Polar Coding for Bit-Interleaved Coded Modulation

Hessam Mahdavifar, *Member, IEEE*, Mostafa El-Khamy, *Senior Member, IEEE*,
Jungwon Lee, *Senior Member, IEEE*, and Inyup Kang, *Member, IEEE*

***Abstract*—A polar coding scheme is proposed for reliable communication over channels with bit-interleaved coded modulation (BICM). In an ideal information-theoretic model, BICM schemes are modeled as a set of multiple binary-input channels that experience different reliabilities. This model is referred to as *multi-channel*. The conventional scheme for encoding information over multi-channels is to encode separately across the individual constituent channels, which requires multiple encoders and decoders. As opposed to the conventional scheme, we take advantage of the recursive structure of polar codes to construct a unified compound polar code. The proposed scheme has a single encoder and decoder and can be used over multi-channels, in particular over channels with BICM. We prove that the scheme achieves the multi-channel capacity with the same error decay rate as in Arıkan's polar codes. Furthermore, due to more levels of channel polarization, we obtain a better finite block length performance compared with the separated polarization scheme. This is confirmed through various simulations over the binary erasure channel and with transmissions over additive white Gaussian noise (AWGN) and fast fading channels with BICM and different modulation orders.**

***Index Terms*—Bit-interleaved coded modulation (BICM), capacity, channel polarization, compound polar code, multi-channels.**

I. Introduction

POLAR codes, introduced by Arıkan in [1], are the first provable capacity-achieving codes for the class of binary-input symmetric discrete memoryless channels with explicit construction and low encoding and decoding complexity. Polar codes and a polarization phenomenon have been successfully applied to various problems such as wiretap channels [2], data compression [3], [4], and multiple-access channels [5], [6].

In this paper, we aim at developing methods to construct polar-based schemes for transmissions with bit-interleaved coded modulation (BICM). The BICM technique was introduced by Zehavi in [7], where he proposed that the reliability of coded modulation over a Rayleigh fading channel can be improved by means of bitwise interleaving at the encoder output. Caire *et al.* elaborated on Zehavi's decoder in [8], where they presented and analyzed a BICM model as a set of parallel independent binary-input channels under the assumption of an infinite-length interleaver. We refer to this information-theoretic model as a multi-channel. Fading channels can also be modeled as multi-channels if the channel state is picked from a finite set and is known at the transmitter. In this case, the transmission takes place over a series of channels bundled together as a multi-channel.

One straightforward solution to the problem of polar code construction for multi-channels is to encode the information separately across the constituent channels using polar encoders specific to each channel. In this paper, we propose a method to do polar coding across multi-channels with a unified single encoder and decoder, which is desirable from a practical point of view. Furthermore, by combining all the channels together and sending one single codeword, efficiently designed for the resulting multi-channel, we achieve a better tradeoff between the rate and the probability of error. This is due to the larger effective transmission block that can be translated into more levels of channel polarization. We call this proposed unified polar-based scheme for transmission over multi-channels as *compound polar codes*.

We presented the channel polarization and the construction of compound polar codes for the case of 2-multi-channels in [9]. Moreover, the scheme for general l-multi-channels is briefly discussed in [9], without numerical simulations and proofs of channel polarization. In a related work, the problem of polarization for quasi-static fading channels was later addressed in [10], where the channel is restricted to have two distinct fading values. One can observe that the quasi-static fading channel with two fading values can be actually modeled as a 2-multi-channel. The work of this paper can be used to address the problem of channel polarization and constructing capacity-achieving codes for quasi-static fading channels with more than two fading values. In an independent, yet relevant, line of research, the multilevel coding (MLC) of [11], [12] was incorporated into polar codes [13], [14] to construct polar-coded modulation schemes.

Using the developed methods in this paper, we construct compound polar codes for transmission over BICM channels and provide simulation results. In particular, BICM with 16-ary quadrature amplitude modulation (16-QAM) and 64-QAM are considered. In fact, 16-QAM BICM with Gray labeling can be approximately modeled as a multi-channel with two constituent channels, simply called 2-multi-channel. A random interleaver concatenated with a designed interleaver for polarization is deployed on top of the compound polar encoder to guarantee

Manuscript received August 22, 2014; revised December 10, 2014, April 9, 2015, and May 29, 2015; accepted May 31, 2015. Date of publication June 10, 2015; date of current version May 12, 2016. The review of this paper was coordinated by Dr. A. J. Al-Dweik.

H. Mahdavifar is with Modem R&D, Samsung Electronics, San Diego, CA 92121 USA (e-mail: h.mahdavifar@samsung.com).

M. El-Khamy is with Modem R&D, Samsung Electronics, San Diego, CA 92121 USA, and also with the Department of Electrical Engineering, Alexandria University, Alexandria 21544, Egypt (e-mail: mostafa.e@samsung.com).

J. Lee is with Modem R&D, Samsung Electronics, San Diego CA 92121 USA, and also with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA (e-mail: jungwon2.lee@samsung.com).

I. Kang is with SOC Development, System LSI, Samsung Electronics, Yongin 446-811, Korea (e-mail: inyup.kang@samsung.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2015.2443772

that channel polarization happens. For the construction of compound polar code, the good bit-channels for transmission of information bits are picked according to the probability of error they observe through the compound polar transformation. Similarly, 64-QAM BICM system with Gray labeling can be approximately modeled as a 3-multi-channel. For the construction of compound polar code, a 3 × 3 building block is considered on top of which n levels of recursive polarization is applied, for some positive integer n. In this case, the length of the constructed code is $N = 3 \times 2^n$. Notice that there are multiple choices for the 3 × 3 building block among which we pick the one with the best polarization rate. The channel polarization and capacity achieving properties of the proposed compound polar codes are proved for the general case of l-multi-channels.

The rest of this paper is organized as follows. In Section II, some background on polar coding and BICM is provided. In Section III, we start by explaining the straightforward separated scheme. Then, we explain our scheme for compound polar transformation for 2-multi-channels, followed by discussions on extending the construction to l-multi-channels. The construction of compound polar codes and the proofs for their capacity achieving property is discussed in Section IV. In Section V, the successive cancelation (SC) decoding of polar codes is extended to compound polar codes. In Sections VI and VII, we provide some numerical analysis for the case of 2-multi-channels where the constituent channels are binary erasure channels (BECs), and present the simulation results for additive white Gaussian noise (AWGN) and fast fading channel with 16-QAM BICM and 64-QAM BICM constellations. Comparisons with some existing schemes and related works are provided in Section VIII. We close this paper by mentioning some directions for future work in Section IX.

II. Preliminaries

A. Polar Codes

Here, we provide an overview of the groundbreaking work of Arıkan [1] and others [15]–[17] on polar codes and channel polarization.

Polar codes are constructed based upon a phenomenon called *channel polarization* discovered by Arıkan [1]. The basic polarization matrix is given as

$$G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{1}$$

Consider two independent copies of a binary-input discrete memoryless channel (B-DMC) $W : \{0,1\} \to \mathscr{Y}$. The two input bits (u_1, u_2), which are drawn from independent uniform distributions, are multiplied by G and then transmitted over the two copies of W. One level of channel polarization is the mapping $(W, W) \to (W^-, W^+)$, where $W^- : \{0,1\} \to \mathscr{Y}^2$, and $W^+ : \{0,1\} \to \{0,1\} \times \mathscr{Y}^2$ with the following channel transformation:

$$W^-(y_1, y_2|u_1) = \frac{1}{2} \sum_{u_2 \in \{0,1\}} W(y_1|u_1 \oplus u_2) W(y_2|u_2)$$
$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2).$$

The *bit-channels* W^- and W^+ are indeed the channels that u_1 and u_2 observe assuming the following scenario: The first bit u_1 is decoded assuming u_2 is noise, and then, u_2 is decoded assuming that u_1 is decoded successfully and is known.

The channel polarization is continued recursively by further splitting W^- and W^+ to get W^{--}, W^{-+}, W^{+-}, W^{++}, etc. This process can be explained best by means of Kronecker power of G. The Kronecker power of G is defined by induction. Let $G^{\otimes 1} = G$ and for any $n > 1$

$$G^{\otimes(n)} = \begin{bmatrix} G^{\otimes(n-1)} & 0 \\ G^{\otimes(n-1)} & G^{\otimes(n-1)} \end{bmatrix}$$

It can be observed that $G^{\otimes(n)}$ is a $2^n \times 2^n$ matrix. Let $N = 2^n$. Then, $G^{\otimes(n)}$ is the $N \times N$ polarization matrix. Following the convention, the random variables are denoted by uppercase letters, whereas their instances are denoted by lowercase letters, except for n and N, which are reserved for specifying the block length of the code. Let $(U_1, U_2, \ldots, U_N)$, denoted by U_1^N, be a block of N independent and uniform binary random variables. The polarization matrix $G^{\otimes(n)}$ is applied to U_1^N to get $X_1^N = U_1^N G^{\otimes n}$. Then, X_i's are transmitted through N independent copies of a B-DMC W. The output is denoted by Y_1^N. This transformation with input U_1^N and output Y_1^N is called the polar transformation. In this transformation, N independent uses of W are transformed into N bit-channels, as described in the following. Let $W^N : \mathscr{X}^N \to \mathscr{Y}^N$ denote the channel consisting of N independent copies of W, i.e.,

$$W^N \left(y_1^N | x_1^N\right) \stackrel{\text{def}}{=} \prod_{i=1}^{N} W(y_i|x_i). \tag{2}$$

The combined channel $\widetilde{W}$ is defined with transition probabilities given by

$$\widetilde{W} \left(y_1^N | u_1^N\right) \stackrel{\text{def}}{=} W^N \left(y_1^N | u_1^N G^{\otimes n}\right). \tag{3}$$

For $i = 1, 2, \ldots, N$, the ith bit-channel is denoted by $W_N^{(i)}$, following the convention settled in [1],[1] and is defined as follows:

$$W_N^{(i)} \left(y_1^N, u_1^{i-1} | u_i\right) \stackrel{\text{def}}{=} \frac{1}{2^{n-1}} \sum_{u_{i+1}^N \in \{0,1\}^{n-i}} \widetilde{W} \left(y_1^N | u_1^N\right) \tag{4}$$

Intuitively, this is the channel that bit u_i observes through an *SC decoder*, deployed at the output. Under this decoding method, proposed by Arıkan for polar codes [1], all the bits u_1^{i-1} are already decoded and are assumed available at the time that u_i is being decoded. The channel polarization theorem states that as N goes to infinity, the bit-channels start polarizing, meaning that they either become a noiseless channel or a pure-noise channel.

