching codes in [1, Sec. 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] needs 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.

The rest of this 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 paper.

II. PRELIMINARIES

A. Arbitrarily Permutated Parallel Channels

Consider the communication model depicted in Fig. 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 \( y_s, x \in \mathcal{Y}_s, \mathcal{X} \), respectively. The encoding operation maps the message \( x_m \) into a set of \( S \) codewords \( \{x_s \in \mathcal{X}^n\}_{s=1}^S \). 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 \( \pi \) is fixed during the entire block transmission). The permutation \( \pi \) 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.

![Fig. 1. Communication over an arbitrarily permutated parallel channel with \( S = 3 \) in this example (taken from [1]).](image)

Formally, the channel is defined by the following family of transition probabilities:

\[
\{P(Y|X; \pi) : Y \in \{\mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_S\}^n, \quad X \in \mathcal{X}^{n \times n}, \pi : |S| \to |S|\}_{n \to 1}
\]

where \( Y = (y_1, y_2, \ldots, y_S) \) are the transmitted codewords, \( \pi : [S] \to [S] \) is the arbitrary permutation mapping of codewords to channels. The decoder produces the estimated message \( \hat{x}_m \) based on the received vectors \( Y \) and the permutation \( \pi \). The case where the decoded message \( \hat{x}_m \) is different from the transmitted message, \( \hat{x}_m \neq x_m \), 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 blocklength \( n \)

\[
\frac{1}{n} \log_2 M \geq R - \varepsilon
\]

\[
P_{e}(\pi) \leq \delta, \text{ for all } S! \text{ permutations } \pi : [S] \to [S]
\]

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

The capacity \( C_{II} \) of this channel model can be derived as a particular case of the compound channel (see, e.g., the review paper [2] and reference therein). Specifically, if there exists an input distribution that achieves capacity for all the parallel channels, then the capacity \( C_{II} \) is given by

\[
C_{II} = \sum_{s=1}^{S} C_s
\]

where \( C_s \) is the capacity of the \( s \)th channel. 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 \( \pi \) provides
a different product channel. Consequently, there are \( S' \) possible product channels. A properly chosen random code was shown to achieve the capacity \( C_{\Pi} \) with decoding by joint typicality for all possible permutations \( \pi \).

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, Sec. 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 nonbinary 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 we have the following.

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

\[
p(y \mid 0) = p(\mathcal{T}(y) \mid 1), \quad \forall y \in \mathcal{Y}.
\]

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, an \( n \times n \) matrix \( G_n \), where \( n = 2^i \), is defined. The matrix \( G_n \) is referred to as the polar generator matrix of size \( n \).

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

The encoding of \( A_n \) information bits to a codeword \( x \in \{0, 1\}^n \) is carried out in two steps. First, a binary \( n \)-length vector \( w \) is defined. Over the indices specified by \( A_n \), the components of \( w \) are set according to the information bits. The rest of the \( |A^n_n| \) bits of \( w \) are predetermined and fixed according to a particular code design (these bits are denoted as “frozen bits” in [4]). Next, a codeword is evaluated according to

\[
x = wG_n. \tag{4}
\]

1These codes can be shown to be coset codes.

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

\[
p_n(y \mid w) = p(y \mid wG_n), \quad \forall y \in \mathcal{Y}^n \text{ and } w \in \mathcal{X}^n. \tag{5}
\]

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

\[
p_n^{(i)}(y, x) = \frac{1}{\mathcal{X}^{i-1}} \sum_{c \in \mathcal{X}^{i-1}} p_n(y \mid [w, x, c]) \tag{6}
\]

where \( y \in \mathcal{Y}^n \), \( w \in \mathcal{X}^{i-1} \), 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

\[
f_{\text{dec}}(p_n^{(i)}, y, w) = \arg \max_{x \in \mathcal{X}} p_n^{(i)}(y, w, x) \tag{7}
\]

where \( p_n^{(i)} \) is a split channel defined in (6), \( y \in \mathcal{Y}^n \), \( w \in \mathcal{X}^{i-1} \), and \( x \in \mathcal{X} \). 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 \( i \in [n] \). If \( i \in A_n \), then the bit \( w_i \) is a predetermined and known bit. Otherwise, we decode the bit \( w_i \) according to \( f_n(p, i, y, (w_1, \ldots, w_{i-1})) \) where \( y \) is the received vector and \( w_1, \ldots, w_{i-1} \) are the already decoded bits. It is shown in [4] that the described successive cancellation decoding procedure may be accomplished with a complexity of \( O(n \log n) \).

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 \( A_n \) such that we have the following.

