



Performance Analysis of Bit-Interleaved Polar Coded Modulation

Kun Zhao^{1,2,3} · Yong Liu^{1,2,3} · Haiqing Du^{1,2,3} · Baoku Yuan^{1,2,3}

Published online: 11 June 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Polar codes have been applied to the construction of bit-interleaved polar coded modulation schemes in recent years. Bit-interleaved polar coded modulation schemes consist of the concatenation of multi-order modulation and polar codes. The structure of polar codes has some similar features to bit-interleaved coded modulation schemes. This similar features motivate us to transform the concatenated transmission schemes into a universal equivalent channel model, which consists of independent parallel channels, through considering polar coding and modulation synthetically in practical applications. We also propose a general polar codes constructing algorithm to design the constituent polar codes for the equivalent channel model. Then we employ a bijective mapper to accomplish the modulation i.e., binary address mapping, from coded bits to signals in constellation with Gray labeling or set partition labeling rule in different order modulations. We analyze and compare the performance of the equivalent channel model with different order modulations under different decoding algorithms. Simulation results show that the performance of our proposed schemes outperforms that of low-density parity-check codes in WiMAX standard and Turbo codes.

Keywords Bit-interleaved coded modulation · Polar codes · Equivalent channel model · Polar codes constructing algorithm

✉ Haiqing Du
duhaiqing@bupt.edu.cn

Kun Zhao
zk2015@bupt.edu.cn

Yong Liu
liuyo@bupt.edu.cn

Baoku Yuan
yuanbaoku@bupt.edu.cn

¹ Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China

² Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China

³ School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

1 Introduction

In 1982, Ungreboeck proposed that coding and modulation can be jointly considered as an entity to improve the performance of communication system in his paper on trellis-coded modulation [1]. Then this concept was developed into a lot of versions in coded and modulation application fields. In [2], Zehavi proposed a theory of replacing the symbol-wise interleaver in Ungreboeck's method with a bit-wise interleaver that implements binary encoding and decoding independently at both sides of channels through decomposing the non-binary input channel into binary channels. The result shown that as a suboptimal (assuming independent between each bit channel) scheme, bit-interleaved coded modulation (BICM) could offer significant advantages at high signal to noise ratio (SNR) region with low computational complexity. At the same time, the encoding and modulation procedures of BICM can be designed simply and independently as they are detached by a bit-wise interleaver [3]. With those remarkable advantages, BICM has been a popular scheme in modern communication systems.

Polar codes are famous for their channel coding properties of low complexity and binary capacity achieving [4], where the capacity-achieving property is an asymptotic character of polar codes with infinite coding length, which is applied into binary-input discrete memoryless symmetric (B-DMS) channels. As the performance of successive cancellation (SC) decoding is disappointing for short length of polar codes, Tal and Vardy proposed the list decoding algorithm for polar codes, i.e., successive cancellation list (SCL) decoding algorithm, to improve the performance of polar codes with moderate coding length. SCL decoding algorithm can completely achieve the maximum likelihood (ML) performance of polar codes at the practical insight [5], which attract more interest in channel coding fields and make polar codes competitive with the state-of-the-art codes i.e., LDPC [6] and Turbo codes [7]. Meanwhile, since iterative belief propagation decoding algorithm has parallel executing property inherently, this decoding has also attracted a lot focus on hardware implement [8–10].

In the following years, polar codes had been widely applied in coded modulation communication system, especially, the modulation of 2^m -order digital modulation system [11, 12], which would not increase spectral efficiency of binary polar coded modulation in some degree. With the proposition of polar coded modulation, BICM has attracted a lot focus again. In [13], authors focused on the polar coded BICM and designed a kind of interleaver which outperformed others with the employed mappings.

Except for those competitive performance, since polar codes can offer the flexibility for practical application in accommodating multilevel coding rates and various transmission channels, polar codes have been selected as official coding scheme for the control channel in the fifth generation (5G) enhanced mobile broadband (eMBB) scenario [14, 15]. However, there are few works on the performance analysis and design of synthetically considering multilevel modulation and polar codes under SC and SCL decoding algorithms for an equivalent channel model with a general polar codes construction in BIPCM system.

In this paper, we aim at offering a general equivalent channel model of BICM with multi-order modulation and universal flexible polar coding construction to optimize the information transmission problem of non-binary discrete, memoryless channels. The main contributions can be listed in the following several parts:

- Through observing a group of BIPCM schemes and the structure features of polar codes, we find that BIPCM can be transformed to an equivalent channel model consist-

ing of several independent parallel channels by considering modulation and polar coding synthetically. We analyze the channel equivalent characters and analytically prove that the equivalent channels are capacity-achieving.

- We propose a general polar codes constructing algorithm and design component polar codes for the equivalent channel model applying the proposed code constructing algorithm to independent parallel channels.
- We use general bijective mapping rules to map coded bits to signal points deriving from various order modulation constellations over the equivalent channel model. Detail simulation results shown the error-correcting performance of different synthetic schemes consisting of multi-order modulation with various decoding algorithms.

We analyze the time and space complexities of those BIPCM schemes. We use simulation results to compare the performance of different schemes consisting of concatenations of polar codes with multi-order modulation with SC, SCL, and log likelihood ratio based (LLR-based) SCL decoding algorithms through AWGN channels.

The construction of this paper consists of the following sections: In Sect. 2, the preliminary of the overall system schemes is introduced, which includes symbol notation, polar codes definition and mapping rules. In Sect. 3, system models of BIPCM and a general equivalent channel model are built, then we analyze and derive the capacities of both type channel models and design the equivalent BIPCM applying the general code constructing algorithm under coding rate constraint. And then we analyze the detail decoding processes of SC, SCL, and LLR-based SCL decoding algorithms, meanwhile, compare the time complexity and space complexity of those decoding algorithms. In Sect. 4, numerical results illustrate that the error-correcting performance of schemes over the equivalent channel model with SCL decoding algorithm outperforms that of LDPC in WiMAX standard and Turbo codes. In addition, we compare the reliability performance of the propose scheme with that of several other schemes under different order modulation. In Sect. 5, we show the summation of this paper.

2 Preliminary

Let $C \triangleq \{1, 2, \dots, N\}$ denote an integer set. The cardinality of set C is represented by $|C| = N$. Boldface upper case letters represent matrix while boldface lower case letters denote vectors. Lower case letters without boldface represent the constant parameters. Upper case letters without boldface denote set or channel depending on the practical applicative scenarios. We set $\mathbf{u}_p^q = [u_p, u_{p+1}, \dots, u_q]$, ($0 \leq p < q$). Set $\mathbf{F} \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ to denote the underlying kernel matrix of polar codes. Matrix dimension denotes by (m, n) .

2.1 Polar Codes

Let the quad tuple $(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ present polar codes, where N denotes the length of polar codes, K represents the length of information bits in polar codes, subsets $\mathcal{A} \subset C$ denote the index sets of information bit channels which are used to transmit information bits and with cardinality $|\mathcal{A}| = K$. Vectors $\mathbf{u}_{\mathcal{A}^c}$ denote the frozen bit sequences. The index sets of frozen bit sequences denote by \mathcal{A}^c with cardinality $|\mathcal{A}^c| = N - K$. In general, we set N to be a power of 2. The number of bit layers of polar codes denotes by $n = \log_2(N)$. The generator matrix of

polar codes define by $\mathbf{G} = \mathbf{B}^N \mathbf{F}^{\otimes n}$, where \mathbf{B}^N presents bit reversal permutation matrix. The given information sequences denote by $\mathbf{u}_A = \mathbf{u}_1^K$ with the informational index sets A in vector \mathbf{u}_1^N . The polar codes are generated by $\mathbf{v}_1^N = \mathbf{u}_1^N \cdot \mathbf{G}$. Binary field denotes by \mathcal{F}_2 .

2.2 General Mapping

In order to generate more reliable and robust Euclidean space codes for the given 2^m -order signal constellation, each signal point is assigned a label consisting of m coded bits. Especially, we assume that $M=2^m$ denotes the order of constellation and each label presented by m -bit digitals. In this paper, we apply a bijective mapping function $x = f(\mathbf{v}_1^m)$ to map binary address index vector $\mathbf{v}_1^m = (v_1, \dots, v_{m-1}, v_m)$, $v_i \in \{0, 1\}$, $1 \leq i \leq m$, into a responding signal point of input alphabet $x \in \mathcal{X}$. Generally, the mapping comes from the procedure of partitioning the signal set \mathcal{X} into several subsets in succession [16].

Labeling rule transforms a channel with 2^m -ary input into m parallel bit channels, which simplifies the processes of encoding and decoding for non-binary channel, i.e., binary encoding and decoding replace that of non-binary channel directly. We just consider two kinds of labeling rules: Gray labeling and set partition labeling. Since Gray labeling could maximize the capacity of BICM transmission scheme [3], we give priority to employ Gray mapping in this paper.

In [17], authors pointed out that natural binary labeling can be transformed to binary reflective Gray labeling using linear transformation which also suits for the reversed direction of one dimensional constellations. To make the best use of set partitioning (SP) and Gray labeling rule, we have:

$$\mathbf{M}_{SP,3} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{M}_{Gray,3} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

The index of each constellation point can be represented by a given number bits. For example, a given 2^m -order ASK (2^m -ASK) constellation is characterized by a binary (M, m) matrix $\mathbf{M}_{SP,m}$ where m bits have the dual representation for the row indices among the set $\{0, \dots, 2^m - 1\}$. Specially, the right-most column denotes the most significant bits.