To measure how good a binary-input channel W is, Arıkan uses the *Bhattacharyya parameter* of W, which is denoted by $Z(W)$ [1] and defined as

$$Z(W) \stackrel{\text{def}}{=} \sum_{y \in \mathscr{Y}} \sqrt{W(y|0) W(y|1)}.$$

[1] The superscript is used with brackets in this notation to distinguish with the notation used for vectors.

It is shown in [1] that, for any B-DMC W

$$1 - I(W) \leq Z(W) \leq \sqrt{1 - I(W)^2} \tag{5}$$

where $I(W)$ is the *symmetric capacity* of W, i.e., the mutual information between the input and the output, given that the input distribution is uniform. If W is symmetric, then $I(W)$ is the capacity of W. Channels with $Z(W)$ close to zero are almost noiseless, whereas channels with $Z(W)$ close to one are almost pure-noise channels. More precisely, it can be proved that the probability of error of a binary symmetric memoryless channel (BSM) is upper-bounded by its Bhattacharyya parameter. Let $[N]$ denotes the set of positive integers less than or equal to N. The set of *good bit-channels* $\mathcal{G}_N(W, \beta)$ is defined, for any $\beta < 1/2$ [15] [16], as follows:

$$\mathcal{G}_N(W, \beta) \stackrel{\text{def}}{=} \left\{ i \in [N] : Z\left(W_N^{(i)}\right) < 2^{-N^\beta}/N \right\}. \tag{6}$$

Theorem 1: [1], [15] For any BSM channel W and any constant $\beta < 1/2$, we have

$$\lim_{N \to \infty} \frac{|\mathcal{G}_N(W, \beta)|}{N} = \mathcal{C}(W)$$

where $\mathcal{C}(W)$ is the capacity of the channel W.

Theorem 1 readily leads to a construction of capacity-achieving *polar codes*. The idea is to transmit the information bits over the good bit-channels while fixing the input to the bad bit-channels to *a priori* known values, e.g., zeros. The following theorem (the second part of [1, Prop. 2]), summarizes the key property of the encoder–decoder pair of polar codes.

Theorem 2: For any $\beta < 1/2$ and any BSM channel W, the polar code of length N associated with the set of good bit-channels $\mathcal{G}_N(W, \beta)$ defined in (6) approaches the capacity of W. Furthermore, the probability of frame error under SC decoding is less than 2^{-N^β}.

The polarization theory is generalized to the case where the initial matrix G is an arbitrary $l \times l$ matrix with $l \geq 3$ [17]. A necessary and sufficient condition on G is provided, which guarantees polarization for any BSM channel. It is proved in [17] that, if G is an invertible matrix, then channel polarization happens if and only if G is not upper triangular.

A method to characterize the *rate of polarization* for a given $l \times l$ matrix G is also provided in [17]. For any B-DMC W, the rate of polarization $E(G)$ is defined as the supremum of all positive numbers β such that the fraction of good bit-channels $\mathcal{G}_N(W, \beta)$, generalized for the polar transformation of length l^N, approaches $I(W)$ as N goes to infinity. It is also shown that $E(G)$ can be characterized in terms of *partial distances* of G, defined as follows.

Definition: [17]: For an $l \times l$ matrix $G = \left[g_1^T, g_2^T, \ldots, g_l^T\right]^T$, the partial distances D_i for $i = 1, 2, \ldots, l$ are defined as

$$D_i \stackrel{\text{def}}{=} d_H\left(g_i, \langle g_{i+1}, \ldots, g_l \rangle\right), i = 1, 2, \ldots, l-1$$

$$D_l \stackrel{\text{def}}{=} d_H(g_l, 0).$$

Theorem 3: [17]: For any BSM W and any $l \times l$ matrix G with partial distances $\{D_i\}_{i=1}^l$, the rate of polarization $E(G)$ is given by

$$E(G) = \frac{1}{l} \sum_{i=1}^{l} \log_l D_i. \tag{7}$$

B. BICM and Multi-channels

In general, a coded modulation scheme is considered the concatenation of an encoder with a memoryless modulator. The modulator maps the encoded bits into symbols from a symbol set $\mathscr{X}$ through a one-to-one mapping, where $\mathscr{X}$ is the input alphabet of the channel over which the modulated symbols are transmitted. Let $|\mathscr{X}| = 2^m$ and elements of $\mathscr{X}$ be represented as binary m-tuples. Let the output of the channel be denoted by $y \in \mathscr{Y}$. In a BICM scheme, the length of the underlying binary code is mN, where N denotes the number of channel uses. The binary codeword generated by the encoder is interleaved by an interleaver. Then, blocks of m bits are mapped to signal constellation symbols of $\mathscr{X}$.

The classical BICM decoder proposed by Zehavi [7] treats each of the m bits in a symbol as independent and computes the reliability information for each bit based on the received symbol by neglecting the bit dependence. This results in mismatched decoding metrics. Moreover, according to the constellation signal space and the bit-to-symbol mapping, bits with different indexes per symbol can observe different channel reliabilities [18], [19].

For any $j \in [m]$ and $x \in \mathscr{X}$, let $b_j(x)$ denote the jth bit of symbol x. For $B \in \{0, 1\}$, let $\mathscr{X}_B^j$ denote the set of all symbols in $\mathscr{X}$ with their jth bit being B. Under the assumption of an infinite-length random interleaver, the BICM capacity is defined as the sum of m independent and parallel binary-input channels, which is denoted by $C_{\mathscr{X}}^{\text{bicm}}$ and is given in the following [19]:

$$C_{\mathscr{X}}^{\text{bicm}} = \sum_{j=1}^{m} E\left\{ \log \frac{\sum_{x' \in \mathscr{X}_B^j} P_{Y|X}(Y|x')}{\frac{1}{2}\sum_{x' \in \mathscr{X}} P_{Y|X}(Y|x')} \right\} \tag{8}$$

where the expected value is taken with respect to the output of channel $Y \in \mathscr{Y}$ and B, assuming a uniform input distribution on $\mathscr{X}$ and, consequently, on B. Furthermore, the notion of generalized mutual information (GMI) is introduced in [19] to measure the asymptotic performance limit of BICM schemes under the presence of suboptimal mismatched decoding metric calculation. It is shown in [20] that under maximum *a posteriori* (MAP) detection, the GMI of the BICM scheme is equal to the BICM capacity given by (8), although the assumption of infinite interleaving is lifted.

Following the framework proposed in [8], for the theoretical analysis of BICM, a single interleaver, also referred to as S-interleaver, is often considered. However, in the original BICM paper of Zehavi [7], multiple interleavers are adapted, e.g., three interleavers that interleave each of the bits' positions in an 8-phase-shift keying (PSK) symbol. The optimal interleaver design has been studied under various models and scenarios [21], [22]. It is shown in [21] that, for AWGN channels, the performance at the error floor region of bit-interleaved turbo-coded modulation can be improved by carefully designing the interleaver. BICM with certain interleaver designs is now a part of most of existing wireless standards, including HSPA and LTE. In our simulation setups in this paper, we will consider a random S-interleaver concatenated with a proposed interleaver designed for channel polarization purposes, which will be clarified later.

The individual bit metrics in each symbol computed by the demodulator can be regarded as statistically independent assuming a perfect interleaver. However, the perfect interleaver requires an infinite length, which makes it practically infeasible. Assuming the perfect interleaver is for theoretical purposes, which simplifies the information-theoretic model. In this model, the BICM system is regarded as a set of m independent and memoryless binary-input channels in parallel, which are connected to the encoder output by a random switch [8]. In fact, the random switch models the perfect interleaver. This motivates us to define the notion of multi-channel. However, this notion suffers from the problem of assuming a perfect interleaver and, hence, falls short to fully describe a practical finite-length BICM system.