1) Rate: \( |A_n| \geq nR. \)

2) Performance: Assume that the information bits \( w_t, t \in A_n \), are chosen in a uniform manner over all possible options in \( \{0, 1\} - A_n \) and fix an arbitrary choice of the predetermined and fixed bits \( w_{-t}, t \in A^n_n \). For every index \( t \in A_n \), the following upper bound is satisfied:

\[
Pr\{E_t(p)\} \leq 2^{-n^3}
\]

where

\[
E_t(p) \triangleq \left\{ p_n^{(i)}(y, (w_1, w_2, \ldots, w_{i-1}) \mid w_t) \leq p_n^{(i)}(y, (w_1, w_2, \ldots, w_{i-1}) \mid w_t + 1) \right\} \tag{8}
\]

and the addition \( w_t + 1 \) 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 \( w_t, t \in \mathcal{A}_c \). In the general case where the parallel channels are not necessarily output-symmetric, this vector cannot 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 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 \text{GF}\((2^m)\) is carried according to (4) where the operations are carried over the finite field \( \text{GF}\((2^m)\), \( \mathbf{w} = \{w_1, \ldots, w_n\} \in \text{GF}\((2^m)\), the symbol \( w_1 \) is an information symbol for every \( l \in \mathcal{A}_n \), and it is predetermined and fixed for every \( l \not\in \mathcal{A}_n \). 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 \( C \) whose minimum distance is \( d \) is called an 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 blocklength \( S \) and a dimension \( k \in [S] \). Let \( S > 0 \) be an integer number, and fix an integer \( m \) such that \( 2^m - 1 > S \). For every \( k \in [2^m - 1] \), there exists a \([2^m - 1, k]\) Reed-Solomon (RS) code over the Galois field \( \text{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 \( \text{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 \( \text{GF}\((2^m)\).

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

III. 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 nonbinary MDS codes. Two alternative schemes are therefore proposed: a scheme based on nonbinary 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 nonbinary MDS codes and binary (noninterleaved) polar codes.

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 \( k_n = |\mathcal{A}_n^{(1)}| + |\mathcal{A}_n^{(2)}| + |\mathcal{A}_n^{(3)}| \). The polar encoding is preceded by mapping \( k_n \) information bits to three length-\(n\) binary vectors

\[
\mathbf{w}_s = \{w_{s,1}, w_{s,2}, \ldots, w_{s,n}\} \in \{0, 1\}^n, \quad s = 1, 2, 3,
\]

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 of the two information bits.

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

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

where \( G_n \) is the generator matrix of the polar code.

The codeword \( \mathbf{x}_s(s) \) is then transmitted over the symmetric channel \( P_s \) (see Definition 2), \( s \in [3] \), as depicted in Fig. 1. The split channels defined in (6) are therefore evaluated with respect to the permuted indices of the transmitted vectors as well. Specifically, let \( \mathbf{y}_s \) denote 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 \( \mathbf{w}_s(s), s \in [3] \). 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

\[
P_{\pi(s),l}(\mathbf{y}_s, w_{\pi(s),1}, \ldots, w_{\pi(s),l-1}, w_l, c|G_n) \quad (9)
\]

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

The \( i \)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

\[
w_{\pi(s),l} = f_{\text{dec}}(P_{\pi(s),l}(\mathbf{y}_s, \{w_{\pi(s),1}, w_{\pi(s),2}, \ldots, w_{\pi(s),l-1}\})) \quad (10)
\]

where \( f_{\text{dec}} \) is the decoding rule in (7), and \( P_{\pi(s),l}, s \in [3] \), are the split channels in (9).
2) Otherwise, if \( l \in \mathcal{A}_n^{(s)} \) and \( l \in \mathcal{A}_n^{(s')} \) for some \( 1 < s < s' \leq 3 \), then decode \( w_{\pi(s),i} \) as in (10) and \( w_{\pi(s'),i} \) as in (10) with \( s \) replaced by \( s' \). Furthermore, set the remaining bit \( w_{\pi(s'),i} \) (where \( s' \neq s \)) to \( w_{\pi(s),i} + w_{\pi(s'),i} \).

3) Otherwise, if \( l \in \mathcal{A}_n^{(s)} \) for a single \( s \in \{3\} \), decode \( w_{\pi(s),i} \) as in (10). Then, set the remaining two bits \( w_{\pi(s),i} \) to \( 0 \).

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 \( C_H \) 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 \( 0 < \beta < \frac{1}{3} \). The error probability \( P_e \) of the provided decoding procedure is upper bounded, via the union bound, by

\[
P_e \leq \sum_{s \in \mathcal{A}} \sum_{i \in \mathcal{A}_n^{(s)}} \Pr \{ E_i P_s \} \quad (11)
\]

where \( E_i P_s \) 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 ensuring that for symmetric channels (see Remark 1), the decoding error probability \( P_e \) in (11) is upper bounded by

\[
P_e \leq n2^{-n^\beta}.
\]

Taking the blocklength \( n \) large enough concludes the proof. \( \blacksquare \)

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 nonbinary polar codes or an interleaving technique.

**B. Coding for \( S > 3 \) Using Nonbinary Polar Codes**

For \( S > 3 \), the binary MDS codes applied in Section III-A must be replaced by MDS codes of blocklength \( S \). The only binary MDS codes are the trivial codes (repetition, SPC, 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 code over the Galois field \( \text{GF}(2^m) \) is chosen, which is denoted by \( C_k \) (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 \( C_0 \).

In order to apply the nonbinary polarization coding scheme, a new set of parallel channels \( \{W_s\}_{s=1}^m \) is defined according to

\[
W_s(y|x) = \prod_{i=1}^m P_s(y_i|b_i)
\]

where \( y = (y_1, \ldots, y_m) \in \mathcal{Y}_s, \ x \in \text{GF}(2^m), \ s \in \{S\}, \ \{b_1(x), \ldots, b_m(x)\} \) is the binary \( m \)-length vector representation of the symbol \( x \in \text{GF}(2^m) \); 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 \( n \) transmissions over the channel \( P_s, s \in \{S\} \). With some abuse of notations, the information index set sequence for each of the nonbinary channels \( W_s, s \in \{S\} \) is also denoted by \( \mathcal{A}_n^{(s)} \). For every \( l \in \{n\} \), define

\[
k_l \Delta \{|s : l \in \mathcal{A}_n^{(s)}\} \quad (12)
\]

The encoding of the parallel nonbinary polarization scheme is carried as follows.