In [17], authors shown that set partitioning and Gray labeling rules could interconvert with each other. In other words, set partitioning labeling could translate into a binary Gray labeling over a $2^m - \text{ASK}$ constellation by multiplying an m ordered binary square matrix \mathbf{T}_m .

$$\mathbf{T}_m = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & 1 \end{bmatrix} \tag{2}$$

So the general linear mapping is derived by:

$$\mathbf{M}_{Gray,m} = \mathbf{M}_{SP,m} \cdot \mathbf{T}_m \tag{3}$$

3 Bit-Interleaved Polar Coded Modulation

Figure 1 depicts the overall operation on the input information vector \mathbf{u}_1^K , where $\mathbf{u}_1^K = [u_1, \dots, u_i, \dots, u_K]$, $u_i \in \{0, 1\}$ denotes the input information vector at the transmitter. The generator matrix of encoder is G_N . Let the notation $W : \mathcal{X} \rightarrow \mathcal{Y}$ denote a DMC with input random variable $X \in \mathcal{X}$ where \mathcal{X} is alphabet set with size of 2^m , and output random variable $Y \in \mathcal{Y}$ with finite size alphabet set \mathcal{Y} and $|\mathcal{Y}| < \infty$.

The mutual information between input X and output Y denotes by $I(X;Y)$. Additionally, x and y denote the realizations of input random variable X and output random variable Y , respectively. For AWGN channel, the original channel W can be modeled by

$$Y = X + N \tag{4}$$

where parameter $N \sim \mathcal{N}(0, \sigma^2)$ denotes Gaussian random variable. We set input X with the probability mass function (PMF) $p_X(x)$ of the input symbols $x \in \mathcal{X}$.

Such a channel with discrete outputs can be represented by the conditional transition PMF $p_{Y|X}(y|x)$. The mutual information of this channel is defined by

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_X(x) p_{Y|X}(y|x) \log_2 \frac{p_{Y|X}(y|x)}{p_Y(y)} \tag{5}$$

where $p_Y(y) = \sum_{x \in \mathcal{X}} p_X(x) p_{Y|X}(y|x)$. Specially, there is no analytic methods to compute $I(X;Y)$, and the results can be achieved by applying numerical integration methods.

Channel with continuous output alphabet \mathcal{Y} with infinite size could be represented by a probability density function (pdf) $f_{Y|X}(y|x)$. The mutual information of this channels can be defined by

$$I(X;Y) = \int_{-\infty}^{+\infty} \sum_{x \in \mathcal{X}} p_X(x) f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)} dy \tag{6}$$

As we all know, mutual information of a symmetric DMC maximize by assuming input random variable X selected with uniform distribution over input alphabets \mathcal{X} . Then the mutual information can be further written as

$$\begin{aligned} I(X;Y) &= \int_{-\infty}^{+\infty} \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)} dy \\ &= I(W) \end{aligned} \tag{7}$$

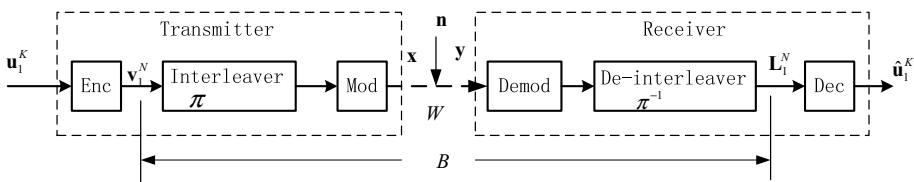


Fig. 1 BIPCM architecture

For the fact that the capacity of a symmetric DMC with uniform input is equal to its mutual information, we obtain the capacity of such a channel:

$$C(W) = \max_{p_X(x)} I(X;Y) = I(W) \tag{8}$$

Similarly, non-binary discrete memoryless channels can be partitioned into several symmetric binary discrete memoryless channels [12]. If we define a binary input symmetric discrete memoryless channel $B : \mathcal{X}' \rightarrow \mathcal{Y}'$, where $\mathcal{X}' = \{0, 1\}$. The m copies of B denote by B^m i.e., vector channel $B^m = \{B \times B \times \dots \times B\}$, which expresses the operating procedure from the output of encoder to the input of decoder. The vector channel can be partitioned into m ordered bit levels [18] $\eta : B^m \rightarrow \{B^{(1)}, B^{(2)}, \dots, B^{(m)}\}$ with set partition (SP) or binary reflected Gray labeling rules.

3.1 Encoding and Decoding of BIPCM

First, at the transmitter, input information vector is encoded by standard polar codes encoder, i.e., information vector \mathbf{u}_1^K multiplies the generator matrix G_N of encoder which generates polar codes \mathbf{v}_1^N with code length N ($N = 2^n$). Polar codes are interleaved by a bitwise interleaver, which was designed in [13]. Then those interleaved bit streams are separated into plenty of subsequences of m bits each. Moreover, each m bits are mapped into signal points in \mathcal{X} , $|\mathcal{X}| = 2^m$ with the labeling mapping $\alpha : \{0, 1\}^m \rightarrow X$, e.g. the general mapping. The linear mapping characterization can denote by matrix operations, i.e., Eqs. (1–3). Labeling positions $\mathbf{b} = (b_1 \dots b_j \dots b_m)$ of symbol x transmit the polar codes through the corresponding bit channel $B^{(j)}$, $j \in \{1, 2, \dots, m\}$ independently.

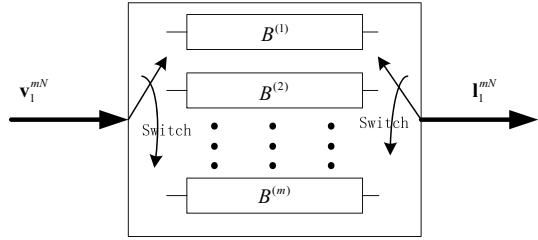
At the receiver, we use $\mathcal{X}_b^j = \{x \in \mathcal{X} : b_j(x) = b\}$ to denote the sets of constellation points with a binary labeling value $b_j \in \{0, 1\}$ at the j -th labeling position. Moreover, we apply the general mapping to accomplish the mapping from coded bits to indices of signal points with Gray or set partition labeling in corresponding constellation i.e., modulation. Then we obtain symbol sequences after modulation. Next, the symbol sequences are sent into AWGN channel W . During this transmission, the symbol sequences are always corrupted by noise e.g. AWGN. Generally, receiver should take some measures to eliminate those noise or interference. After symbol sequences are transmitted over wireless channels, the demodulator calculates the probabilities of labeling positions $b_j(x)$ basing on the received symbols. In generally, the probabilities will be converted to log-likelihood ratios (LLRs) under the assumption that the transmitted prior value is 0 or 1. Finally, those LLRs values are de-interleaved and sent to decoders which implement several decoding algorithms and output the estimates of information sequence $\hat{\mathbf{u}}_1^K$.

3.2 Equivalent Channel Model

According to the mutual information definition, the capacity of equivalent channels of BIPCM can be achieved by the proof in [19, 20]. Considering the coding construction of polar codes and BICM system, we transform the BIPCM model into m independent parallel equivalent binary input channels $\mathbf{B} = \{B^{(1)}, \dots, B^{(m)}\}$ in Fig. 2.

Let coded bit sequence \mathbf{v}_1^{mN} be interleaved by bit-wise interleaver $\pi : i \rightarrow i'$ where index $i = (j - 1) \cdot N + k$ and $i' = k \cdot 2^m + j$, i.e., $i' = \pi(i)$ for all $1 \leq j \leq m$, $1 \leq k \leq N$. The interleaved bit streams are separated into groups of subsequence with m bits each.

Fig. 2 Block graph of the equivalent channel model of the vector channels B^m in Fig. 1



Binary labeling mapping $\alpha : \{0, 1\}^m \rightarrow \mathcal{X}$, accomplishes the mapping operations step by step. Each label position $b_j(x)j \in \{1, 2, \dots, m\}$ is corresponding to the bit-wise index of symbol x which distributes uniformly over the subset of modulation constellation with the fixed label position j , i.e., $b_j(x)$ denotes a realization of Bernoulli random variable with probability $1/2$ in the j -th equivalent channel $B^{(j)}$.

$$\mathbf{I}_1^{mN} = \{llr_1(v_1), \dots, llr_i(v_i), \dots, llr_{mN}(v_{mN})\} \tag{9}$$

$$llr_i(v_i) = \log_e \frac{B(\mathbf{y}, \hat{v}_1^{i-1} | v_i = 0)}{B(\mathbf{y}, \hat{v}_1^{i-1} | v_i = 1)} \tag{10}$$

where vector $\mathbf{y} = (y_1, y_2, \dots, y_N)$ denote the received symbol sequences which are also the input vector of demodulator in Fig. 1.