A multi-channel consists of multiple B-DMCs and is defined as follows. Let $W_i : \mathscr{X} \to \mathscr{Y}_i$, for $i = 1, 2, \ldots, l$, denote the l given B-DMCs (indeed, $\mathscr{X} = \{0, 1\}$). Then, the corresponding l-multi-channel $(W_1 \cdot W_2 \cdots W_l) : \mathscr{X}^l \to \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_l$ is another DMC whose transition probability for $x_1^l = (x_1, x_2, \ldots, x_l) \in \mathscr{X}^l$ and $y_1^l = (y_1, y_2, \ldots, y_l) \in \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_l$ is given by

$$(W_1 \cdot W_2 \cdots W_l) \left(y_1^l | x_1^l\right) = \prod_{i=1}^{l} W_i(y_i | x_i).$$

In fact, each binary sequence of length l is transmitted over this multi-channel in such a way that each bit is transmitted through one of the l channels. In general, for any N that is a multiple of l, a sequence of N bits is transmitted over this multi-channel in such a way that each channel carries N/l bits. It is known to both the transmitter and the receiver which channel carries which bits in the sequence. An arbitrary interleaver and deinterleaver can be deployed at the transmitter and the receiver; thus, the ordering of channels does not matter.

The symmetric capacity of the l-multi-channel $(W_1.W_2 \cdots W_l)$ is equal to the sum of the symmetric capacities of the constituent channels W_i's, i.e.,

$$I\left((W_1 W_2 \cdots W_l)\right) = \sum_{i=1}^{l} I(W_i). \tag{9}$$

This coincides with the definition of BICM capacity in (8) under the model of the independent and parallel binary channels. In the following, we consider the symmetric capacity per bit, which is the expression in (9) normalized by l. Moreover, note that, if the constituent channels are symmetric, then the symmetric capacity is same as the capacity [23, Ch. 7].

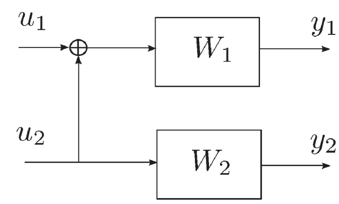


Fig. 1. Proposed building block of our scheme.

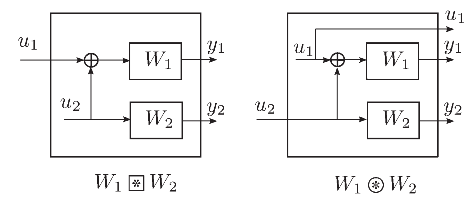


Fig. 2. Bit-channels of the proposed building block.

III. Compound Polar Transformation for Multi-channels

Here, the proposed scheme for the compound polar transformation in the case of 2-multi-channels is discussed first followed by methods for extension to the case of transmission over l-multi-channels.

A. *Compound Polar Transformation for 2-Multi-channels*

Suppose that two BSM channels W_1 and W_2 are given. The proposed 2×2 building block of our scheme is shown in Fig. 1. This block can be split into two bit-channels by generalizing the definition of channel combining suggested in [1]. Formally, for any two B-DMCs $W_1 : \mathscr{X} \to \mathscr{Y}_1$ and $W_2 : \mathscr{X} \to \mathscr{Y}_2$ (indeed, $\mathscr{X} = \{0, 1\}$), let $W_1 \boxast W_2 : \mathscr{X} \to \mathscr{Y}_1 \times \mathscr{Y}_2$ denote another B-DMC whose transition probability for any $(y_1, y_2) \in \mathscr{Y}_1 \times \mathscr{Y}_2$ and $u \in \mathscr{X}$ is given by

$$W_1 \boxast W_2(y_1, y_2 | u) = \frac{1}{2} \sum_{x \in \mathscr{X}} W_1(y_1 | u \oplus x) W_2(y_2 | u). \tag{10}$$

In addition, $W_1 \circledast W_2 : \mathscr{X} \to \mathscr{Y}_1 \times \mathscr{Y}_2 \times \mathscr{X}$ denote another B-DMC whose transition probability for any $(y_1, y_2) \in \mathscr{Y}_1 \times \mathscr{Y}_2$ and $x, u \in \mathscr{X}$ is given by

$$W_1 \circledast W_2(y_1, y_2, x | u) = \frac{1}{2} W_1(y_1 | u \oplus x) W_2(y_2 | u). \tag{11}$$

In fact, for a B-DMC W, $W^- = W \boxast W$, and $W^+ = W \circledast W$. The channels $W_1 \boxast W_2$ and $W_1 \circledast W_2$ are shown in Fig. 2.

The process of combining the two constituent channels in the first step can be regarded as one level of channel polarization. Then, the channel polarization is applied recursively on top of this building block to polarize $W_1 \boxast W_2$ and $W_1 \circledast W_2$. The scheme is shown in Fig. 3. The interleaver block π is designed in such a way that the first half of encoded block is transmitted through W_1 and the second half through W_2. We will prove that this particular interleaver will guarantee the channel polarization.

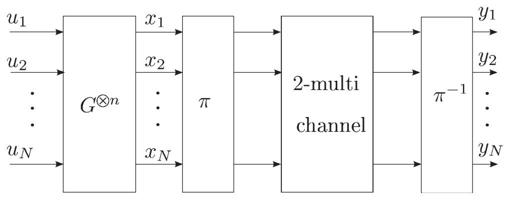


Fig. 3. Proposed scheme for 2-multi-channels with length N.

Remark: The channel combining operations $W_1 \boxast W_2$ and $W_1 \circledast W_2$ are not commutative channel operations. However, it turns out that the Bhattacharyya parameters of the combined channels are symmetric with respect to W_1 and W_2, i.e.,

$$Z(W_1 \boxast W_2) = Z(W_2 \boxast W_1)$$
$$Z(W_1 \circledast W_2) = Z(W_2 \circledast W_1).$$

Therefore, as far as the bound on the probability of error, the ordering of W_1 and W_2 does not matter. This will be shown by Theorem 8.

The transformation with input u_1^N and output y_1^N, as shown in Fig. 3, is called the *compound polar transformation*, which can be regarded as a generalization of Arıkan's polar transformation defined in Section II.

It will be shown in the following section that this compound polar transformation is equivalent to two separated polar transformations of length $N/2$ for $W_1 \boxast W_2$ and $W_1 \circledast W_2$ independently. Therefore, the channel polarization theorem can be established for this proposed transformation accordingly. Consequently, the polar code construction is done by picking the indexes corresponding to the set of good bit-channels for transmitting information bits. The definition of good bit-channels will also be provided in the following section.

B. Extending the Construction to l-Multi-channels

Suppose that an l-multi-channel with l constituent B-DMCs $W_1, W_2, \ldots, W_l$ is given. A generalization of the compound polar construction provided earlier would be straightforward using the results of [17]. Let G_0 be an $l \times l$ matrix that guarantees the channel polarization for the polar transformation $G_0^{\otimes n}$ according to the criteria proposed in [17]. Then, one can easily generalize the proposed construction for 2-multi-channels to l-multi-channels to construct codes of length $N = l^n$. However, the problem with this construction would be the decoding complexity of the SC decoding that increases by a factor of 2^l. The reason is that the 2×2 Arıkan's polarization butterfly is replaced by an $l \times l$ kernel across which the complexity of calculating the likelihood ratios (LRs) is 2^l. To maintain the complexity of the SC decoder as low as that of the regular polar codes, we propose an alternative method, which is described as follows. We take an $l \times l$ building block and apply n levels of regular 2×2 polarization levels on top of the $l \times l$ building block to construct compound polar codes of length $N = l2^n$. This idea is elaborated here, and the choice of building block will be discussed in Section VII.

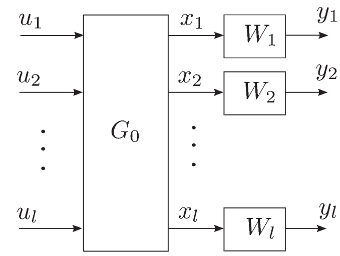


Fig. 4. Building block for the general case of l-multi-channels.

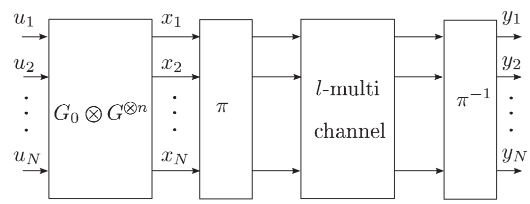


Fig. 5. Compound polar transformation for the general case of l-multi-channels.

Fix an $l \times l$ invertible matrix G_0 as the initial matrix of the construction. We will propose certain criteria to have one *good level of polarization* using G_0 in Section VII. The $l \times l$ building block of the construction corresponding to G_0 is shown in Fig. 4. In this figure, $x_1^l = u_1^l \cdot G_0$, and then, $x_1, x_2, \ldots, x_l$ are transmitted through $W_1, W_2, \ldots, W_l$, respectively.

For $n \geq 0$, the scheme with length $N = l2^n$ is constructed as follows. The polar transformation $G^{\otimes n}$ is applied on top of the building block shown in Fig. 4. The block diagram of this compound transformation is shown in Fig. 5. The polar transformation matrix is $G_0 \otimes G^{\otimes n}$, which is an $l2^n \times l2^n$ matrix. Through this transformation, the input sequence u_1^N is multiplied by $G_0 \otimes G^{\otimes n}$ and then interleaved by the interleaver π. We design the interleaver π in such a way that the first N/l encoded bits are transmitted through W_1, the second N/l encoded bits are transmitted through W_2, etc. This particular choice of π guarantees the channel polarization, which is proved in the following. This is the compound polar transformation for an l-multi-channel.