1) For every channel index \( s \in \{S\} \) and every information index \( i \in \mathcal{A}_n^{(s)} \), denote by \( a_i^{(s)} \) the symbol in \( \text{GF}(2^m) \) corresponding to \( n \) information bits.

2) For every \( l \in \{n\} \), choose the unique codeword \( e^{(l)} = (c_1^{(l)}, c_2^{(l)}, \ldots, c_S^{(l)}) \in C_k \), satisfying \( c_s^{(l)} = a_i^{(s)} \) for every \( s' \in \{s : l \in \mathcal{A}_n^{(s')}\} \).

3) Compute \( S \) polar codewords \( x_s, s \in \{S\} \), according to

\[
x_s = (c_1^{(s)}, c_2^{(s)}, \ldots, c_S^{(s)}) \cdot G_n
\]

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

The decoding procedure is further detailed in Fig. 2 via an illustrative example. The codeword \( x_s, s \in \{S\} \), is transmitted over the channel \( W_s \), and let \( y_s \) 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 \( l \in \mathcal{A}_n^{(s)} \), let \( c_i^{(l)} = e^{(l)} = f_{cc}(W_{s,n}^{(l)}, y_s, c_1^{(s)}, \ldots, c_1^{(s)}, c_2^{(s)}, \ldots, c_2^{(s)}, \ldots, c_{l-1}^{(s)}).

2) Find the unique codeword \( e = (c_1, c_2, \ldots, c_S) \in C_k \), satisfying \( c_s^{(l)} = c_i^{(s)} \) for every \( s' \in \{s : l \in \mathcal{A}_n^{(s')}\} \).

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

**Proposition 2:** The parallel nonbinary polar coding scheme achieves the capacity of the considered model.

**Proof:** The ability to choose a unique codeword in \( C_k \), \( l \in \{n\} \), follows directly from the fact that an \((S,k)\) MDS code
Fig. 2. Illustration of the nonbinary parallel polar encoding procedure in the particular case of $\mathcal{S} = 4$. The grid of rectangles illustrates the symbols $c_i^{(s)}$, for $1 \leq i \leq 8$ and $s \in [4]$, as defined in the encoding procedure. Each row of squares represents the vector $(r_1^{(s)}, r_2^{(s)}, \ldots, r_8^{(s)})$, where each of the squares represents a symbol. A filled square represents a symbol $c_i^{(s)}$ for which $i \in \mathcal{A}_n^{(s)}$, $s \in [4]$. For the depicted grid, $\mathcal{A}_n^{(1)} \cap [8] = \{2, 3, 4, 7, 8\}$, $\mathcal{A}_n^{(2)} \cap [8] = \{2, 6, 7, 8\}$, $\mathcal{A}_n^{(3)} \cap [8] = \{3, 5, 7, 8\}$, and $\mathcal{A}_n^{(4)} \cap [8] = \{2, 3, 5, 6, 7, 8\}$. 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 $i \notin \mathcal{A}_n^{(s)}$, $s \in [4]$. The symbols represented by the empty squares are determined such that each column forms an MDS codeword. The four vertical rectangles mark four of these codewords: $c_1^{(1)}$, $c_4^{(3)}$, and $c_6^{(1)}$. Codewords $c_1^{(1)}$ and $c_3^{(1)}$ belong to codes of dimensions 0 and 4, respectively (a constant vector and the whole space). Accordingly, in $c_1^{(1)}$, all four squares are empty to represent four predetermined and fixed symbols, while in $c_3^{(1)}$, all four squares are filled squares, representing four arbitrary information symbols (an arbitrary vector in the whole space). The codeword $c_3^{(1)}$ is represented by a code of dimension 1, accordingly $c_3^{(1)} = c_3^{(1)} = c_3^{(1)}$ (the empty squares equal to the value of the single filled square). The codeword $c_3^{(1)}$ belongs to a code of dimension 3 where the three filled squares completely determine the value of the single empty square according to $c_3^{(1)} = -c_3^{(1)} = -c_3^{(1)}$. 

can correct up to $S - k_1$ erasures. For a DMC $W$ with an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, define the events

$$E_d^d(W) = \begin{cases} (w, y) \in \mathcal{X}^n \times \mathcal{Y}^m : \\ W_n^{(1)}(y, (w_1, \ldots, w_{n-1}) w_t) \\ \leq W_n^{(1)}(y, (w_1, \ldots, w_{n-1}) w_t + d) \end{cases}, \quad (13)$$

$l \in [n], d \in \mathcal{X}$

where $W_n^{(1)}$, $l \in [n]$ are the split channels of $W$. The error probability $P_e$ for the nonbinary decoding procedure is upper bounded by

$$P_e = \sum_{s \in [S]} \sum_{i \in \mathcal{X}^n} \sum_{i \in \mathcal{X}^n} \Pr(E_d^d(W_s)). \quad (14)$$

It follows from [8] and [9] that the probability of the event $E_d^d(W_s)$ can be made exponentially low as the blocklength increases 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] 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 nonbinary polar codes in Section A in the Appendix. The symmetry of $W_s$, $s \in [S]$ according to Definition 4 for nonbinary 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 Nonbinary 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 nonbinary case is stated in Definition 4 in Section A of the Appendix).