Meanwhile, we set vector $\mathbf{x} = (x_1, \dots, x_N)$ as the output symbol vector of modular. Demodular demodulates the received symbol sequence by the following ‘‘argmax’’ rule:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \tag{11}$$

where notation $q(\mathbf{x}, \mathbf{y})$ denotes the decoding metric, which is defined by the product of symbol metrics $q(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^N q(x_k, y_k)$. The decoder is equal to a function of denotation from output alphabet \mathcal{Y} to the sequence space which estimates the transmitted sequence $\hat{\mathbf{x}}$ in (11).

The nature model of the symbol decoding metrics is defined by:

$$q(x_k, y_k) = \prod_{j=1}^m p_j(b_j(x_k), y_k) \tag{12}$$

$$p_j(b_j(x_k), y_k) = \sum_{x' \in X_j^p(x_k)} p(y_k | x') \tag{13}$$

$$p_j(y_k | b_j(x_k)) = p_j(b_j(x_k), y_k) \cdot p_j(b_j(x_k)) \tag{14}$$

where $q(x_k, y_k)$ denotes symbol probability of equivalent channels W with input symbol x_k and output symbol y_k . Notation $p_j(b_j(x_k), y_k)$ denotes the crossover probability of the j -th label position $b_j(x_k)$ in symbol x_k with the given received symbol y_k . Especially, we do not consider the mutual information between joint label positions in one symbol. Term $p_j(y_k | b_j(x_k))$ denotes the prior probability of received symbol y_k corresponding to $b_j(x_k)$ at the j -th label position. The form $p_j(b_j(x_k))$ denotes the distribution probability of $b_j(x_k)$.

The capacity of channel W can be rewritten by:

$$\begin{aligned}
 I(W) &= I(X;Y) \\
 &\leq \sum_{j=1}^a I(B^{(j)}) \\
 &= \sum_{j=1}^m \sum_{k=1}^N \sum_{x_k \in \mathcal{X}} \sum_{y_k \in \mathcal{Y}} p_j(b_j(x_k)) p_j(y_k | b_j(x_k)) \\
 &\quad \times \log_2 \frac{p_j(y_k | b_j(x_k))}{\sum_{b_j(x_k)=0}^1 p_j(y_k | b_j(x_k))}
 \end{aligned} \tag{15}$$

where $I(W)$ denotes the mutual information of channel W with input alphabet \mathcal{X} and output alphabet \mathcal{Y} . Through chain rule of mutual information [21], we obtain the inequality relation (a) in (15).

At receiver, demodulator receives the symbol sequences, and then employs equations in Eqs. (13, 14) and (16, 17) to calculate the probabilities $p_j(b_j(x_k), y_k)$ of each labeling position without considering the mutual information between bit levels in each received symbol. The log likelihood ratios of the prior probability of label position $b_j(x_k)$ with the input symbol x_k and output symbol y_k . According to equality (10), the following formula holds:

$$llr(y_k | b_j(x_k)) = \log \frac{p_j(y_k | b_j(x_k) = 0)}{p_j(y_k | b_j(x_k) = 1)} \tag{16}$$

where $llr(y_k | b_j(x_k))$ will feed back to the input of de-interleaver. Then we implement the interleaving operation for all $llr(y_k | b_j(x_k))$. Thus, we have the following relation:

$$llr_{\pi(i)}(v_{\pi(i)}) = llr(y_k | b_j(x_k)) \tag{17}$$

where $\pi(i) = k \cdot 2^m + j$. After the de-interleaved operation, we obtain the LLRs sequence \mathbf{I}_1^{mN} .

Then demodulator calculates the decoding metrics, which are sent to de-interleaver forming a length mN LLRs sequence \mathbf{I}_1^{mN} . Finally, the estimates \mathbf{u}_1^N of input message can be retrieved by decoders. The process of decoding is described in detail in the following section.

3.3 Design of BPCMC Basing on Equivalent Channel Model

Firstly, we need to generate an efficient code construction. In general, we could apply channel polarization (Bhattacharyya parameter estimation), Monte Carlo method and Gaussian approximation (GA) to choose the index set \mathcal{A} of information bit channels. Let $(N, K^{(j)}, \mathcal{A}^{(j)}, \mathbf{u}_{\mathcal{A}^{(j)c}})$ denote the constituent polar codes for the j -th bit level of the equivalent channel model in Fig. 2.

We set B to denote the underlying channel for each bit level (also called equivalent channel) with $\mathbf{B} = \{B^{(1)}, \dots, B^{(m)}\}$. We assume that the k -th bit channel at φ -th decoding layer among the component polar codes with code length N for the j -th bit level of the equivalent model denotes by $B_{k,\varphi}^{(j)}$. Bhattacharyya parameter denotes by $Z(B_{k,\varphi}^{(j)}) = Z_{k,\varphi}^{(j)}$. The recursive calculation relational expressions are given as following:

$$\begin{cases} Z_{k-1,\varphi}^{(j)} \leq 2Z_{\lfloor k/2 \rfloor, \varphi-1}^{(j)} - (Z_{\lfloor k/2 \rfloor, \varphi-1}^{(j)})^2, & b_k^{(j)} = 0 \\ Z_{k,\varphi}^{(j)} = (Z_{\lfloor k/2 \rfloor, \varphi-1}^{(j)})^2, & b_k^{(j)} = 1 \end{cases} \tag{18}$$

for $j = \{1, \dots, m\}k = \{1, \dots, N\}$, and $0 \leq \varphi \leq n$, $n = \log_2 N$. Especially, $b_k^{(j)} = b_j(x_k)$ denotes the binary value of the j -th label position in the k -th received symbol, and $Z_0 = Z(B^{(j)})$.

Through polarization operation, the cumulative symmetric capacities and reliability are remained in following relational formulas [4]:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N I(B_k^{(j)}) = I(B^{(j)}) \tag{19}$$

$$\frac{1}{N} \sum_{k=1}^N Z(B_k^{(j)}) \leq Z(B^{(j)}) \tag{20}$$

The fraction of the left side of formula (19) tends to $I(B^{(j)})$ with probability 1, when the component code length $N \rightarrow \infty$. On the contrary, when the length of the component polar codes is finite, the following relation holds

$$\frac{1}{N} \sum_{k=1}^N I(B_k^{(j)}) \leq I(B^{(j)}) \tag{21}$$

Especially, Eq. (19) holds for the independence among different bit levels of the equivalent channel model in Fig. 2.

In BIPCM, let $R^{(j)}$ denote the code rate of ordered bit channel $B^{(j)}$ in bit vector channels $\mathbf{B} = \{B^{(1)}, \dots, B^{(m)}\}$. We assume that channel W in Fig. 1 depicts the underlying 2^m -ary discrete memoryless channel. Then source bits of BIPCM scheme are encoded by component binary polar encoders with code rate $R^{(j)}$ for the j -th bit level. The total code rate is defined by

$$R_{BIPCM} = \sum_{j=1}^m R^{(j)} \tag{22}$$

$$R^{(j)} = \frac{|\mathcal{A}^{(j)}|}{N} \tag{23}$$

$$R^{(j)} \leq I(B^{(j)}) \tag{24}$$

where $j \in \{1, 2, \dots, m\}$ denotes the index of bit levels. In this process, each $R^{(j)}$ is employed with respect to the capacity $I(B^{(j)})$ of the j -th duplicated bit channel $B^{(j)}$ of the underlying channel B .

The set $\mathcal{A}^{(j)}$ of indices of information bit channels is chosen based on the following rule with constraint condition (22):

$$\max \sum_{j=1}^m \sum_{k \in A^{(j)}} I(B_k^{(j)}) \tag{25}$$

Instead of exactly analyzing the term $I(B_k^{(j)})$, we approximate every bit capacity among all of the bit channels $\sum_{j=1}^m \sum_{k \in A \cup A^c} I(B_k^{(j)})$ by channel polarization with recursively computing the Bhattacharyya parameters to estimate the capacities or bit error rate (BER) of all the bit channels. Then we pick out the indices of K best bit channels satisfying the constraint conditions on different design requirements (19), (20), (24) to constitute the information bit index set $A^{(j)}$.

According to Arikan’s Proposition 2 [4], we denote the probabilities of block error for DMC channel W employing polar codes $(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ by $P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ with the following definitions:

$$P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c}) = P_r\{\hat{u}_i \neq u_i | \hat{\mathbf{u}}_1^{i-1} = \mathbf{u}_1^{i-1}, i \in \mathcal{A}, u_i \in F_2\} \tag{26}$$

$$\mathcal{A} = \sum_{j=1}^m A^{(j)} \tag{27}$$

where the block error probabilities satisfy the following constraint:

$$P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c}) \leq \sum_{j=1}^m \sum_{k \in A^{(j)}} Z_k^{(j)} \tag{28}$$

In order to obtain Bhattacharyya parameter values, firstly, we calculate the capacity $I(B^{(j)})$ of each channel employing formula (15), then define the corresponding Bhattacharyya parameter using $Z_0 = 1 - I(B^{(j)})$. When channel $B^{(j)}$ belongs to binary erasure channel (BEC), Bhattacharyya parameter of this channel is equal to its erasure probability. Then the reliability parameter of polar codes $(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ could be chosen by the following rule:

$$\min \sum_{j=1}^m \sum_{k \in A^{(j)}} Z_k^{(j)} \tag{29}$$

In this paper, under the joint rate constraint in (22), we generate code construction corresponding to m independent parallel channels in Fig. 2. In other words, the best strategy for assigning individual rates for m independent parallel component polar codes is equal to pick out the information bit sets with the minimum summary results of formula (29).