IV. Construction of Compound Polar Codes and the Proof of Capacity Achieving Property

Here, the construction of compound polar codes, based on the compound polar transformation explained earlier, is discussed. Then, we prove that the constructed codes are capacity achieving using the channel polarization theorem.

Let $W_i : \mathscr{X} \to \mathscr{Y}_i$, for $i = 1, 2, \ldots, l$, denote the l given B-DMCs (indeed, $\mathscr{X} = \{0, 1\}$). Given the compound scheme at a certain length N shown in Fig. 5, the individual bit-channels $(W_1 \cdot W_2 \cdots W_l)_N^{(i)}$ are defined. It is assumed that u_i's are samples of independent uniform binary random variables. Let $\widetilde{W}_N : \mathscr{X}^N \to \mathscr{Y}_1^{2^n} \times \mathscr{Y}_2^{2^n} \times \cdots \times \mathscr{Y}_l^{2^n}$ denote the channel from u_1^N to y_1^N, i.e.,

$$\widetilde{W}_N \left(y_1^N | u_1^N\right) = \prod_{i=1}^{l} \prod_{j=1}^{2^n} W_i \left(y_{(i-1)2^n+j} | x_{(i-1)2^n+j}\right)$$

where $x_1^N = u_1^N(G_0 \otimes G^{\otimes n})$. Then, for $i = 1, 2, \ldots, N$, $(W_1 \cdot W_2 \cdots W_l)_N^{(i)} : \mathscr{X} \to \mathscr{Y}_1 \times \mathscr{Y}_2, \ldots, \mathscr{Y}_l \times \mathscr{X}^{i-1}$ denote a B-DMC whose transition probability for any $(y_1, y_2, \ldots, y_l) \in \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_l$ and $(u_1, u_2, \ldots, u_i) \in \mathscr{X}^i$ is given by

$$(W_1 \cdot W_2 \cdots W_l)_N^{(i)} \left(y_1^N, u_1^{i-1} | u_i\right) \stackrel{\text{def}}{=} \frac{1}{2^{N-1}} \sum_{\overline{u} \in \{0,1\}^{N-i}} \widetilde{W}\left(y_1^N | (u_1^{i-1}, u_i, \overline{u})\right). \quad (12)$$

For notational convenience, for $i = 1, 2, \ldots, l$, let

$$\dot{W}_i = (W_1 \cdot W_2 \ldots W_l)_l^{(i)} \quad (13)$$

be the ith bit-channel in the building block shown in Fig. 4.

Lemma 4: For any l BSM channels $W_1, W_2, \ldots, W_l$

$$\sum_{i=1}^{l} \mathcal{C}(\dot{W}_i) = \sum_{i=1}^{l} \mathcal{C}(W_i).$$

See proof in the Appendix.

For any vector $u_1^N = (u_1, u_2, \ldots, u_N)$, let $u_{1,e}^N$ and $u_{1,o}^N$ denote the even-indexed and odd-indexed entries of the vector u_1^N, respectively. The following lemma states the recursive relation between the bit-channels using the channel combining operations. This lemma can be regarded as an extension of [1, Prop. 3] to compound polar transformation.

Lemma 5: For any l B-DMCs $W_1, W_2, \ldots, W_l$, and any $n \geq 1, N = l2^n, 1 \leq i \leq N/2$

$$(W_1 \cdot W_2 \cdots W_l)_N^{(2i-1)} = (W_1 \cdot W_2 \cdots W_l)_{N/2}^{(i)} \boxast (W_1 \cdot W_2 \cdots W_l)_{N/2}^{(i)}$$

$$(W_1 \cdot W_2 \cdots W_l)_N^{(2i)} = (W_1 \cdot W_2 \cdots W_l)_{N/2}^{(i)} \circledast (W_1 \cdot W_2 \cdots W_l)_{N/2}^{(i)}.$$

See proof in the Appendix.

In the next lemma, we establish a key property in the recursive structure of the proposed compound polar code, which is a direct consequence of the choice of the interleaver π in Fig. 5. In BICM schemes, π will be concatenated with the BICM channel interleaver. We will clarify more about this concatenation in the numerical results of the following section. For simplicity, suppose that the channel interleaver is the identity, i.e., in the absence of π, the encoded bits will be transmitted consecutively over $W_1, W_2, \ldots, W_l, W_1, W_2, \ldots$, etc. Then, $\pi : [N] \to [N]$ is specified as follows. For any $i \in [N]$, let $i = (j-1)2^n + r$ with $1 \leqslant r \leqslant 2^n$. Then

$$\pi(i) = rl + j. \quad (14)$$

The interleaver π guarantees the copies of $\dot{W}_j$, for $j = 1, 2, \ldots, l$ are combined together recursively, which leads to n levels of polarization. This recursive structure will be used in Theorem 7 to establish the channel polarization property.

Lemma 6: Let $W_1, W_2, \ldots, W_l$ be l given B-DMCs. Then, for any $n \geqslant 0$, $N = l2^n$, $1 \leqslant i \leqslant N$, let $i = (j-1)2^n + r$ with $1 \leqslant r \leqslant 2^n$. Then

$$(W_1 \cdot W_2 \cdots W_l)_N^{(i)} = [\dot{W}_j]_{2^n}^{(r)}$$

See proof in the Appendix.

The definition of good bit-channels is extended to the case of compound polar transformation, i.e., for any $\beta < 1/2$ and $N = l2^n$

$$\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta) \stackrel{\text{def}}{=} \left\{ i \in [N] : Z\left((W_1 \cdot W_2 \cdots W_l)_N^{(i)}\right) < 2^{-N^\beta}/N \right\}. \quad (15)$$

The following theorem establishes the channel polarization theorem for the compound polar transformation shown in Fig. 5.

Theorem 7: For any l BSM channels $W_1, W_2, \ldots, W_l$ and any constant $\beta < 1/2$, we have

$$\lim_{N \to \infty} \frac{|\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)|}{N} = \frac{1}{l} \sum_{i=1}^{l} \mathcal{C}(W_i).$$

Proof: If $N = l2^n$ is large enough, then we can pick β' such that $\beta < \beta' < 1/2$ and

$$\frac{2^{-\left(\frac{N}{l}\right)^{\beta'}}}{\frac{N}{l}} < \frac{2^{-N^\beta}}{N}.$$

Then, definitions of good bit-channels given in (6) and (15) together with Lemma 6 imply that

$$|\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)| \geqslant \sum_{i=1}^{l} \left| \mathcal{G}_{2^n}(\dot{W}_i, \beta') \right|. \quad (16)$$

Then, by Theorem 1 and Lemma 4

$$\lim_{N \to \infty} \frac{|\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)|}{N} \geqslant \sum_{i=1}^{l} \lim_{N \to \infty} \frac{\left|\mathcal{G}_{2^n}(\dot{W}_i, \beta')\right|}{N} = \frac{1}{l} \sum_{i=1}^{l} \mathcal{C}(\dot{W}_i) = \frac{1}{l} \sum_{i=1}^{l} \mathcal{C}(W_i).$$

On the other hand, we have

$$2^n \sum_{i=1}^{l} I(W_i) = \sum_{i=1} I\left((W_1 \cdot W_2 \cdots W_l)_N^{(i)}\right) \quad (17)$$

$$\geqslant \sum_{i \in \mathcal{G}_N(W_1, \ldots, W_l, \beta)} I\left((W_1 \cdot W_2 \cdots W_l)_N^{(i)}\right)$$

$$\geqslant \sum_{i \in \mathcal{G}_N(W_1, \ldots, W_l, \beta)} 1 - Z\left((W_1 \cdot W_2 \cdots W_l)_N^{(i)}\right) \quad (18)$$

$$\geqslant |\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)| - 2^{-N^\beta} \quad (19)$$

Equation (17) is by the chain rule on the mutual information between the input and the output of the scheme shown in Fig. 3.

Equation (18) follows by (5). Equation (19) holds by definition of the set of good bit-channels. Therefore

$$\lim_{N\to\infty} \frac{|\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)|}{N} \leqslant \frac{1}{l} \sum_{i=1}^{l} \mathcal{C}(W_i)$$

which completes the proof of theorem. ■

The encoding of our scheme, which we call *compound polar code* is discussed next. For a given $\beta < 1/2$, if $i \in \mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)$, then u_i is an information bit. Otherwise, u_i is set to zero. The encoded message x_1^N is transmitted through $W_1, W_2, \ldots, W_l$, as shown in Fig. 5. For the decoding, the SC decoding defined in [1] is performed, where the details are discussed in the following section. Then, the following theorem follows from Theorem 7, and its proof is similar to [1, Th. 2].

Theorem 8: For any $\beta < 1/2$ and any l BSM channels $W_1, W_2, \ldots, W_l$, the compound polar code of length N associated with the set of good bit-channels $\mathcal{G}_N(W_1, W_2, \ldots, W_l, \beta)$ defined in (15) approaches the average of the capacities of $W_1, W_2, \ldots, W_l$. Furthermore, the probability of frame error under SC decoding is less than 2^{-N^β}.