C. 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 $\zeta_k$ denote an MDS code over $GF(2^{m_k})$ of dimension $k$, and let $k_1$ be defined as in (12). The encoding process is carried as follows.

1) For every information index $i \in \mathcal{A}_n^{(s)}$, and every channel index $s \in [S]$: pick $m$ information bits, denoted by $u_i^{(s)}$, $l_i^{(s)} \in \mathcal{A}_{n_1}^{(s)}$. The grid of rectangles refers to Fig. 2 with the difference that, due to the transmission permutation, codeword $x_i$ is transmitted over channel $P_s$, codeword $x_i$ is transmitted over channel $P_s$, and codewords $x_i$ and $x_i$ are transmitted over channels $P_s$ and $P_s$, respectively. All the symbols in $c_i^{(s)}$ (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., $c_i^{(1)}$, note that $4 \notin \mathcal{A}_n^{(1)}$). As another consequence of the channel permutation, some information symbols cannot be decoded via (7). Such information symbols are represented by filled and rotated squares (e.g., $c_i^{(3)}$, note that $4 \notin \mathcal{A}_n^{(1)}$). Consequently, at the fourth decoding stage, even though the message symbol is $c_i^{(1)}$ (represented by a filled rotated square), due to the transmission permutation, only the symbol $c_i^{(1)}$ 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 $c_i^{(3)}$ is the rotated square and $c_i^{(4)}$ is now $x$-marked (as $6 \notin \mathcal{A}_n^{(1)}$ and $6 \notin \mathcal{A}_n^{(4)}$). The message symbols (filled squares) $c_i^{(3)}$ and $c_i^{(4)}$ can be decoded via (7) but $c_i^{(3)}$ cannot. Nevertheless, the nonmessage symbol (filled $x$-marked square) $c_i^{(3)}$ is decoded via (7) and due to the MDS structure of columns, the message symbol (rotated square) $c_i^{(3)} = -c_i^{(3)} = -c_i^{(3)}$. All the symbols (filled squares) in $c_i^{(3)}$ are decoded via (7) as $\beta \in \mathcal{A}_n^{(1)}$ for every $s \in [s]$. 

Fig. 3. Illustration of the nonbinary parallel polar decoding procedure in the particular case of $\mathcal{A} = 8$ and $\mathcal{S} = 4$. The grid of rectangles refers to Fig. 2 with the difference that, due to the transmission permutation, codeword $x_i$ is transmitted over channel $P_s$, codeword $x_i$ is transmitted over channel $P_s$, and codewords $x_i$ and $x_i$ are transmitted over channels $P_s$ and $P_s$, respectively. All the symbols in $c_i^{(s)}$ (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., $c_i^{(1)}$, note that $4 \notin \mathcal{A}_n^{(1)}$). As another consequence of the channel permutation, some information symbols cannot be decoded via (7). Such information symbols are represented by filled and rotated squares (e.g., $c_i^{(3)}$, note that $4 \notin \mathcal{A}_n^{(1)}$). Consequently, at the fourth decoding stage, even though the message symbol is $c_i^{(1)}$ (represented by a filled rotated square), due to the transmission permutation, only the symbol $c_i^{(1)}$ 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 $c_i^{(3)}$ is the rotated square and $c_i^{(4)}$ is now $x$-marked (as $6 \notin \mathcal{A}_n^{(1)}$ and $6 \notin \mathcal{A}_n^{(4)}$). The message symbols (filled squares) $c_i^{(3)}$ and $c_i^{(4)}$ can be decoded via (7) but $c_i^{(3)}$ cannot. Nevertheless, the nonmessage symbol (filled $x$-marked square) $c_i^{(3)}$ is decoded via (7) and due to the MDS structure of columns, the message symbol (rotated square) $c_i^{(3)} = -c_i^{(3)} = -c_i^{(3)}$. All the symbols (filled squares) in $c_i^{(3)}$ are decoded via (7) as $\beta \in \mathcal{A}_n^{(1)}$ for every $s \in [s]$. 

Remark 2 (Coding for Nonbinary 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 nonbinary case is stated in Definition 4 in Section A of the Appendix).
For every $i \in [n]$, choose the unique codeword $c^{(i)} = (c_1^{(i)}, c_2^{(i)}, \ldots, c_2^{(i)}) \in C_{k_i}$ for which the binary representation of $c^{(i)}_s \in \mathbb{GF}(2^{m})$ is equal to

$$(u^{(s)}_{(i-1)m+1}, \ldots, u^{(s)}_{(i-1)m+m})$$

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

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

$$\left(\left(u^{(s)}_{(i-1)m+1}, \ldots, u^{(s)}_{(i-1)m+m}\right) \in \{0, 1\}^m \right.$$

as the binary representation of the symbol $c^{(s)}_s$.

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

$$x_{g, s} = (u^{(s)}_{g+m}, u^{(s)}_{g+m+2}, \ldots, u^{(s)}_{(n-1)m+g+q}) \in G_m$$

and $G_m$ is the binary polar generator matrix.