3.4 General Polar Codes Constructing Algorithm

The above part mainly introduces the Bhattacharyya parameters as the upper bounds of block error rate (BLER) of component polar codes to design the BIPCM schemes over the equivalent channels. In fact, there exist several other measures to estimate the quality of polarized channels e.g. Monte Carlo method, Gaussian approximation and so on. To simplify the description, we describe the design progress of polar codes construction with all those measures in the following general polar codes constructing algorithm.

Algorithm 1: Polar codes construction for BIPCM

Input: Reliability parameters $\xi(B_k^{(j)})$ of polarized channels $B_k^{(j)}$ for $k = 1, 2, \dots, N; j = 1, \dots, m$ and the total rate $R < m$

Output: Information bit sets of constituent polar codes over different levels $\mathcal{A}^{(j)}$ and $\sum_{j=1}^m |\mathcal{A}^{(j)}| = \lceil NR \rceil$

Initialize $\mathcal{A}^{(j)}$ with zero for any $j = 1, \dots, m$

If $\sum_{j=1}^m |\mathcal{A}^{(j)}| < \lceil NR \rceil$ implement

$index(j, k) \leftarrow \arg \max_{j, k': k' \in \mathcal{A}^{(j)}} \xi(B_k^{(j)});$

$\mathcal{A}^{(j)} \leftarrow \mathcal{A}^{(j)} \cup \{k\};$

End if

Return $\mathcal{A}^{(j)}$

In this algorithm, the reliability parameters $\xi(B_k^{(j)})$ can be Bhattacharyya parameters, BER, and BLER of polarized channels. The BER and BLER can be estimated by Monte Carlo method or Gaussian approximation.

Complexity of BIPCM. Assuming that all of component polar codes have the same block length N . The equivalent independent channel model has m levels (corresponding to 2^m -order modulation). The list size of SCL decoding algorithm denotes by L . The complexities of encoding and decoding of BIPCM can be defined by the following formulas.

The time complexities of encoding and decoding of BIPCM

$$\nabla_{T,E} = O(mN \log N) \tag{30}$$

$$\nabla_{T,D}^{SC} = O(mN \log N) \tag{31}$$

$$\nabla_{T,D}^{SCL} = O(mLN \log N) \tag{32}$$

where $\nabla_{T,E}$ denotes the time complexity of encoding of BIPCM; $\nabla_{T,D}^{SC}$ denotes the time complexity of SC decoder; $\nabla_{T,D}^{SCL}$ denotes the time complexity of SCL decoder.

The space complexities of encoding and decoding of BIPCM denote by the following relations:

$$\Delta_{S,E} = O(N \log N) \tag{33}$$

$$\Delta_{S,D}^{SC} = O(N \log N) \tag{34}$$

$$\Delta_{S,D}^{SCL} = O(LN \log N) \tag{35}$$

where $\nabla_{S,E}$ denotes the space complexity of encoding of BIPCM; $\nabla_{S,D}^{SC}$ denotes the space complexity of SC decoder; $\nabla_{S,D}^{SCL}$ denotes the space complexity of SCL decoder.

We assume that our operation system can implement the infinite precious arithmetic in unit cost.

3.5 Decoding Processes of Different Decoding Algorithms

Considering the construction of polar codes, we can regard all component polar codes for m equivalent channels as a sequence of mN independent uses of the B-DMC.

Since polar codes belong to a class of recursive concatenated codes, we can calculate those polarized probabilities of bit channels recursively. Decoding process also can be implemented by the similar recursive features in reverse direction.

Successive Cancellation Decoding: In this section we research the process of decoding in Fig. 1. We denote the layer orders of the total coded bit sequence by λ , $1 \leq \lambda \leq \theta$, $\theta = \log_2(mN)$. There exist $\psi = 2^\lambda$ bit channels in the λ -th layer. The τ -th bit channel denotes by $W_\lambda^{(\tau)}$, $1 \leq \tau \leq \psi$ in the λ -th layer. Then the transmission probability can be defined by following recursive forms [22]:

$$W_\lambda^{(\tau)}\{\mathbf{y}_1^\psi | \mathbf{u}_1^{\tau-1}\} = \sum_{u_{\tau+1}=0}^1 W_{\lambda-1}^{(\lceil \tau/2 \rceil)}\left\{\mathbf{u}_{1,e}^{\tau-1} \oplus \mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_1^{\psi/2}\right\} W_{\lambda-1}^{(\lfloor \tau/2 \rfloor)}\left\{\mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_{\psi/2+1}^\psi\right\} \quad (36)$$

$$W_\lambda^{(\tau+1)}\{\mathbf{y}_1^\psi | \mathbf{u}_1^\tau\} = W_{\lambda-1}^{(\lceil \tau/2 \rceil)}\left\{\mathbf{u}_{1,e}^{\tau-1} \oplus \mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_1^{\psi/2}\right\} W_{\lambda-1}^{(\lfloor \tau/2 \rfloor)}\left\{\mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_{\psi/2+1}^\psi\right\} \quad (37)$$

For the τ -th bit channel, the SC decoder outputs the decisions employing the following decision rule:

$$\hat{u}_\tau = \begin{cases} 0, & i \in \mathcal{A}^c \\ \arg \max_{u_\tau \in \mathcal{F}_2} W_\psi^{(\tau)}\{\mathbf{y}_1^{mN}, \mathbf{u}_1^{\tau-1} | u_\tau\}, & i \in \mathcal{A} \end{cases} \quad (38)$$

To improve the reliability and decrease the implementation complexity of the proposed schemes, we achieve the soft decision messages LLRs by combing the decoding messages in each layer employing following recursive calculations:

$$LLR_\lambda^{(\tau)}(\mathbf{u}_1^{\tau-1} | \mathbf{y}_1^\psi) = 2 \operatorname{arc} \tanh \left(\tanh \left(\left(LLR_{\lambda-1}^{(\lceil \tau/2 \rceil)}(\mathbf{u}_{1,e}^{\tau-1} \oplus \mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_1^{\psi/2}) \right) / 2 \right) \right) \times \tanh \left(\left(LLR_{\lambda-1}^{(\lfloor \tau/2 \rfloor)}(\mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_{\psi/2+1}^\psi) \right) / 2 \right) \quad (39)$$

$$LLR_\lambda^{(\tau+1)}(\mathbf{u}_1^\tau | \mathbf{y}_1^\psi) = (-1)^{u_\tau} LLR_{\lambda-1}^{(\lceil \tau/2 \rceil)}\left(\mathbf{u}_{1,e}^{\tau-1} \oplus \mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_1^{\psi/2}\right) + LLR_{\lambda-1}^{(\lfloor \tau/2 \rfloor)}\left(\mathbf{u}_{1,o}^{\tau-1} | \mathbf{y}_{\psi/2+1}^\psi\right) \quad (40)$$

where $LLR_\lambda^{(\tau)}(\mathbf{u}_1^{\tau-1} | \mathbf{y}_1^\psi) = \ln \frac{W_\lambda^{(\tau)}\{u_1^{\tau-1}, u_\tau=0 | \mathbf{y}_1^\psi\}}{W_\lambda^{(\tau)}\{u_1^{\tau-1}, u_\tau=1 | \mathbf{y}_1^\psi\}}$. The decision rule for index $\tau \in A$ over LLRs of the θ -th layer reads

$$\hat{u}_\tau = \begin{cases} 0 & LLR_\theta^\tau(\mathbf{u}_1^{\tau-1} | \mathbf{y}_1^{mN}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (41)$$

when $\lambda = 0$, the fractions of the right hand sides of two recursive Eqs. (39), (40) are initialized with the initial LLRs $\mathbf{I}_1^{mN} = \{llr_1(v_1), \dots, llr_i(v_i), \dots, llr_{mN}(v_{mN})\}$ of the equivalent

labeling position sequence \mathbf{y}_1^{mN} for the received symbol sequence \mathbf{Y}_1^N . While reaching the last layer i.e., $\lambda=\theta$, decoders start to estimate elements of input information vector \mathbf{u}_1^K . This successive cancellation decoding process has time complexity $O(mN \log N)$ referring to (31). If the values of LLRs are quantized by Q_{SC} bits, space complexity (34) of SC decoding can be derived as [23]:

$$\Delta_{SC} = (2mN + 1) \cdot Q_{SC} + mN - 1 \tag{42}$$

However, when the length of polar codes is finite, Arikan’s SC decoding is sub-optimal compared with ML decoding of polar codes. Then we turn to the asymptotic ML decoding performance algorithm, i.e., SCL decoding algorithm.