V. Decoding of Compound Polar Codes

The decoding method for compound polar codes is the SC decoder similar to Arıkan's polar codes, as mentioned earlier. The SC decoding for compound polar codes can be implemented by extending and modifying that of original polar codes in [1]. Let $N = l2^n$ and suppose that u_1^N is the input vector, as shown in Fig. 5. Let y_1^N denote the received word. For $i = 1, 2, \ldots, N$, if $W_N^{(i)}$ is not a good bit-channel, then the decoder knows that ith bit u_i is set to zero; therefore, $\hat{u}_i = u_i = 0$. Otherwise, the decoder computes the likelihood $L_N^{(i)}$ of u_i, given the channel outputs y_1^N and previously decoded $\hat{u}_1^{i-1}$, i.e.,

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) = \frac{(W_1, W_2 \ldots, W_l)_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|u_i = 0)}{(W_1, W_2 \ldots, W_l)_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|u_i = 1)}.$$

Then, it makes the hard decision based on $L_N^{(i)}$.

The likelihood functions $L_N^{(i)}$ can be computed recursively as follows. A straightforward calculation using the bit-channel recursion formulas in Lemma 6 for $n \geq 1$ gives the following recursive formulas:

$$\begin{aligned}
&L_N^{(2i-1)}(y_1^N, \hat{u}_1^{2i-2}) \\
&= \frac{1 + L_{N/2}^{(i)}(y_{1,o}^N, \hat{u}_{0,e}^{2i-2} \oplus \hat{u}_{0,o}^{2i-2}) L_{N/2}^{(i)}(y_{1,e}^N, \hat{u}_{0,o}^{2i-2})}{L_{N/2}^{(i)}(y_{1,o}^N, \hat{u}_{0,e}^{2i-2} \oplus \hat{u}_{0,o}^{2i-2}) + L_{N/2}^{(i)}(y_{1,e}^N, \hat{u}_{0,o}^{2i-2})} \\
&L_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}) \\
&= L_{N/2}^{(i)}(y_{1,o}^N, \hat{u}_{0,e}^{2i-2} \oplus \hat{u}_{0,o}^{2i-2})^{1-\hat{u}_{2i-1}} L_{N/2}^{(i)}(y_{1,e}^N, \hat{u}_{0,o}^{2i-2}).
\end{aligned}$$

The only question is how to initiate the SC decoder for $n = 0$, when $N = l$. A naive way of computing transition probabilities $\dot{W}_i(y_1^l, u_1^{i-1}|u_i)$ using (12) results in the complexity $O(2^l)$. The recursive steps can be done using Arıkan's refined SC decoding algorithm with complexity $O(N \log N/l)$. Therefore, the total complexity is $O\left(N\left(\log N - \log l + 2^l\right)\right)$. As N grows large, the dominating term is $N \log N$; therefore, the total complexity of the SC decoding algorithm is $O(N \log N)$.

Notice that we did not take into account a bit-reversal transformation as in [1]. As a result, we split y_1^N into $y_{1,o}^N$ and $y_{1,e}^N$ for recursive computation of log LRs (LLRs) as opposed to splitting y_1^N into $y_1^{N/2}$ and $y_{N/2+1}^N$ in [1].

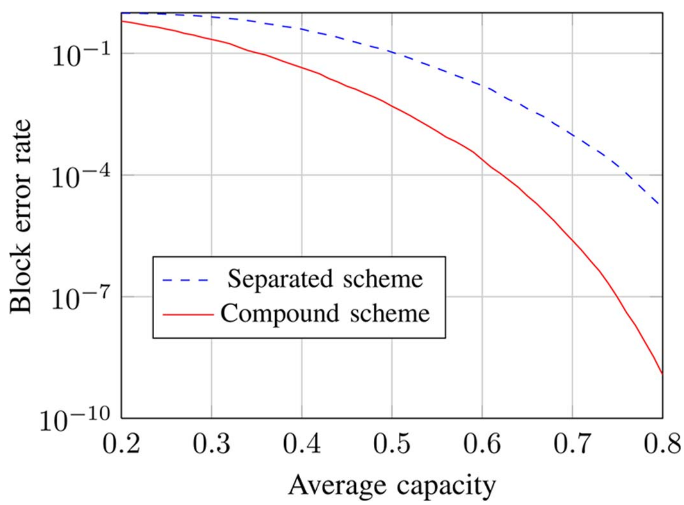


Fig. 6. Improvement in probability of error in the case of two BECs.

VI. Analysis of the Proposed Scheme for 2-Multi-Channels

A. Numerical Analysis for Binary Erasure Channels

The proposed compound polar code and the conventional separated scheme are compared for the case of 2-multi-channel $(W_1 \cdot W_2)$ where both the constituent channels W_1 and W_2 are BECs. Let $W_1 = \text{BEC}(p_1)$ and $W_2 = \text{BEC}(p_2)$, where p_1 and p_2 are the erasure probabilities. It is assumed that p_1 and p_2 vary such that $p_1 - p_2 = 0.1$. A rate of $0.8 \times$ capacity for each of the constituent channels W_1 and W_2 and the multi-channel $(W_1 \cdot W_2)$ whose capacity is $\mathcal{C} = 1 - (p_1 + p_2)/2$ per bit is fixed. The transmission block length is also fixed to 2^{14} (2^{13} each in the case of separated scheme). Let $P_{e,1}$ and $P_{e,2}$ be the computed error probability using the Bhattacharyya parameters of the selected good bit-channels for each of the two subblocks of length 2^{13} in the separated scheme. Then, the total probability of frame error assuming that the inputs are independent is bounded by

$$P_e = 1 - (1 - P_{e,1})(1 - P_{e,2})$$

We compare this with the probability of error computed using the Bhattacharrya parameters of the selected good bit-channels in the proposed compound polar code constructed with respect to the compound polar transformation in Fig. 3. The result is shown in Fig. 6.

B. Simulation Results for 16-QAM BICM Over AWGN

As stated previously, transmission with 16-QAM BICM and Gray labeling can be approximately modeled as transmission over a 2-multi-channel. Among the four bits in each transmitted symbol, two of them go through one channel denoted by W_1, and the other two go through another channel denoted by W_2. Under the Gray labeling, the first and the second bits in each symbol go through W_1, and the third and the fourth bits go through W_2.

In the proposed compound scheme, a compound polar code of length N is constructed by picking the good bit-channels in the compound polar transformation specified later. A random S-interleaver, which can be any arbitrary interleaver, of length N is applied to the encoded block. There will be another interleaver π designed specifically for channel polarization as discussed in Section III. The concatenation of random interleaver and π map the first subblock of length $N/2$ of the codeword to the first and second bits of the $N/4$ symbols and map the second subblock of length $N/2$ of the codeword to the third and fourth bits of the $N/4$ symbols. For instance, if the first interleaver is the identity interleaver, then π is as follows. For any $i \in [N]$, let $i = (N/2)j_1 + 2j_2 + r$, where $1 \leqslant 2j_2 + r \leqslant N/2$ and $r \in \{1, 2\}$. Then, $\pi(i) = 4j_2 + 2j_1 + r$. The constructed $N/4$ symbols will be transmitted by using $N/4$ independent uses of an AWGN channel.

To pick the best bit-channels for the construction of compound polar code, a simulation-based Monte Carlo method is used to estimate the probability of error of the individual bit-channels in the compound polar transformation depicted in Fig. 3. A SC decoder is used and for $i = 1, 2 \ldots, N$, it is assumed that all the previous input bits indexed by $1, 2, \ldots, i-1$ are provided to the decoder by a genie at the time that the ith bit is decoded to estimate the probability of error of the ith bit, i.e., $P_e((W_1 \cdot W_2)_N^{(i)})$, using Monte Carlo simulation. We use the estimated probability of error of the bit-channels to pick the best ones for code construction instead of using Bhattacharryaa parameters, which are exponentially hard to compute. For instance, to construct a compound polar code of rate 0.5, the bit-channels are sorted based on their estimated probability of error in an ascending order. Then, the first $N/2$ of them are picked to carry the information bits.

For the code construction and simulation, the block length is fixed to $2^{10} = 1024$, and the rate is fixed to 1/2. In the conventional separated scheme, two independent polar codes for W_1 and W_2 are constructed separately. Notice that the total rate is 1/2 and is not necessarily the individual rates for each constituent channel. To have a fair comparison, we aim at optimizing the performance of the separated scheme by carefully assigning rates to be transmitted on each channel. In this paper, we use the energy-per-bit-to-noise-power-spectral-density ratio (E_b/N_0) metric to specify the SNR. For $E_b/N_0 =$ 5 dB, the probability of error of all bit-channels are estimated for each of the constituent channels. Then, the rates are picked such that the total probability of error is minimized. It turns out that the rate 0.621 on the stronger channel W_1 and 0.379 on the weaker channel W_2 minimizes the total probability of error at 5 dB. The compound polar code is constructed by picking the best bit-channels at $E_b/N_0 = 5$ dB. The comparison between the two methods is shown in Fig. 7. One can observe that our compound polar code is about 1.5 dB better than the conventional separated scheme at moderate SNRs. As SNR increases, the gap becomes larger, and the curves start diverging.