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

$$x^{(s)} = (x_{1,s}, x_{2,s}, \ldots, x_{m,s}).$$

The concatenated codeword $x^{(s)}$ is transmitted over the channel $P_s$, $s \in [S]$, and let $y^{(s)} = (y^{(s)}_1, \ldots, y^{(s)}_m)$ denote the received vector at the output of this channel.

Assuming that the bits $u^{(s)}_{g+m}$ and $u^{(s)}_{g+m+2}$ are already decoded, the bits $u^{(s)}_{g+m}$ are decoded sequentially at the $k$th stage (for $k = 1, \ldots, n$) as follows.

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

$$u^{(s)}_{l-1+m+g} = f_{dec}(P_s, y_1, y_2, \ldots, y_m)$$

where $f_{dec}$ and $P^{(i)}_s$ are defined in (7) and (9), respectively.

2) Find the unique codeword $c = (c_1, c_2, \ldots, c_S) \in C_{k}$ for which the symbol $c^{(s)}_s$ is equal to

$$\left(\left(u^{(s)}_{(i-1)m+1}, \ldots, u^{(s)}_{(i-1)m+m}\right) \in \{0, 1\}^m \right.$$

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

3) For every $s \in [S]$ for which $l \notin A^{(s)}_s$, the bits $\left(\left(u^{(s)}_{(i-1)m+1}, \ldots, u^{(s)}_{(i-1)m+m}\right) \in \{0, 1\}^m \right.$ are set according to the binary representation of the symbol $c^{(s)}_s \in \mathbb{GF}(2^m)$.

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 $C_{k_i}$, $i \in [n]$, follows directly from the fact that an $(S, k_i)$ MDS code can correct up to $S - k_i$ erasures. For every channel $s \in [S]$, $m$-interleaved polar codes of blocklength $n$ are applied. Hence, the code rate $R_n$ of the parallel binary-interleaved polar scheme is given by

$$R_n = \sum_{s=1}^{S} \frac{m |A^{(s)}_s|}{mn} = \frac{1}{n} \sum_{s=1}^{S} A^{(s)}_s.$$

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 to 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 nonbinary 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 $C_n$ be the capacity of the channel $P_n$, $s \in [S]$. Fix $0 < \beta < \frac{1}{2}$ and a set of rates $R_1, \ldots, R_k$ such that $0 \leq R_s \leq C_s$ for every $s \in [S]$. Then, there exists a sequence of information sets $A^{(s)}_n \subseteq [n], s \in [S]$ and $n = 2^k$ where $i \in \mathbb{N}$, satisfying the following properties.

1) Rate

$$A^{(s)}_n \geq nR_s, \forall s \in [S].$$

2) Monotonicity

$$A^{(s)}_n \subseteq A^{(s+1)}_n \subseteq \cdots \subseteq A^{(k)}_n.$$

3) Performance

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

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

$$\mathcal{E}_l(p_s) = \left\{ p^{(l)}_n(y, w^{(l-1)}_l | w_l) \leq p^{(l)}_n(y, w^{(l-1)}_l | w_l + 1) \right\}$$

for all $l \in [n]$.

Proof: See Section B in the Appendix.
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 $\{\mathbf{u}_{s,1}\}, s, l \in [S]$ where the vector $\mathbf{u}_{s,1}$ is of length $k_n^{(S-l+2)} - k_n^{(S-l+1)}$ bits. The vectors $\mathbf{u}_{s,1}, s \in S'$ and $\mathbf{u}_{s,2} = (u_{s,2}(1), u_{s,2}(2), \ldots, u_{s,2}(k_{S-1} - k_S)), s \in [S-1]$, are set to information bits. Next, the vector $\mathbf{u}_{S,2}$ is determined (the following steps are accompanied with the illustration in Fig. 4).

1) Construct the $(S-1) \times K_{S-1}$ matrix over $GF(2^m)$, $C^{(2)}$, 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

$$
\begin{align*}
&\{u_{i,2}((j-1)m+1), u_{i,2}((j-1)m+2), \ldots, u_{i,2}(jm)\}, \ i \in [S-1], \ j \in [K_{S-1}]
\end{align*}
$$

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

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

3) A $K_{S-1}$-length vector $\mathbf{u}_{S,2}$ over $GF(2^m)$ 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 Fig. 4).

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

Let $2 < l \leq S$, and assume that the vectors $\mathbf{u}_{s,l'}, s \in [S], l' < l$, are already defined. The vectors $\mathbf{u}_{s,1}, s \in [S]$ are defined as follows.

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

2) Construct the $(S - (l - 1)) \times K_{S-(l-1)}$ matrix $C^{(l)}$ over $GF(2^m)$ from the row vectors in step 1, where the $(i,j)$ element of $C^{(l)}$ is defined by the $m$ bits

$$
\begin{align*}
&\{u_{i,1}((j-1)m+1), u_{i,1}((j-1)m+2), \ldots, u_{i,1}(jm)\}, \ i \in [S-(l-1)], \ j \in [K_{S-(l-1)}]
\end{align*}
$$

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

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

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

$$
\begin{align*}
\mathbf{x}_s &= \sum_{l=1}^{S} \mathbf{u}_{s,1} G_n \left( A_n^{(S-(l-1))} \setminus A_n^{(S-(l-2))} \right) \\
&\quad + b G_n \left( [n] \setminus A_n^{(1)} \right), \ s \in [S].
\end{align*}
$$

Here, $A_n^{(s+1)} = b$, $b$ is a binary predetermined and fixed vector, and $G_n$ is the polar generator matrix.