SCL decoding algorithm: Instead of making hard decision on the input of channel $W'_{\lambda-1}^{(\lfloor \tau/2 \rfloor)}$, we consider $W'_\lambda^{(\tau)}$ and $W'_\lambda^{(\tau+1)}$ in a decoding tree simultaneously. However, with the increase of decoding level, the number of decoding paths is increasing exponentially. SCL decoding algorithm offers a lower complexity decoding scheme which just considers the first L largest decoding paths and prunes the remaining paths. The paths are composed of likelihood ratios $W'_\lambda^{(\tau)}(\mathbf{y}_1^\psi, \mathbf{u}_1^{\tau-1} | u_\tau = 0) / W'_\lambda^{(\tau)}(\mathbf{y}_1^\psi, \mathbf{u}_1^{\tau-1} | u_\tau = 1)$. Here, L denotes the number of decision paths, i.e., the list size of SCL decoding algorithm. A rule of thumb is that the larger the decision path number L is, the better the reliability of polar codes with SCL decoding. However, SCL decoding algorithm has $O(LmN \log N)$ time complexity. The increase of list size will result to the increment of time complexity linearly. With the increase of index τ , the likelihood ratio of the index will tend to infinite small, which can result in numerical underflow.

- I. In each SC decoding layer, we uniform N probabilities illustrating this decoding algorithm, since there are N probabilities for each layer of $\log N$ layers. The complexity of normalized unit is $O(N \log N)$.

In order to avoid numerical underflow, we use the relationships in (39) and (40) to calculate the posterior probabilities for bit channels [24], build and span the decision decoding tree.

- II. Another way to avoid underflow: During the update process of the term $W'_\lambda^{(\tau)}$, we regard the maximum term $\max\{W'_\lambda^{(\tau)}(\mathbf{y}_1^Z | \mathbf{u}_1^{\tau+1}) | r \in \mathcal{A}\}$ as a normalized factor to scale the residual likelihood values in the λ -th layer. Additionally, in case of reducing the complexity of normalized steps, we can perform the computations in logarithm region.
- III. In [24], authors proposed a method (LLR-based SCL decoding algorithm) of calculating path metrics corresponding to (39) and (40): making use of the posterior soft decisions on $LLR_\lambda^{(g)}$, $g \in \{1, \dots, \tau\}$

$$PM_i^\tau = \sum_{g=1}^{\tau} \ln \left(1 + e^{-(1-2\hat{u}_g[I]) \cdot LLR_\psi^{(g)}[I]} \right) \tag{43}$$

$$LLR_\psi^{(g)}[I] = \log_e \frac{W_\psi^{(g)}\{u_1^{g-1}, u_g = 0 | \mathbf{y}_1^Z\}}{W_\psi^{(g)}\{u_1^{g-1}, u_g = 1 | \mathbf{y}_1^Z\}} \tag{44}$$

where PM_l^τ denotes the τ -th path metric at the l -th decision path among the L largest paths. Notation $LLR_\psi^{(g)}[l]$ denotes the g -th LLRs of the information bits in the ψ -th layer at the l -th decision path. Term $\hat{u}_g[l]$ denotes the estimate of the g -th bit at the l -th decision path.

The path metrics $PM_l^{(\tau)}$ are updated successively by decoders by:

$$PM_l^{(\tau)} = \varphi\left(PM_l^{(\tau-1)}, LLR_\psi^{(\tau)}[l], \hat{u}_\tau[l]\right) \tag{45}$$

where we have the setting for function φ with

$$\begin{aligned} \varphi : R^2 \times \{0, 1\} &\rightarrow R \\ \varphi(a, b, c) &= a + \ln\left(1 + e^{-(1-2c)b}\right) \end{aligned} \tag{46}$$

where a , b , and c represent $PM_l^{(\tau-1)}$, $LLR_\psi^{(\tau)}[l]$, and $\hat{u}_\tau[l]$ respectively.

The method in [24] is one kind of hard-friendly algorithm can boost the throughput of per unit area. Time complexity is still $O(LmN \log N)$. If we assume that path metrics are quantized by Q_{PM} bits, the memory requirement of this decoding can be calculated as [24]:

$$\Delta_{PM} = ((mN - 1) \cdot L + mN) \cdot Q_{SC} + L \cdot Q_{PM} + (2mN - 1) \cdot L \tag{47}$$

Either SCL decoding algorithm or LLR-based SCL decoding algorithm, can improve the error-correcting performance of BIPCM schemes employing aided cyclic redundancy check (CRC), which just costs negligible compute complexity and little loss of coding rates.

4 Performance Evaluation

In this section, we use MATLAB 2016b simulation platform and win7 64bit operation system to complete the performance evaluation of groups of bit-interleaved polar coded modulation schemes.

Before using Monte Carlo simulation to polarize bit channels and search for good bit channels which have the lowest bit error ratios. We assume that the transmission codewords consist of all zero only. During each iteration in all BIPCM schemes with different decoding algorithms, we just apply SC decoding algorithms to estimate the sent messages (all zero). When achieving BER of each bit channel, we sort those results and pick out the indices of the given number most reliability channels to compose the information sets. Unfortunately, the computation complexity of Monte-Carlo code construction is $O(M \cdot N \log N)$, M denotes the iterative number. So, in order to achieve the frozen bit index sets accurately, we need huge training data and cost vast runtime.

In this section, we simulate several bit-interleaved polar coded modulation schemes employing BPSK, 4-ASK (ASK4), 8-ASK (ASK8) and 16-ASK(ASK16) with Gray labeling for constant code length $N=2048$, $R=0.5$ under SC, simple successive cancellation list (SSCL), and SCL decoding algorithms. Hereafter, we use the shorthand SSCL to represent the SCL decoding algorithm with only one single decoding path, since the number of decoding paths of SCL decoding algorithm can range from 1 to the length of component polar codes. Specially, the schemes with ASK8 modulation cannot be labeled by Gray labeling rule, we employ a quasi-Gray labeling, since the labeling cannot minimizing the number of signals [25] i.e., Gray labeling condition cannot be met. In different BIPCM schemes, we want to maximize the code performance (referring to 22), so we fix different

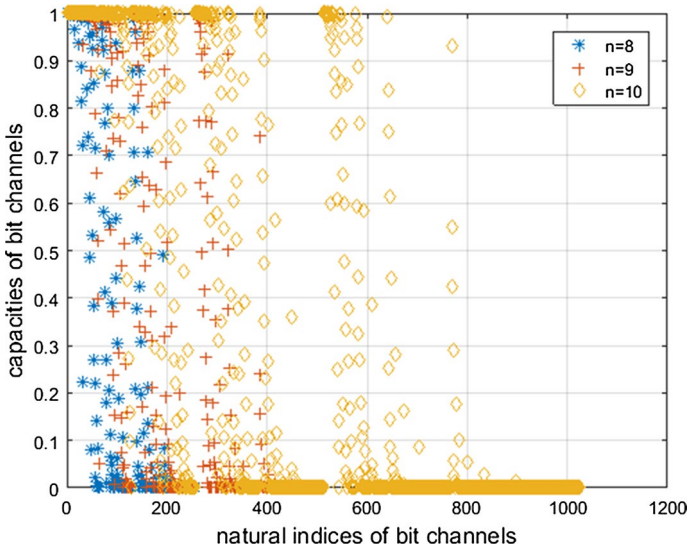


Fig. 3 Capacities of bit channels with natural indices

target signal to noise ratios and the number of iteration as 1×10^5 for Monte-Carlo code construction.

We apply algorithm 1 to seek index sets of the information bit channels: Bhattacharyya parameters, Monte Carlo method and Gaussian approximation are used to estimate the performance of polarized channels and construct the index sets of information and frozen bit channels with the construction (29).

Figure 3 describes the distribution of capacities $I(W_{mN}^{(i)})$ of all polarized bit channels with code length 256, 512, and 1024 of polar codes respectively. In other words, it depicts the polarization effect of the binary erasure channels with erasure probability $\epsilon = 0.32$ according to natural channel indices ranging from 256 to 1024. With the increase of polar codes length, the proportion of channels whose capacities fall into the middle region between 0 and 1, is becoming fewer and fewer. On the contrary, the proportion of channels with capacities 0 or 1 is more than that falling into middle region between 0 and 1. This polarization phenomenon belongs to the well-known channel polarization [4].

Figure 4 describes the capacities sorted increasingly for bit channels with normalized indices. With the increase of polar codes length, the slopes of curves become steeper and steeper. In other words, the rate of polarization approaches optimality with the increment of polar codes length. This is also corresponding to Fig. 3. Both of Figs. 3 and 4 show that increasing the length of polar codes can improve the reliability of polar codes with the cost of complexities of encoding and decoding.

Figure 5 depicts the capacity comparison of each bit level of 16-ASK modulation. The solid curves denote the capacities of different levels in polar codes and 16-ASK modulation. The dot and dash curves denote the capacities of BIPCM schemes. The rates of component polar codes for different levels are randomly distributed in this figure. Apparently, Fig. 5 shows that the assignment is not optimal to design component codes. One should select the rates of different levels according to (24).