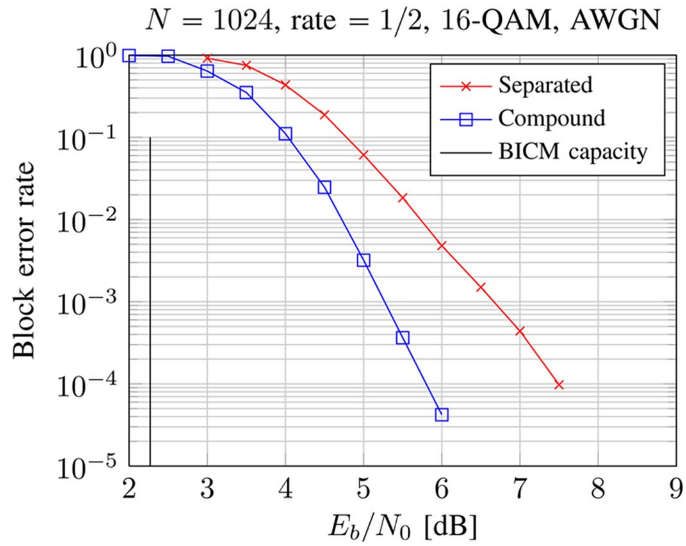


Fig. 7. Performance of the proposed scheme with 16-QAM BICMover AWGN channel.

To compare the simulation results with a relevant asymptotic performance limit, the BICM capacity given in (8) is used, which is equal to GMI under MAP detection. The achievable information rate of a given BICM scheme with a certain detection method can be computed using Monte Carlo simulations by adopting the GMI as the performance measure [24]. The BICM capacity in Fig. 7 is derived using Monte Carlo simulations assuming MAP detection and is taken from [25]. The BICM capacities are also provided for other cases later in the paper, which are also from [25].

C. Simulation Results for 16-QAM BICM Over Fast Fading Channel

Here, the simulation results for the performance comparison between the proposed compound scheme and the separated scheme over the fast fading channel are provided. A Rayleigh fast fading channel is considered, in which $y = ax + n$, where x is the input symbol, y is the received symbol, n is the additive complex Gaussian noise, and a is picked from a Rayleigh distribution with $E\{|a|^2\} = 1$. The random variable a is independent across transmissions, and a full channel state information is assumed at the receiver. The parameters of the simulation setup are: $N = 1024$ and rate $= 0.5$, similar to that earlier. The larger block length of $N = 4098$ is also tested for both the compound scheme and the separated scheme. The method of construction is also the same as in the previous simulation. For the case of $N = 1024$, the compound code is constructed based on the estimated probability of error of the individual bit-channels at $E_b/N_0 = 7.5$ dB. The construction for the separated scheme is also done at $E_b/N_0 = 7.5$ dB, and the rates assigned to the two constituent channels are optimized in terms of the total probability of error. The information rate to be transmitted

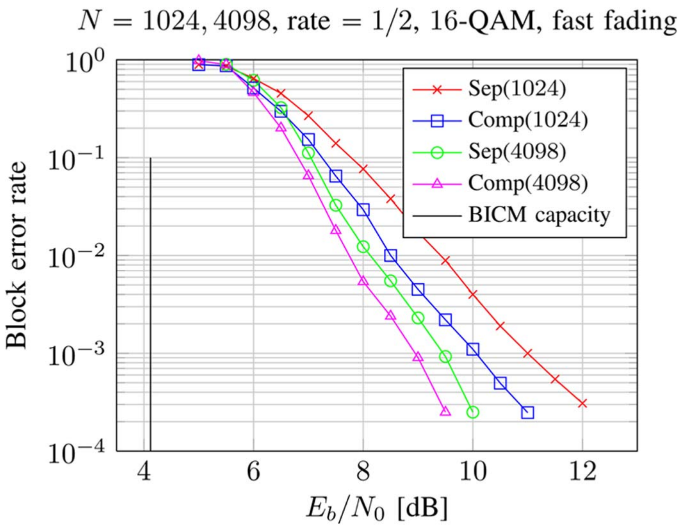


Fig. 8. Performance of the proposed scheme with 16-QAM BICM over fast fading channel.

on the stronger channel is picked as 0.617, and that of the weaker channel is picked as 0.383. For the case of $N = 4098$, the compound code and the separated code are constructed based on the estimated probability of error of the individual bit-channels at $E_b/N_0 = 6$ dB. The rates assigned to the two constituent channels in the separated scheme are optimized in terms of the total probability of error. The information rate to be transmitted on the stronger channel is picked as 0.602, and that of the weaker channel is picked as 0.398. The comparison between the two methods is shown in Fig. 8.

VII. Compound Polar Codes for 3-Multi-channels

Here, the method for choosing the initial matrix in the construction of compound polar transformation is discussed. Then, the simulation results for 64-QAM BICM with transmission over the AWGN channel, which can be modeled as a 3-multi-channel, are provided.

A. Choosing the Initial Matrix G_0 for the Compound Polar Code

Here, our criteria for picking the initial matrix G_0 is discussed for a general value of l, where l is the number of constituent channels. If l is power of two, i.e., $l = 2^m$, then $G_0 = G^{\otimes m}$ is picked. The advantage of this particular G_0 is the resulting low-complexity decoding algorithm. Basically for $N = l2^n$, the Arıkan's SC decoding structure for a polar code of length 2^{n+m} can be exploited with complexity $O(N \log N)$. For other values of l, the method described in [17] and reviewed in Section II is used to pick the $l \times l$ matrix with best polarization rate. This is explained next for the case of $l = 3$.

Based on the aforementioned criteria, the matrix G_0 is picked such that the polarization rate is maximized according to Theorem 3. By a brute-force search, it turns out that $(1/3)\log_3 4 \approx 0.42$ is the maximum polarization rate, which is slightly less than 0.5 of the 2×2 Arıkan's polarization matrix G. The following G_0 is an example that achieves this maximum polarization rate:

$$G_0 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1. \end{bmatrix}$$

It is easy to check that partial distances of G_0 are as follows: $D_1 = 1, D_2 = 2, D_3 = 2$. Then, by Theorem 3, the polarization rate of G_0 is given by

$$\frac{1}{3}(\log_3 1 + \log_3 2 + \log_3 2) = \frac{1}{3}\log_3 4.$$

The polarization building block corresponding to G_0 for transmission over the multi-channel $(W_1 \cdot W_2 \cdot W_3)$ is shown in Fig. 9.

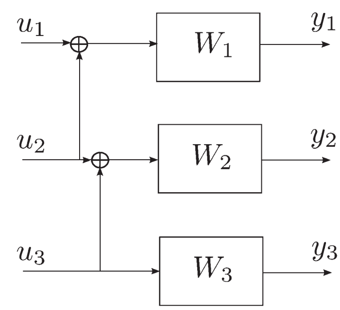


Fig. 9. Proposed building block for 3-multi-channels.

In the SC decoding, the LRs of the building block, which are used to initiate the SC decoding, are computed as follows. Let L_1, L_2, L_3 denote the channel LRs given at the output. Then

$$LR(u_1|y_1,y_2,y_3) = \frac{L_1 + L_2 + L_3 + L_1L_2L_3}{1 + L_1L_2 + L_1L_3 + L_2L_3}$$

$$LR(u_2|y_1,y_2,y_3,u_1) = L_1^{1-2u_1}\frac{1 + L_2L_3}{L_2 + L_3}$$

$$LR(u_3|y_1,y_2,y_3,u_1,u_2) = L_2^{1-2u_2}L_3.$$

B. Simulation Results for 64-QAM BICM

Simulation results of compound polar codes with 64-QAM BICM is provided next, where the polarization building block is as shown in Fig. 9. Notice that 64-QAM BICM with Gray labeling can be approximately modeled as a 3-multi-channel $(W_1 \cdot W_2 \cdot W_3)$. The transmission block length is fixed to $N = 3 \times 2^9 = 1536$, and the transmission is assumed over the AWGN channel. A random S-interleaver of length N is applied to the encoded block followed by a designed π for the polarization purpose. The concatenation of the random interleaver and π map the ith subblock of the encoded codeword of length 512 to the bits $2i$ and $2i+1$ of the symbols, for $i = 1, 2, 3$. The constructed compound polar code is compared with the separated scheme, where assigned rates to the constituent channels in the separated scheme are optimized. It turns out that the following choices for the rates minimize the

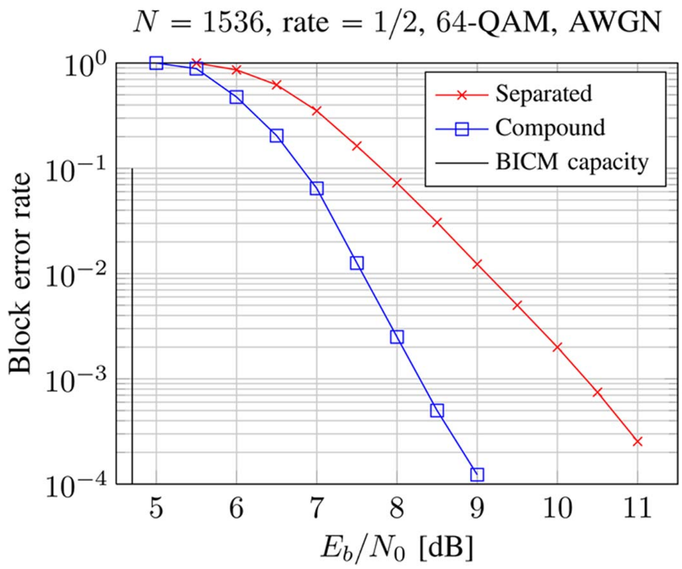


Fig. 10. Performance of the proposed scheme with 64-QAM BICM over AWGN channel.

probability of error at the design SNR ($E_b/N_0 = 7$ dB): $R_1 = 0.24$, $R_2 = 0.53$, and $R_3 = 0.73$, where R_i is the assigned rate for transmission over W_i. The simulation results are shown in Fig. 10. One can observe that, at high SNRs, there is about 2-dB improvement in SNR using the compound polar code.