The decoding process starts with the observations received from channel $P_1$ (see Fig. 4). A polar successive cancellation decoding, with respect to the information index set $A_1^{(1)}$, is applied to the received vector. This allows the decoding of the vectors $\mathbf{u}_{n^{(1)},l}, l \in [S]$. Next, the decoding proceeds to successive cancellation decoding procedure for the vector received at the output of channel $P_2$ (i.e., the channel with the second largest capacity). This decoding procedure is capable of decoding $A_2^{(1)}$ bits based on $n - A_2^{(2)}$ predetermined and fixed bits. For the current decoding procedure, $n - |A_1^{(1)}|$ of these bits are the predetermined and fixed bits in $b$. The rest of $A_1^{(1)} - |A_2^{(1)}|$ bits are based on the bits decoded at the previous decoding stage. Specifically, the bit vector $\mathbf{u}_{n^{(1)},l}, s \in [S]$, is fully determined. Moreover, based on the codewords $\mathbf{c}_j, j \in [K_{1,2}]$, the vectors $\mathbf{u}_{n^{(1)},l}, s \in [S]$ are determined for all $s \geq 2$ as well.

Next, the remaining $S - 2$ decoding stages are followed. Note that after the $(S - 1)$th decoding stage, where $2 < s < S$, the vectors $\mathbf{u}_{n^{(s)},l}, l$ for either $1 \leq l < s$ and $l \in [S]$, or $l > s$ 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 $\mathbf{u}_{n^{(s)},l}$ for all $l \in [S]$ and the vectors $\mathbf{u}_{n^{(s)},l}, s \in [S]$ for all $s' \geq s$. In order to apply the polar successive cancellation decoding procedure to the vector received over the channel $P_s$, the bits in $b$ and $\{u_{n^{(s)},l} : l > S - (s - 2)\}$ must be known. The vector $b$ is clearly known. In addition, the bits in $\{u_{n^{(s)},l} : l > S - (s - 2)\}$ are already decoded in previous stages. It is left to determine the bits in $\{u_{n^{(s)},l} : l > S - (s - 2)\}$ 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 $P_2$ is a degraded version of $P_1$, and $P_3$ is a degraded version of $P_1$. We first describe the encoding.

1) The $k_1$ information bits that are used to encode $x_1$ are (arbitrarily) partitioned into three subsets: $u_{1,1} \in \mathcal{X}^{k_1}$, $u_{1,2} \in \mathcal{X}^{k_2}$, and $u_1 \in \mathcal{X}^{k_3}$. 2) The $k_2$ information bits used to encode $x_2$ are (arbitrarily) partitioned into two subsets: $u_{2,1} \in \mathcal{X}^{k_1}$ and $u_{2,2} \in \mathcal{X}^{k_2}$. In addition, $u_2$ (used for encoding $x_1$) is also involved in the encoding of $x_2$.

3) The codewords $x_1$ and $x_2$ are defined as follows:

- $x_1 = u_{1,1}G_n \left( A_n^{(3)} \right) + u_{1,2}G_n \left( A_n^{(2)} \right) + bG_b \left( n \right) \left\lfloor A_n^{(1)} \right\rfloor$
- $x_2 = u_{2,1}G_n \left( A_n^{(3)} \right) + u_{2,2}G_n \left( A_n^{(2)} \right) + bG_b \left( n \right) \left\lfloor A_n^{(1)} \right\rfloor$

where $b \in \mathcal{X}^{k_2}$ is a predetermined and fixed vector.

4) The encoding of the codeword $x_3$ is based on the remaining $k_3$ information bits, denoted by $u_3 \in \mathcal{X}^{k_3}$:

- $x_3 = u_3G_n \left( A_n^{(3)} \right) + u_{1,2}G_n \left( A_n^{(2)} \right) + bG_b \left( n \right) \left\lfloor A_n^{(1)} \right\rfloor$

The order of decoding the information bits for all possible assignments of codewords over a set of three parallel and degraded 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 $\left\{ A_n^{(i)} \right\}_{i \in S'}$ satisfying the conditions in Corollary 1. It is noted that the index sets in $\mathcal{A}_n^{(i)}$, $s \in [S]$, 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 CONCLUSION

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 nonbinary polar codes (see [8] and [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 nonbinary 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].
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.

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

APPENDIX

A. On the Symmetry Property of Nonbinary Polarization

Definition 4 (Nonbinary 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 \( T : \mathcal{Y} \times \mathcal{X} \rightarrow \mathcal{Y} \) which satisfies the following properties.

1) For every \( x \in \mathcal{X} \), the function \( T(\cdot, x) : \mathcal{Y} \rightarrow \mathcal{Y} \) is bijective.

2) For every \( x_1, x_2 \in \mathcal{X} \) and \( y \in \mathcal{Y} \), the following equality holds:

\[
p(y|x_1) = p(T(y, x_2 - x_1)|x_2).
\]

Lemma 2 (Message Independence Property for Nonbinary Symmetric-Channel Polarization): Let \( p \) be a symmetric DMC and \( \mathbf{0} \) be the \( n \)-length all-zero vector over \( \mathcal{X} \). Denote by \( P_n(\mathcal{E}^t_i(p)|\mathbf{u}) \) the probability of the event \( \mathcal{E}^t_i \) in (13), assuming that \( \mathbf{w} = \mathbf{u} \) in (5). Then

\[
P_n(\mathcal{E}^t_i(p)|\mathbf{u}) = P_n(\mathcal{E}^t_i(p)|\mathbf{0})
\]

for every \( \mathbf{u} \in \mathcal{X}^n \) and \( t \in [n] \).