We assume that the transmission information messages consist of all zero and employ algorithm 1 with BER deriving from Monte Carlo method for bit channels of polar

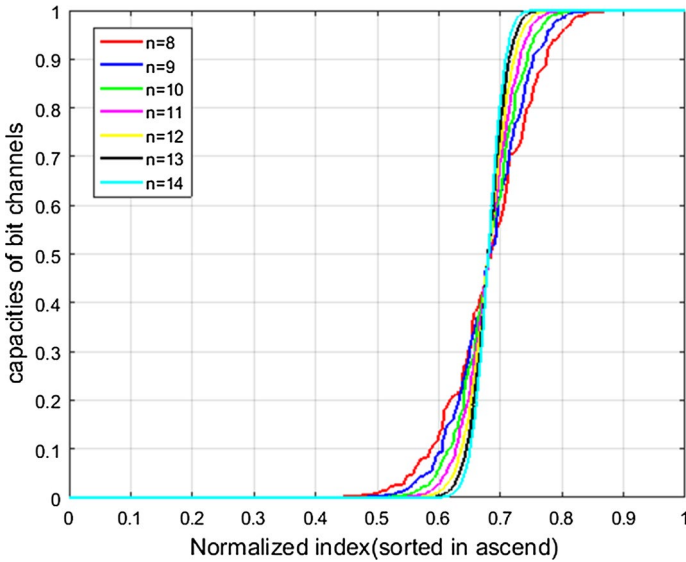


Fig. 4 Distributions of capacities over bit channels with the increase of polar codes length

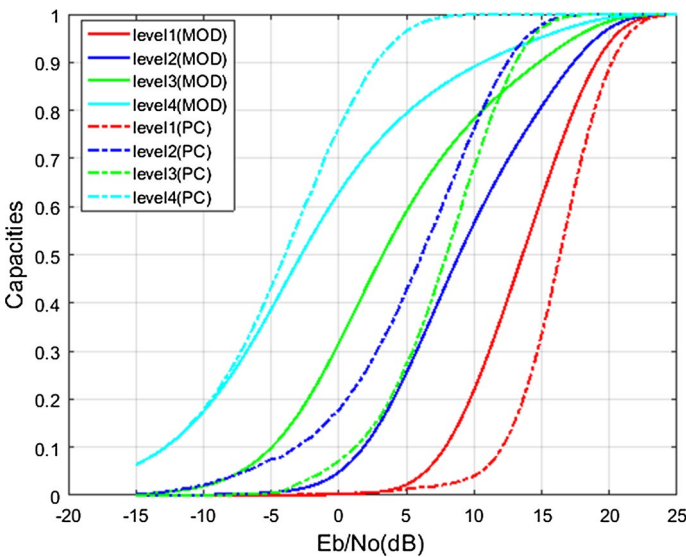


Fig. 5 Capacities of different mapping levels for 16-ASK modulation and 4 levels of polarized channels

codes $(2048, 1024, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$. We obtain the information bit sets \mathcal{A} and frozen bit sets \mathcal{A}^c . Unfortunately, the coding construction with Monte Carlo method has the time complexity $O(\delta \cdot N \log N)$ which increases linearly with the iterative number δ . So we need to consider the trade-off between reliability and complexity of polar codes synthetically.

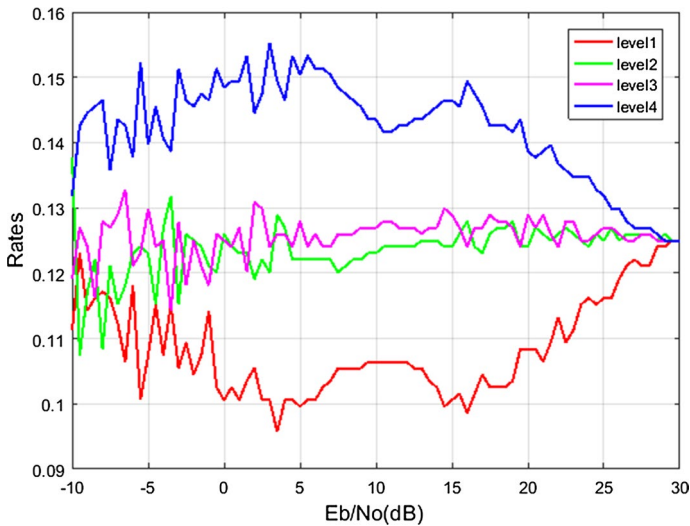


Fig. 6 Code rates of component polar codes for different levels of 16-ASK modulation over the equivalent channel model under Monte Carlo coding construction

Figure 6 describes the rate distributions of component polar codes for different levels of 16-ASK modulation in BIPCM. We set the length of component polar codes to be equal to 1024 and overall code rate to be equal to 0.5. The rates are derived from the ratios of the size of information bit sets for component polar codes to the total length of polar codes referring to (22) and (23) employing Monte Carlo code construction. These curves come from the natural set partitioning rule. The right-most column presents the least significance column corresponding to $\mathbf{M}_{SP,3}$ in (1). The code rate for level 1 is lowest, because its Euclidean distance of the corresponding subsets of constellation partition is comparative minimum which results in the worst error correction performance. On the contrary, level 4 has the best reliability for the protection of large subset distance. At the highest SNR region, the code rates tend to identical values, which show that code rates would be uniformly distributed on each bit level with increasing signal transition power.

Another method of approaching BLER of BIPCM is Gaussian approximation [26] which applies the LLRs mean of each polarized channel and approximates the BLER of the corresponding channels [27]. And then, algorithm 1 employs those BLER to construct the information set \mathcal{A} through selecting the K best sub-channels among N channels.

Figure 7 depicts the rate distributions of constituent polar codes for 4 levels of 16-ASK modulation over the equivalent channel model. The polar codes have the overall code length 2048 and code rate 0.5. Figure 7 shows us the rate distributions for each component code over 4 levels of modulation referring to (22) and (23). From Fig. 7, we observe that level 1 has the lowest code rate and level 4 owns the highest code rate over the whole SNR region. The curve tracks of level 1 and level 4 are related with the partition procedure of the 16-ASK modulation constellation which is similar to Fig. 6. In addition, we can find that at the lowest and highest SNR regions, the distributions of code rates for each bit level become most separated. Then, we always apply Monte Carlo code construction to simulate the random process of channel polarization considering the stable property of Monte Carlo theory with large training set.

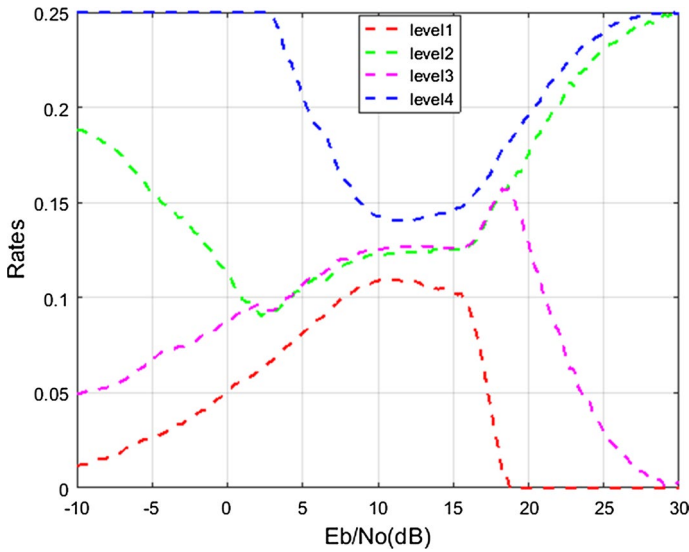


Fig. 7 Rate distributions of information bit indices for component polar codes on 4 levels for 16-ASK modulation over the equivalent channel model

We design the component polar codes for the equivalent channel model applying algorithm 1 which employs BER deriving from Monte Carlo simulation with iteration 1×10^5 over BPSK, 4-ASK (ASK4), 8-ASK (ASK8) and 16-ASK (ASK16) with Gray labeling, and polar codes with code length $N=2048$, $R=0.5$ under SC, SSCL, and SCL decoding algorithms respectively. Among different BIPCM schemes, to maximize the overall summary capacities referring to (25), we fix several different target SNR for different modulation schemes [13].

Figure 8 depicts that the error-correcting performance of polar codes, LDPC, and Turbo codes over BICM with BPSK modulation. All kinds of codes have coding rate 0.5. Especially, Turbo codes employ generator polynomials $g(d) = [1, (1 + d + d^2)/(1 + d)]$. The length of polar codes is 2048 and that of LDPC is 2304, which belongs the fixed construction of WiMAX standard. The error-correcting performance of SCL with $L=16$ outperforms other schemes and has 0.25 dB gain comparing with LDPC at BLER level 10^{-3} , 1.5 dB for Turbo codes. The convergent feature of LDPC with min-sum decoding algorithms is better than polar codes with SCL decoding algorithm at SNR region being more than 2.2 dB. The reliability of scheme with SCL decoding algorithm ($L=1$) is similar to that of SC decoding algorithm. Turbo codes with max and linear log maximum posterior probability have the worst convergent property. However, at high SNR region, the convergence of SC decoding algorithm also becomes worse and worse.