VIII. Related Works and Comparisons

Here, we provide an overview of the recent relevant work of [13] and [14] and then discuss the differences and similarities. A performance comparison with our compound polar scheme, and the state of the art of turbo codes are provided.

Two frameworks are proposed in [14] for combining binary polar codes with M-ary modulations by partitioning the M-ary channel into binary channels. The first approach, also called sequential binary partition (SBP), is inspired by the MLC idea, and an observation made by Arıkan that the MLC structure is closely related to polar coding on a conceptual level. Let the M-ary DMC channel be denoted by W and $M = 2^m$. Then, an order-m SBP φ of W transforms W into m B-DMCs, also called *bit-channels*. This transformation preserves the symmetric capacity of W, which is referred to as coded modulation capacity in the literature, by following the chain rule of mutual information. The SBP resembles the Arıkan's transformation of polarization kernel into binary bit-channels. A product concatenation of two SBPs φ and ψ, denoted by $\varphi \otimes \psi$, is introduced. Then, the polar transformation of length $N = 2^n$ is considered an N-SBP, which is denoted by κ^n, and the concatenation $\varphi \otimes \kappa^n$, of total length mN, is used for constructing *polar coded modulation*. We refer to this scheme as the SBP-based scheme.

As mentioned in [14], one potential drawback of the SBP-based scheme for practical applications is the necessity for using several short component codes, i.e., m codes, whose code rates is varying depending on the bit levels. The component codewords need to be decoded sequentially at the receiver, i.e., the hard-decoded information of previous codewords will be used to demodulate and to decode the next codewords. Hence, this scheme cannot be applied to BICM schemes that require a single codeword and parallel LLR calculation. Our construction of compound polar codes also can not be used under the SBP scenario. However, the advantage of the SBP-based scheme is that the channel polarization theorem for each of the bit-channels results in the achievability of coded modulation capacity for the total scheme.

In the second framework, which is more suitable for BICM application, a parallel binary partition (PBP) of W is considered. We refer to the second scheme as the PBP-based scheme. In this case, an m-order PBP $\overline{\varphi}$ of W is mapping W into m mutually independent B-DMCs whose sum of symmetric capacities is in general smaller than that of W. This operation, when, for instance, applied to 16-QAM, becomes similar to our multi-channel model described in Section II, as well as that of our previous construction in [9]. It is assumed that m is a power of 2 and that the polar transformation $\kappa^{n+\log m}$ is concatenated with a PBP $\overline{\varphi}$. It is argued in [14] that the resulting transform can be represented as $(\overline{\varphi} \odot \kappa^{\log m}) \otimes \kappa^n$, where $\overline{\varphi} \odot \kappa^{\log m}$ is described as a *degraded* SBP. Furthermore, a method for empirically optimizing the bit labeling rule is proposed in [14] by maximizing the variance of bit-channels' symmetric capacities.

In contrast, the compound polar code construction, proposed in this paper, is guaranteed to polarize on each constituent channel of the multi-channel. Further polarizing the constituent channels together via the polarization building block gives an extra step of polarization. This results in improved finite-length performance compared with the PBP construction, which did not discuss the joint polarization of the constituent channels. This is verified by simulating a PBP scheme based on the polarization structure demonstrated in [14, Fig. 6]. The simulation setup is the same as the setup of Section VI-B, and Gray labeling is assumed for bit-to-symbol mapping. A polar code of rate 0.5 and length 1024 is constructed using the Monte Carlo simulation, assuming the underlying polar modulation transform of the PBP scheme. The simulation results are shown in Fig. 11. It can be observed that our compound polar code offers a better performance of up to 0.5-dB SNR gain.

It is well known that different bit-to-symbol mappings affect the performance of BICM systems [26]. Note that, although the proposed constructions assume Gray labeling, our compound polar scheme and the interleaver π can be modified for a different labeling after identifying the indexes of the different BICM channels with different reliabilities.

For the PBP-based scheme of [14], it is assumed that m is a power of 2. It is suggested in [14] that, if m is not a power of 2, then a different polarization kernel will be used. One drawback of this approach, as mentioned in Section III-B, is the exponentially increased decoding complexity with respect to the size of the polarization kernel. Instead, we proposed in Section III-B to combine an initial matrix G_0, in this case of size $m \times m$, with n levels of 2×2 polarization kernels. This scheme is applied to the 64-QAM modulation as shown in Section VII, where $m = 6$ is not a power of 2.

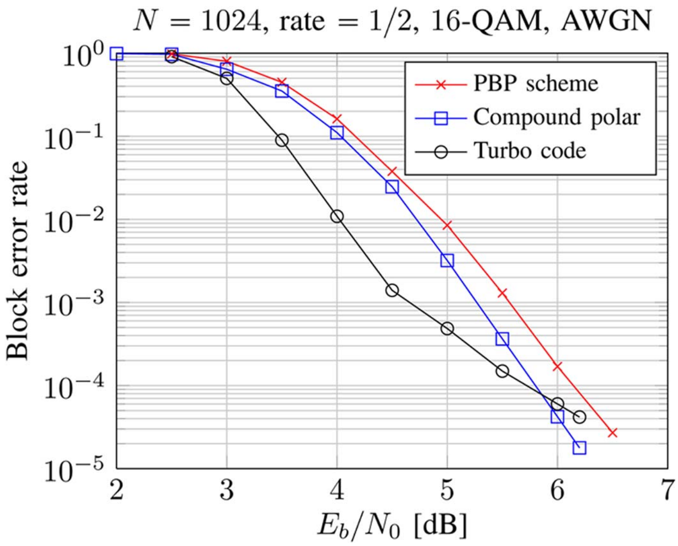


Fig. 11. Performance comparison between the proposed compound polar code and other schemes.

The performance of the proposed compound polar code is also compared with LTE turbo codes at the same rate and block length. The turbo code is considered with the information block length of 512 and the puncturing that makes the code block length 1024 and the rate 0.5. The simulation results are shown in Fig. 11, which indicate that, in the waterfall region of turbo code, the polar code falls short, which is widely known. Multiple methods have been proposed to improve this performance that can be applied also on top of the compound polar code, which are discussed in Section VIII. However, the advantage of polar codes is that they do not show an error floor, which makes them beat the turbo code at its error floor region where the frame error rate is around 10^{-4}.

IX. Conclusion and Future Work

In this paper, we have proposed a compound polar scheme to be used over a set of parallel channels, which are called multichannels. We extended the channel polarization theorem for multi-channels and proposed a framework for the construction of polar-based schemes for multi-channels, which are named compound polar codes. In particular, BICM channels can be approximately modeled as multi-channels. We constructed compound polar codes for 16-QAM and 64-QAM BICM and simulated the performance over AWGN and fast fading channels. The compound polarization scheme showed considerable performance gains over the conventional separated polarization scheme.

Notice that the existing methods for improving the performance of single-channel polar codes can be applied on top of our proposed compound polar codes. For instance, the encoders of individual polar codes can be made systematic as suggested by Arıkan [27] to improve the bit error rate. Moreover, concatenation of polar codes with other block codes [28] or list decoding of polar codes [29] can be used to boost the finite-length performance. Multidimensional decoding methods, as proposed in [30], or the relaxed polarization method [31] can also be extended for compound polar codes to reduce their decoding latency and computational complexity. Furthermore, BICM with iterative decoding, referred to as BICM-ID (cf. [26]), can improve the performance of the BICM compound polar coded scheme.

There are some open problems and directions for future work. For the code construction in this paper, a simulation-based Monte Carlo method is used to pick the best bit-channels. However, the question is how to extend Tal-Vardy method [32] for the construction of polar codes to construct compound polar codes for transmission over BICM channels. Another interesting problem is what is the best channel ordering in a multi-channel, as shown in Fig. 4, that results in the best performance of the corresponding compound polar code. As discussed before, for the case of 2-multi-channels, the ordering does not matter, which may not be the case for an arbitrary number of constituent channels l. Another open problem is regarding the best choice for the initial $l \times l$ polarization matrix G_0. In this paper, our criterion was to pick G_0 that results in the best polarization rate when the same G_0 is used for the rest of channel polarization steps. However, in our proposed scheme for compound polar codes, Arıkan's 2×2 polarization matrix G is used for channel polarization on top of the polarization building block determined by G_0. Note that the proof of channel capacity achieving property of our scheme is independent of the choice of G_0; however, the finite-length performance of the scheme is affected by this choice. However, for Arıkan's construction, G_0 may be chosen to maximize the polarization exponent [17]; other choices of G_0 need to be explored for the compound polar codes construction.