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

\[
T(y, x) = (T(y_1, x_1), T(y_2, x_2), \ldots, T(y_n, x_n)).
\]

Subtraction of a vector is also defined itemwise, i.e., \(-x = (x_1, \ldots, x_n) = (-x_1, \ldots, x_n)\). Based on the symmetry property of that channel, for every \( t \in [n], y \in \mathcal{Y}^n, w \in \mathcal{X}^{n-t}, \) \( w \in \mathcal{X}^t, \) and \( a \in \mathcal{X}^n \), we have

\[
p_n(\{y, (w_1, \ldots, w_{t-1})|w_t\})
\]

\[
= \frac{1}{\mathcal{X}^{|t-1|}} \sum_{c \in \mathcal{X}^{|t-1|}} p_n(\{y, (w_1, \ldots, w_{t-1})|c, G_{n,t}\})
\]


(18)

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

\[
p_n(\{y, (w_1, \ldots, w_{t-1})|w_t\})
\]

\[
= p_n(\{T(y, aG_n, (w_1, \ldots, w_{t-1}) + (a_1, \ldots, a_{t-1})|w_t + a_t\}, (18)
\]

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

\[
(\mathbf{w}, \mathbf{y}) \in \mathcal{E}^t_i(p) \iff (\mathbf{a} + \mathbf{w}, T(\mathbf{y}, \mathbf{a} \cdot G_n)) \in \mathcal{E}^t_i(p). \quad (19)
\]

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

\[
1_{\mathcal{E}^t_i(p)}(\mathbf{u}, \mathbf{y}) = 1_{\mathcal{E}^t_i(p)}(\mathbf{0}, T(\mathbf{y}, \mathbf{u} \cdot G_n))
\]

\[
= \sum_{y \in \mathcal{Y}^n} p_n(\mathbf{y}|\mathbf{u} \cdot G_n) 1_{\mathcal{E}^t_i(p)}(\mathbf{0}, T(\mathbf{y}, \mathbf{u} \cdot G_n))
\]

\[
= \sum_{y \in \mathcal{Y}^n} p_n(\mathbf{y}|\mathbf{0}) 1_{\mathcal{E}^t_i(p)}(\mathbf{0}, y)
\]

\[
= P_n(\mathcal{E}^t_i(p)|\mathbf{0})
\]

where \( (a) \) follows from (5), \( (b) \) follows from (19) by plugging \( \mathbf{a} = \mathbf{u} \), and \( (c) \) follows since \( T(\mathbf{y}, \mathbf{x}) \) 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

\[
P_2(\mathbf{y}_2|x) = \sum_{\mathbf{y}_1 \in \mathcal{Y}_1} P_1(\mathbf{y}_1|x)D(\mathbf{y}_2|\mathbf{y}_1), \quad \forall x \in \mathcal{X}, \mathbf{y}_2 \in \mathcal{Y}_2.
\]

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 blocklength \( n \), the split channels of \( P_1 \) and \( P_2 \) are denoted by \( P^{(i, n)}_1 \) and \( P^{(i, n)}_2 \), respectively, for all \( i \in [n] \). Assume that the channel \( P_2 \) is a stochastically degraded version of channel \( P_1 \). Then, for every \( i \in [n] \), the split channel \( P^{(i, n)}_2 \) is a stochastically degraded version of the split channel \( P^{(i, n)}_1 \).
Proof: The proof follows by induction (see [10] and [11]).

Definition 6 (Stochastically Degraded Parallel Channels): Let \( \{P_i\}_{i=1}^S \) be a set of \( S \) parallel memoryless channels. The channels \( \{P_i\}_{i=1}^S \) 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 \( P_{s_i} \) is a stochastically degraded version of \( P_i \) for every \( i \in [S-1] \).

Proof of Corollary 1: From [5], it follows that there exists a sequence of sets \( \mathcal{A}^{(s)}_n \) satisfying \( |\mathcal{A}^{(s)}_n| \geq nR_s \) and \( Pr(\mathcal{E}_n(P_s)) \leq 2^{-n^\delta} \). 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 \( \{\mathcal{A}^{(s')}_n\} \), \( s > s' \), can be chosen such that the following properties are met.

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

\[
\mathcal{A}^{(s')}_n \subseteq \mathcal{A}^{(s')}_{n-1} \subseteq \cdots \subseteq \mathcal{A}^{(s'+1)}_n. \tag{20}
\]