Figure 9 depicts the performance of BIPCM schemes with BPSK modulation under SC, SSCL and SCL decoding algorithms applying Gray labeling. At the same time, we fix code length as 2048 and coding rate as 0.5. At 1×10^5 BER levels, the scheme with SCL $L=4$ decoding algorithm outperforms SSCL scheme about 0.7 dB, and has about 0.8 dB performance gain than the scheme with SC decoding algorithm. The similar phenomenon emerges at 1×10^{-3} BLER levels. Meanwhile, both groups of BER and BLER curves almost have identical trend, since the calculations of BER have a lot relations with BLER respectively. Simultaneously, the performance of BER and BLER with SSCL and

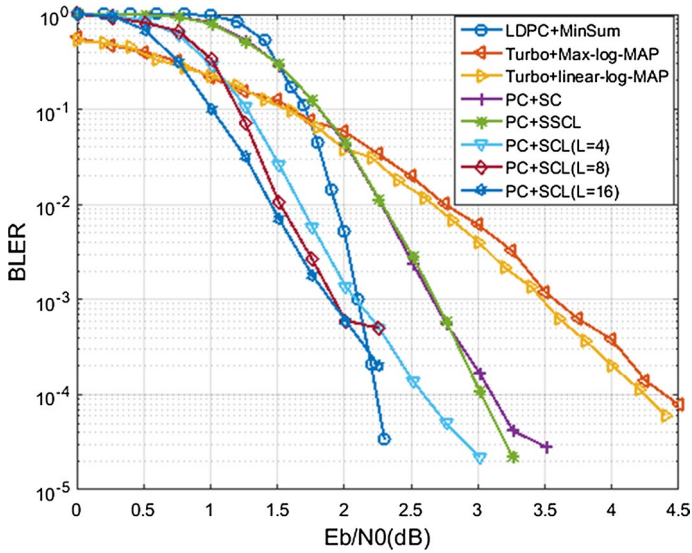


Fig. 8 Error-correcting performance of polar codes and LDPC over BICM schemes with BPSK modulation

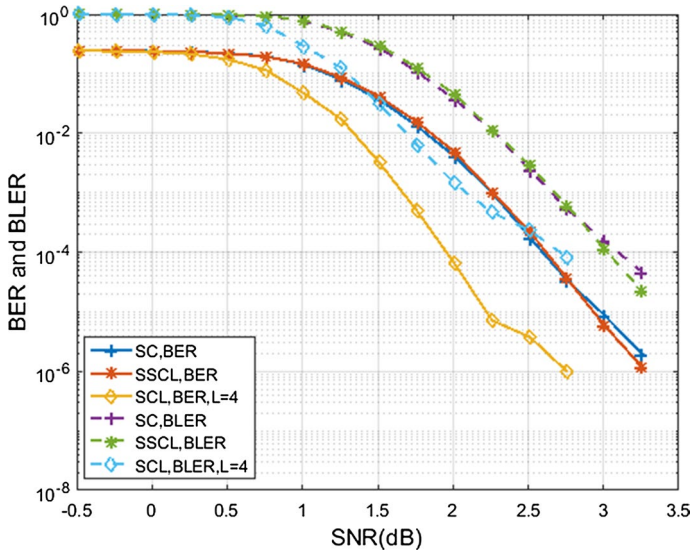


Fig. 9 The BER and BLER performance of polar codes and BPSK modulation under SC, SSCL, and SCL decoding algorithms

SC decoding schemes has the similar tendency. But at higher SNR region, the convergent property of SSCL scheme is better than that with SC decoding.

Figure 10 depicts that polar coded ASK4 BICM scheme under three different decoding algorithms. The error-correcting performance of scheme with SCL decoding algorithm outperforms the others. By the comparison between Figs. 9 and 10, we find that the

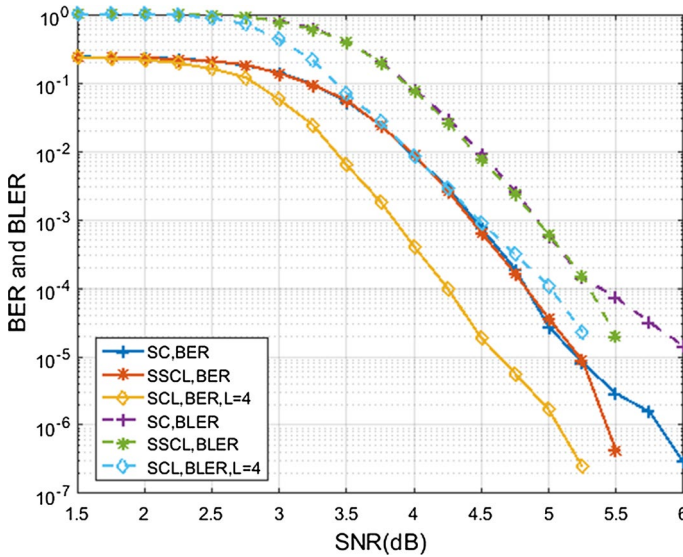


Fig. 10 The BER and BLER comparisons among different decoding algorithms with code length $N=2048$, code rate $R=0.5$ in polar coded and ASK4 BICM scheme

convergent property of SCL become better and better at the end of curves with the equivalent partition of encoded bit sequence in Fig. 2. It can be seen that the scheme with SCL decoding $L=4$ algorithm at $BER 1 \times 10^{-5}$ has 0.9 dB gain comparing with the other two BICM schemes employing SC and SSCL decoding algorithms. The BLER of the scheme with SCL decoding reaches 1×10^{-3} at SNR 4.5 dB with 0.5 dB gain more than the other two competitive schemes. The curves of the BICM schemes with SC and SSCL decoding algorithms are almost coincide. However, in the end of SC decoding algorithm, the speed of decrease becomes slower than SSCL decoding algorithm, i.e., the convergent character of SC algorithm is worse than SSCL algorithm.

Figure 11 shows that the BIPCM scheme with SCL improves about 0.9 dB and 0.7 dB gain comparing with the schemes with SSCL and SC decoding algorithms respectively at $BER 1 \times 10^{-5}$. It induces that in polar coded modulation schemes, the error correcting performance of SSCL decoding algorithm does not keep better than SC decoding algorithm with the increase of modulated order at all time.

Figure 12 depicts that with the increase of modulating order, the reliability improvements of the equivalent channel model with SCL decoding algorithm are still remarkable, achieve 0.5 dB and 0.6 dB gain outperforming the competitive schemes with SSCL decoding and SC decoding, at $BER 10^{-3}$ and $BLER 10^{-5}$ individually. Nevertheless, there exist different fluctuations in some degree, in SCL decoding schemes and SC decoding schemes at high SNR region. Conversely, the profiles of BER and BLER curves of SSCL decoding schemes are more stable than that of other decoding algorithms.

By the observations on above figures, we find that the BER and BLER performance of the schemes with SCL decoding algorithm have the best convergent property. At the same time, the performance of SSCL decoding schemes has the comparative stable graphic tracks over the schemes with SC decoding algorithm.

Figure 13 depicts the BER and BLER performance comparisons of BIPCM schemes with multi-order modulation under SCL ($L=4$) decoding algorithm. Since we employ

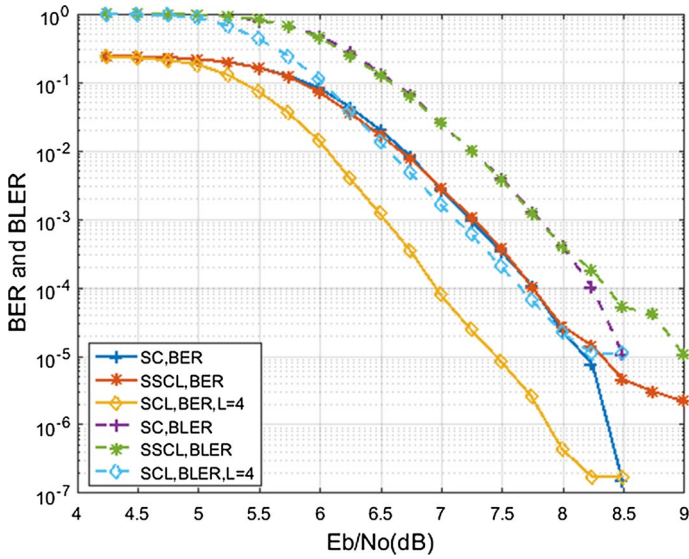


Fig. 11 The BER and BLER comparisons among different decoding algorithms at code length $N=2048$, code rate $R=0.5$ under polar coded and ASK8 BICM scheme

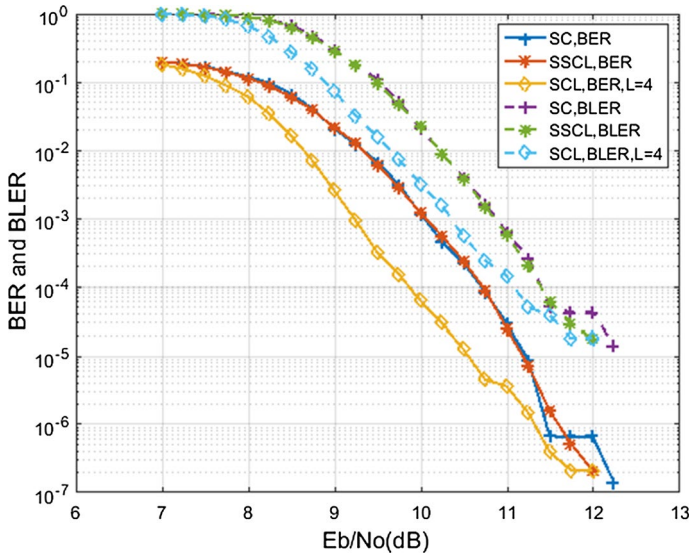


Fig. 12 The BER and BLER compare between different decoding algorithms under ASK16 over BICM scheme