Appendix
Proof of Lemma 4

Let $U_1, U_2, \ldots, U_l$ be l independent uniform binary random variables and $X_1^l = U_1^l \cdot G_0$. Since G_0 is invertible, $X_1, X_2, \ldots, X_l$ are also l independent uniform binary random variables. Let Y_1^l denote the outputs of the channels $W_1, W_2, \ldots, W_l$ with inputs $X_1, X_2, \ldots, X_l$, respectively, as shown in Fig. 4. Then, the proof follows from the following series of equalities:

$$\begin{aligned}\sum_{i=1}^{l} \mathcal{C}(W_i) &= \sum_{i=1}^{l} I(X_i; Y_i) = I(X_1^l; Y_1^l) = I(U_l^1; Y_1^l) \\ &= \sum_{i=1}^{l} I(U_i; Y_1^l | U_1^{i-1}) \\ &= \sum_{i=1}^{l} I(U_i; Y_1^l, U_1^{i-1}) \\ &= \sum_{i=1}^{l} \mathcal{C}(\dot{W}_i)\end{aligned}$$

where we used the chain rule on the mutual information and the fact that W_i and $\dot{W}_i$ are all symmetric and the symmetric capacity is therefore equal to the capacity for each of them. ■

Proof of Lemma 5:

$$\begin{aligned}
&(W_1 \cdot W_2 \cdots W_l)_N^{(2i)} \\
&= \frac{1}{2^{N-1}} \sum_{u_{2i+1}^N} \widetilde{W}_{N/2}\left(y_{1,o}^N | u_{1,o}^N \oplus u_{1,e}^N\right) \widetilde{W}_{N/2}\left(y_{1,e}^N | u_{1,e}^N\right) \\
&= \frac{1}{2} \frac{1}{2^{N/2-1}} \sum_{u_{2i+1,e}^N} \widetilde{W}_{N/2}\left(y_{1,e}^N | u_{1,e}^N\right) \\
&\quad \cdot \frac{1}{2^{N/2-1}} \sum_{u_{2i+1,o}^N} \widetilde{W}_{N/2}\left(y_{1,o}^N | u_{1,o}^N \oplus u_{1,e}^N\right) \\
&= (W_1 \cdot W_2 \cdots W_l)_{N/2}^{(i)} \circledast (W_1 \cdot W_2 \cdots W_l)_{N/2}^{(i)}
\end{aligned}$$

where we only used the definitions of bit-channels in (12) and the channel operation $\circledast$ in (11). Using similar arguments, the other equation can be derived. ■

Proof of Lemma 6: The proof is by induction on n. The base of induction, $n = 0$ and $i = j$ is clear by definition of $\dot{W}_i$. Suppose that it holds for n. For $1 \le i \le N = l2^n$, let $i = (j-1)2^n + r$ with $1 \le r \le 2^n$. Then, $2i = (j-1)2^{n+1} + 2r$ with $1 \le r \le 2^{n+1}$, and we have

$$\begin{aligned}
&(W_1 \cdot W_2 \cdots W_l)_{2N}^{(2i)} \\
&= (W_1 \cdot W_2 \cdots W_l)_N^{(i)} \circledast (W_1 \cdot W_2 \cdots W_l)_N^{(i)} && (20) \\
&= [\dot{W}_j]_{2^n}^{(r)} \circledast [\dot{W}_j]_{2^n}^{(r)} && (21) \\
&= [\dot{W}_j]_{2^{n+1}}^{(2r)}. && (22)
\end{aligned}$$

Equation (20) follows by Lemma 5, (21) is the induction hypothesis, and finally, (22) follows by definition of channel combining operation. Similarly

$$\begin{aligned}
&(W_1 \cdot W_2 \cdots W_l)_{2N}^{(2i-1)} \\
&= (W_1 \cdot W_2 \cdots W_l)_N^{(i)} \boxast (W_1 \cdot W_2 \cdots W_l)_N^{(i)} \\
&= [\dot{W}_j]_{2^n}^{(r)} \boxast [\dot{W}_j]_{2^n}^{(r)} \\
&= [\dot{W}_j]_{2^{n+1}}^{(2r-1)}.
\end{aligned}$$

This proves the induction hypothesis for $n+1$, which completes the proof of the lemma. ■

REFERENCES

[1] E. Arıkan, "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.

[2] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

[3] E. Arikan, "Source polarization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 899–903.

[4] E. Abbe, "Randomness and dependencies extraction via polarization," in *Proc. ITA Workshop*, 2011, pp. 1–7.

[5] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012.

[6] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Techniques for polar coding over multiple access channels," in *Proc. 48th Annu. CISS*, Princeton, NJ, USA, 2014, pp. 1–6.

[7] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.

[8] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[9] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Compound polar codes," in *Proc. ITA Workshop*, San Diego, CA, USA, Feb. 2013, pp. 1–6.

[10] J. J. Boutros and E. Biglieri, "Polarization of quasi-static fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 769–773.

[11] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.

[12] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[13] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Multilevel polar-coded modulation," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1302–1306.

[14] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Polar-coded modulation," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4108–4119, Oct. 2013.

[15] E. Arikan and E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 1493–1495.

[16] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytech. Féd. Lausanne, Laussane, Switzerland, 2009.

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

[18] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2-QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, Jul. 2009.

[19] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. M. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, Jun. 2009.

[20] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 1/2, pp. 1–153, 2008.

[21] E. Rosnes and O. Ytrehus, "On the design of bit-interleaved turbo-coded modulation with low error floors," *IEEE Trans. Commun.*, vol. 54, no. 9, pp. 1563–1573, Sep. 2006.

[22] A. Alvarado, E. Agrell, L. Szczecinski, and A. Svensson, "Exploiting UEP in QAM-based BICM: Interleaver and code design," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 500–510, Feb. 2010.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[24] T. T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 437–447, Feb. 2011.

[25] A. Yedla, M. El-Khamy, J. Lee, and I. Kang, "Performance of spatially-coupled LDPC codes and threshold saturation over BICM channels," *arXiv preprint arXiv:1303.0296*, to be published.

[26] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.

[27] E. Arıkan, "Systematic polar coding," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 860–862, Aug. 2011.

[28] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Performance limits and practical decoding of interleaved reed-solomon polar concatenated codes," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1406–1417, May 2014.

[29] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 1–5.

[30] H. Mahdavifar, M. El-Khamy, J. Lee, and I. Kang, "Fast multi-dimensional polar encoding and decoding," in *Proc. ITA Workshop*, San Diego, CA, USA, Feb. 2014, pp. 1–5.

[31] M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang, "Relaxed channel polarization for reduced complexity polar coding," in *Proc. IEEE WCNC*, New Orleans, LA, USA, Mar. 2015, pp. 219–224.

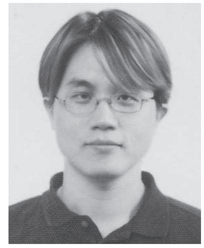
[32] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

Hessam Mahdavifar (S'10–M'12) received the B.Sc. degree from Sharif University of Technology, Tehran, Iran, in 2007 and the M.Sc. and Ph.D. degrees from the University of California, San Diego, CA, USA, in 2009 and 2012, respectively, all in electrical engineering.

Since July 2012, he has been with Modem R&D, Samsung Electronics, San Diego, first as a Senior Research Engineer and then as a Staff Research Engineer. His main research interests include coding and information theory with applications to privacy, storage, and wireless communications.

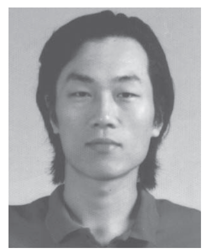
Mostafa El-Khamy (SM'12) received the B.Sc. and M.Sc. degrees from Alexandria University, Alexandria, Egypt, in 1999 and 2001, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology (VirginiaTech), Pasadena, CA, USA, in 2003 and 2006, respectively.

From 2006 to 2008, he was with Qualcomm R&D, San Diego, CA, where he designed new femtocell technologies for Third-Generation Partnership Project systems. Since 2008, he has been with the faculty of the Department of Electrical Engineering, Alexandria University, where he is currently an Adjunct Associate Professor. He has also held visiting faculty positions with VirginiaTech-MENA, Alexandria; University of Ottawa, Ottawa, ON, Canada; the German University in Cairo, New Cairo, Egypt; and Egypt–Japan University for Science and Technology, Alexandria. Since 2011, he has been with Modem R&D, Samsung Electronics, San Diego, where he is currently a Principal Engineer working on the design and research of wireless modem technologies.



Jungwon Lee (S'00–M'05–SM'12) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2005.

From 2003 to 2010, he was with Marvell Semiconductor Inc., as a Principal Engineer, where he developed HD Radio, Bluetooth, Wi-Fi, WiMAX, and Long-Term Evolution chips. Since 2010, he has been with Modem R&D, Samsung Electronics, San Diego, CA, USA, where he is currently a Senior Director with Samsung Mobile Solutions Laboratory, in charge of a systems group for second-, third-, and fourth-generation cellular modem and Global Navigation and Satellite System chips. He is also currently an Associate Adjunct Professor with the University of California, San Diego. He is the author or coauthor of more than 80 papers and holds over 190 U.S. patents.



Inyup Kang (M'11) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1985 and 1987, respectively, and the Ph.D. degree in electrical engineering from the University of California, Los Angeles, CA, USA, in 1996.

From 1996 to 2009, he was with Qualcomm, San Diego, CA, USA, where he led generations of cellular modem chipsets as Vice President of Technology. Since 2010, he has been working with SOC Development, System LSI, Samsung Electronics, Yongin, Korea, as Executive Vice President in charge of R&D and commercialization of cellular baseband/radio-frequency chipset and software for second, third, and fourth generations. His main research interests include the fundamental and theoretical aspects of cellular wireless communications.