An information index set \( \{\mathcal{A}^{(s')}_n\} \) 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 \( Pr(\mathcal{E}_n(P_s)) < 2^{-n^\delta} \). These are the rate and performance properties in (15) and (17) for the particular case of \( s = S \). Choose an arbitrary index \( l \in \mathcal{A}^{(s'+1)}_n \). Since \( P_{s'+1} \) is a degraded version of \( P_s \), then according to Lemma 3, the split channel \( P^{(s'+1)}_{s'+1,n} \) is a degraded version of the split channel \( P^{(s')}_{s'+n} \). It is clearly suboptimal to first degrade the observation vector \( y \in \mathcal{Y}_n \) to create a vector \( \tilde{y} \in \mathcal{Y}_n^{s'+1} \), and only then decode the corresponding information bit. Consequently, \( \mathcal{E}_{n}(P_s) \subseteq \mathcal{E}_{n}^{s'+1}(P_s) \), which implies that if \( P^{(l)}_{s'+1,n} \) satisfies (17), so does \( P^{(l)}_{s'+n} \). It follows that \( l \) is a valid index for \( \mathcal{A}^{(s')}_n \) in terms of the performance property. That is, for every \( l \in \mathcal{A}^{(s')}_{n+1} \), it follows that \( Pr(\mathcal{E}_n(P_{s'})) < 2^{-n^\delta} \). Therefore, we set \( \mathcal{A}^{(s')}_n = \mathcal{A}^{(s'+1)}_{n+1} \cup \mathcal{A}^{(s')}_{n} \). The set sequences \( \{\mathcal{A}^{(s')}_n\}, s \geq s' \), satisfy the following properties.

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

The proof follows by induction.

ACKNOWLEDGMENT

The anonymous reviewers and the Associate Editor, Dr. Uri Erez, are gratefully acknowledged for constructive comments that improved the lucidity of the presentation.

REFERENCES

[1] F. M. J. Willems and A. Gorokhov, “Signaling over arbitrarily permuted parallel channels,” IEEE Trans. Inf. Theory, vol. 54, no. 3, pp. 1374–1382, Mar. 2008.

[2] A. Lapidoth and P. Narayan, “Reliable communication under channel uncertainty,” IEEE Trans. Inf. Theory, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[3] A. Hitron, A. Khina, and U. Erez, “Transmission over arbitrarily permuted parallel Gaussian channels,” in Proc. IEEE Int. Symp. Inf. Theory, Boston, MA, USA, Jul. 2012, pp. 2671–2675.

[4] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[5] E. Arikan and E. Telatar, “On the rate of channel polarization,” in Proc. IEEE Int. Symp. Inf. Theory, Seoul, Korea, Jun. 2009, pp. 1493–1495.

[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, U.K.: Cambridge Univ. Press, 2006.

[8] R. Mori and T. Tanaka, “Channel polarization on q-ary discrete memoryless channels by arbitrary kernels,” in Proc. IEEE Int. Symp. Inf. Theory, Austin, TX, Jun. 2010, pp. 894–898.

[9] E. Sasoglu, E. Telatar, and E. Arikan, “Polarization for arbitrary discrete memoryless channels,” in Proc. IEEE Inf. Theory Workshop, Taormina, Italy, Oct. 2009, pp. 144–148.

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

[11] S. B. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” IEEE Trans. Inf. Theory, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[12] S. H. Hassani, S. B. Korada, and R. Urbanke, “The compound capacity of polar codes,” in Proc. 47th Annu. Allerton Conf. Commun., Control Comput., Monticello, IL, USA, Sep. 2009, pp. 16–21.

[13] S. B. Korada, E. Sasoglu, and R. Urbanke, “Polar codes: Characterization of exponent, bounds, and constructions,” IEEE Trans. Inf. Theory, vol. 56, no. 12, pp. 6253–6264, Dec. 2010.

[14] S. B. Korada and E. Sasoglu, “A class of transformations that polarize binary-input memoryless channels,” in Proc. IEEE Int. Symp. Inf. Theory, Seoul, Korea, Jun. 2009, pp. 1478–1482.

[15] E. Arikan and G. Markarian, “Two-dimensional polar coding,” presented at the 10th Int. Symp. Commun. Theory Appl., Amblisbile, U.K., Jul. 2009.

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

Eran Hof (S’08) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2001, 2005, and 2010, respectively. During 2000–2001, he was with RAFAEL (Haifa, Israel) as a Research Assistant. During 2001–2007, he served in the Israel Defence Forces as a Research Engineer. Since 2011 he is with the Samsung Semiconductor Israel R&D Center (Ramat-Gan, Israel).

Igal Sason (S’98–M’02–SM’11) was born in Israel in 1969. He received the B.Sc. and Ph.D. degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 1992 and 2001, respectively. During 2001–2003, he was a Scientific Collaborator in the Faculty of Computer Science and Communications, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. Since 2003, he has been with the Department of Electrical Engineering at the Technion where he is currently an Associate Professor. His research interests are in coding theory and information theory, especially codes defined on graphs and iterative decoding. Another recent research interest of him concerns relationships between information theory and probability theory. Dr. Sason is a co-recipient of the 2003 IEEE Communications Society and Information Theory Society Joint Paper Award. During 2003–2006 he was a recipient of the Alon Fellowship for young faculty members and the Career Development Chair. From 2008 to 2011 he served as an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.
Shlomo Shamai (Shitz) (S’80–M’82–SM’89–F’94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai is an IEEE Fellow and a member of the Israeli Academy of Sciences and Humanities. He is the recipient of the 2011 Claude E. Shannon Award. He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless Communications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.

Chao Tian (S’00–M’05–SM’12) received the B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2000 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Cornell University, Ithaca, NY in 2003 and 2005, respectively.

Dr. Tian was a postdoctoral researcher at Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007. He joined AT&T Labs-Research, Florham Park, New Jersey in 2007, where he is now a Senior Member of Technical Staff. His research interests include multi-user information theory, joint source-channel coding, quantization design and analysis, as well as image/video coding and processing. Dr. Tian is currently an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.