Monte Carlo method to construct the information sets which are used to translate information bits, we set distinct beginning points of simulation for different modulation schemes to maximize the coding performance with respect to (25). We can see that under the same decoding algorithms, the BIPCM scheme with BPSK modulation expresses the fastest decreasing speeds at both of BER and BLER regions. The

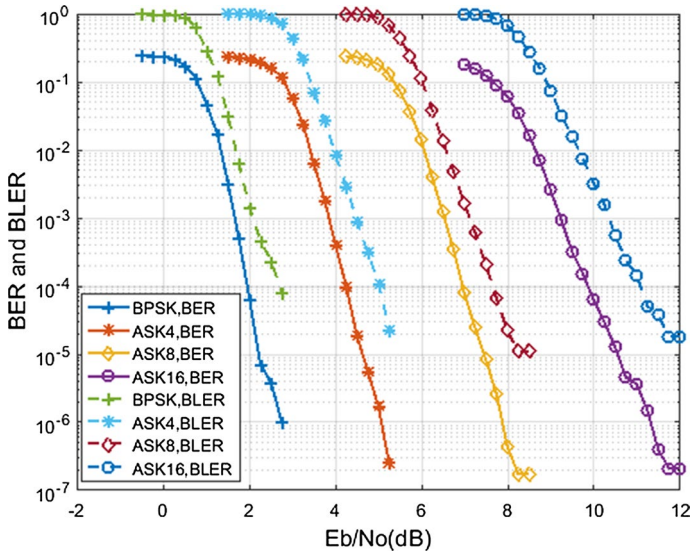


Fig. 13 The BER and BLER performance of polar coded multi-order modulation under SCL $L=4$ decoding algorithm

convergent gaps among BPSK, ASK4, ASK8, ASK16 schemes widen with the increase of spectrum efficiency, i.e., the increase of modulation order. In the other word, in order to improve the spectrum efficiency under the same BER and BLER performance, we have to cost more power to process signals in a finite range.

5 Conclusion

Through analyzing polar codes construction and the design of bit-interleaved coded modulation, we find that the construction of BIPCM is equal to an equivalent channel model which is composed of several independent parallel transmission channels through synthetically considering coding and modulation in practical applications. We propose a general code constructing algorithm to design the constituent polar codes for those equivalent parallel channels. And then, we employ a general bijective mapping to accomplish the bit address mapping from coded bits to signal points in constellation with Gray labeling or set partition labeling rules under different polar decoding algorithms. At the receiver, de-modulators calculate the decoding metrics composed of LLRs of the received symbol sequences from AWGN channels. Decoders apply SC, SSCL and SCL decoding algorithms to estimate the translated information bits and recover the original messages. Simulation results show that the BIPCM schemes applying SCL decoding algorithm have the best error-correcting performance among all the decoding algorithms. The schemes with SSCL and SC decoding algorithms have the similar performance. However, at high SNR region, the schemes with SC decoding algorithm become more unstable than SSCL schemes. Further work will focus on non-binary input discrete memoryless channels with component non-binary polar codes in BIPCM.

Acknowledgements Manuscript submitted December 14, 2017. This work was supported by the National Natural Science Foundation of China under Grant 61671073.

References

1. Ungerboeck, G. (1982). Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28(1), 55–67.
2. Zehavi, E. (1989). 8-PSK trellis codes on Rayleigh channel. In *Proceedings of IEEE military communications conference* (pp. 536–540).
3. Caire, G., Taricco, G., & Biglieri, E. (1998). Bit-interleaved coded modulation. *IEEE Transactions on Information Theory*, 44(3), 927–946.
4. Arkan, E. (2009). Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7), 3051–3073.
5. Tal, I., & Vardy, A. (2011). List decoding of polar codes. In *Proceedings of IEEE international symposium on information theory* (pp. 1–5).
6. MacKay, D. J. C. (1999). Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45, 399–431.
7. Berrou, C., Glavieux, A., & Thitimajshima, P. (1993). Near Shannon limit error correcting coding and decoding: Turbo-codes. In *Proceedings of IEEE international conference on communications (ICC)*, (Geneva, Switzerland) (Vol. 2, pp. 1064–1070).
8. Leroux, C., Tal, I., Vardy, A., & Gross, W. J. (2011). Hardware architectures for successive cancellation decoding of polar codes. In *Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP)*, (pp. 1665–1668).
9. Arikan, E. (2008). A performance comparison of polar codes and Reed–Muller codes. *IEEE Communications Letters*, 12(6), 447–449.
10. Arikan, E. (2010). Polar codes: A pipelined implementation. In *Proceedings of 4th international symposium broadcasting communications (ISBC)*, (pp. 11–14).
11. Afser, H., & Tirpan, N. (2014). Bit interleaved polar coded modulation. *IEEE wireless communications and networking conference (WCNC)*, (pp. 480–484).
12. Seidl, M., Schenk, A., Stierstorfer, C., & Huber, J. B. (2013). Polar-coded modulation. *IEEE Transactions on Communications*, 61(10), 4108–4119.
13. Shin, D. M., Lim, S. C., & Yang, K. (2012). Mapping selection and code construction for 2^m -ary polar coded modulation. *IEEE Communications Letters*, 16(6), 905–908.
14. GPP. (2016). Study on new radio access technology physical layer aspects. TR 38.802.
15. Chandesaris, L., Savin, V., Declercq, D. (2018). Lasting successive-cancellation based decoders for multi-level polar coded modulation. In *IEEE conferences, 2018 25th international conference on telecommunications (ICT)*, (pp. 264–268).
16. Pottie, G. J., & Taylor, D. P. (1989). Multi-level codes based on partitioning. *IEEE Transactions on Information Theory*, 35(1), 87–98.
17. Alvarado, A., Graell i Amat, A., Brannstrom, F., & Agrell, E. (2012). On the equivalence of TCM encoders. In *Proceedings of IEEE international symposium on information theory (ISIT)*, (pp. 2401–2405).
18. Wachsmann, U., Fischer, R. F. H., & Huber, J. B. (1999). Multilevel codes: Theoretical concepts and practical design rules. *IEEE Transactions on Information Theory*, 45, 1361–1391.
19. Martinez, A., Guillen, A., Fabregas, I., Caire, G., & Willems, F. (2009). Bit-interleaved coded modulation revisited: A mismatched decoding perspective. *IEEE Transactions on Information Theory*, 55, 2756–2765.
20. Martinez, A., & Willems, F. (2006). A coding theorem for bit-interleaved coded modulation. In *27th symposium on information theory in the Benelux (WIC 2006)*, Noordwijk, the Netherlands.
21. Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory* (pp. 22–23). New York: Wiley.
22. Trifonov, P., Miloslavskaya, V., & Morozov, R. (2018). Fast sequential decoding of polar codes. Available: <https://arxiv.org/abs/1703.06592v1>.
23. Leroux, C., Raymond, A. J., Sarkis, G., & Gross, W. J. (2013). A semi-parallel successive-cancellation decoder for polar codes. *IEEE Transactions on Signal Processing*, 61(9), 289–299.
24. Stimming, A. B., Parizi, M. B., & Bury, A. (2015). LLR based successive cancellation list decoding of polar codes. *IEEE Transactions on Signal Processing*, 63(19), 5165–5179.
25. Stierstorfer, C., & Fischer, R. F. H. (2007). (Gray) mappings for bit-interleaved coded modulation. In *Proceedings of IEEE vehicle technology conference* (pp. 1703–1707).
26. Mori, R., & Tanaka, T. (2009). Performance of polar codes with the construction using density evolution. *IEEE Communications Letters*, 13(7), 519–521.
27. Wu, D., Li, Y., & Sun, Y. (2014). Construction and block error rate analysis of polar codes over AWGN channel based on Gaussian approximation. *IEEE Communication Letters*, 18(7), 1099–1102.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Kun Zhao received the B.S. degree in electronic information engineering major from Zhengzhou University, Zhengzhou, Henan province, China, in 2002, and M.S. degree in information and communication engineering of Kunming University of Science and Technology, Kunming, Yunnan province, China, in 2015. Currently, he is pursuing the Ph.D. degree in the school of information and communication engineering of Beijing University of Posts and Telecommunications, Beijing China, from 2015. His research interests are information theory and coding theory.



Yong Liu is Professor, doctoral supervisor, graduated and received a Ph.D. degree in Beijing University of Posts and Telecommunications, Beijing, China, in 2002. He was learning computer network security technology in German University, and studied data communication technology in Japan, and then developed several multimedia communication projects cooperatively in Silicon Valley, USA.



Haiqing Du was born in 1976, Ph. D, Lecturer. She received her Ph.D. degree in communications and information systems major from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2010, and she has been a visiting scholar at the department of electrical engineering from University of Washington, Seattle, WA 98195, USA in 2013. Currently she is a lecturer in Communication Network Technology Center, School of Information and Communication Engineering, BUPT. Her major research interests include wireless video transmission and vision-based localization and object tracking. In recent years, she has contributed to about 7 research projects. She has been awarded 6 national invention patents, and has published more than 10 papers.



Baoku Yuan is lecturer, received B.S. degree in engineering in computer applications major from Beijing University of Posts and Telecommunications, Beijing, China, in 1999.