

# An Ultra-Reliable Low-Latency Non-Binary Polar Coded SCMA Scheme

Shufeng Li, *Member, IEEE*, Mingyu Cai, Libiao Jin, Yao Sun, *Senior Member, IEEE*, Hongda Wu, *Student Member, IEEE*, and Ping Wang, *Senior Member, IEEE*

**Abstract**—The joint transmission scheme of polar codes and sparse code multiple access (SCMA) has been regarded as a promising technology for future wireless communication systems. However, most of the existing polar-coded SCMA (PC-SCMA) systems suffer from high latency caused by the feedback iteration and list decoding. In addition, the error performance of PC-SCMA systems is unsatisfactory for ultra-reliable transmission. Inspired by the compelling benefits of non-binary polar codes, in this paper, we design a non-binary polar-coded SCMA (NB-PC-SCMA) system with a free order matching strategy to address the issues of delay and reliability. Specifically, we first formulate a joint factor graph for NB-PC-SCMA and propose a non-binary successive cancellation list (NB-SCL) and damping based joint iterative detection and decoding (NSD-JIDD) multiuser receiver to improve the BER and latency performance. Then, a lazy-search based NB-SCL (L-NB-SCL) decoding is proposed to reduce the computational complexity by modifying the path search pattern of the list decoder. After that, we optimize the update of user nodes for SCMA detection to improve the convergence error and finally propose the optimized NSD-JIDD (OSD-JIDD) algorithm, which can avoid redundant operations by exploiting L-NB-SCL decoding. Simulation results show that the proposed NB-PC-SCMA system achieves better bit error rate (BER) performance and considerable latency gain when compared to its counterparts. In particular, the proposed OSD-JIDD can achieve similar BER performance of NSD-JIDD with less complexity.

**Index Terms**—Polar Codes, SCMA, Ultra-Reliable Low-Latency Communication, Multiuser Receiver, Joint Iterative Detection and Decoding.

## I. INTRODUCTION

WITH the rapid development of wireless communication networks, improving the spectrum efficiency with limited physical resources has become an urgent call for B5G/6G communication systems. One of the key technologies to solve this problem is the non-orthogonal multiple access (NOMA) scheme [1]. Sparse code multiple access (SCMA), a potent code-domain NOMA scheme, is a new candidate for future communication systems attributed to the superior capacity of

This work was supported in part by the National Natural Science Foundation of China under Grant 61601414, and in part by the Fundamental Research Funds for the Central Universities under Grant CUC210B032.

Shufeng Li, Mingyu Cai and Libiao Jin are with the State Key Laboratory of Media Convergence and Communication, School of Information and Communication Engineering, Communication University of China, Beijing 100024, China (e-mail: lishufeng@cuc.edu.cn; cccmy1c@cuc.edu.cn; libiao@cuc.edu.cn).

Yao Sun is with James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, U.K. (e-mail: Yao.Sun@glasgow.ac.uk)

Hongda Wu and Ping Wang are with the Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, ON M3J 1P3, Canada (e-mail: hwu1226@eeecs.yorku.ca; pingw@yorku.ca)

overload tolerance and resource reuse [2]. At the receiver end, the message passing algorithm (MPA) is adopted to detect individual user symbols.

However, the SCMA system still fails to meet the throughput and reliability for future networks [3]. In [4], the amalgamation with existing techniques, e.g., channel coding, was introduced to improve the performance of SCMA. So far, turbo coded SCMA (TC-SCMA) [5]–[7] and low-density parity-check (LDPC) coded SCMA (LDPC-SCMA) [8]–[11] systems have been widely investigated to obtain a coding gain. While these works conceived impressive joint detection and decoding (JDD) algorithms based on “turbo principle” for co-design systems, the above coding schemes are difficult to meet the ultra-reliable transmission for future communication systems and suffer a high computational complexity.

As a standard code for control channels for 5G New Radio [24], polar code proposed by E. Arikan in [25] is the first coding scheme that can provably achieve the capacity of the binary input discrete memoryless channel with low complexity. In particular, polar codes can provide excellent error correction capability and higher spectral efficiency, which is competitive in the scenario of ultra-reliable low-latency communication (URLLC) [26] for future wireless networks.

Benefiting from the above achievements, polar-coded SCMA (PC-SCMA) systems have been investigated in the literature. A typical JDD scheme directly combines soft-input soft-output (SISO) polar decoder and SCMA detector [12]–[16]. Specifically, the authors in [12] proposed a JDD algorithm amalgamating MPA with belief propagation decoding to obtain performance gains. As a further development, a joint iteration detection and decoding (JIDD) receiver using a soft cancellation (SCAN) decoder was proposed in [13] without inner iteration, which laid a foundation for JDD-based PC-SCMA receivers. In [14]–[16], some modified versions of JIDD were proposed to accelerate the convergence and reduce the computational complexity. However, none of these traditional SISO schemes can break the BER performance limit of the SCAN decoder.

In contrast, an alternative JDD scheme amalgamates the soft-input hard-output (SIHO) polar decoder with the SCMA detector [17]–[23]. To be more specific, a sequential user partition based JDD receiver was proposed in [17], [18] with limited receiver performance due to the feedback of hard outputs. The authors in [19] first proposed a JDD scheme using a SISO-based successive cancellation (SC) decoder to achieve performance gains. Furthermore, a JIDD employing an SC-list (SCL) decoder was presented in [20], [21], which shows a

TABLE I  
OVERVIEW OF EXISTING LITERATURE ON CHANNEL CODED SCMA SYSTEMS WITH JDD SCHEMES.

Contributions	This work	PC-SCMA											TC-SCMA		LDPC-SCMA				
		[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[6]	[7]	[8]	[9]	[10]	[11]
Joint factor graph	✓	✓	✓	✓	✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓
SISO	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Impact of user load	✓						✓						✓	✓	✓	✓	✓	✓	✓
BER improvement	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fading channels	✓		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Complexity reduction	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Latency reduction	✓						✓			✓				✓	✓	✓	✓	✓	✓
Early termination	✓				✓	✓			✓				✓	✓	✓	✓	✓	✓	✓
Damping technique	✓		✓	✓	✓	✓								✓	✓	✓	✓	✓	✓
Convergence analysis	✓		✓	✓	✓					✓	✓			✓	✓	✓	✓	✓	✓
Effect of channel estimation													✓	✓					
Non-binary coding	✓																✓	✓	✓
Non-binary coding with FOMS	✓																		

better BER performance by designing the SCL decoder's extrinsic messages for turbo iteration. A joint channel estimation and decoding scheme was also proposed for fading channels [22], [23].

However, most of the SIHO schemes still lags behind the latency and BER target of URLLC. As a promising solution, non-binary polar codes (NB-PCs) can polarize discrete memoryless channels with arbitrary q-ary alphabets for ultra-reliable transmission [27]. Importantly, NB-PCs can save decoding latency by symbol-level operation instead of bit-level counterparts [28], providing potential for URLLC application. Moreover, NB-PCs with different non-binary kernels over  $GF(q)$  were investigated in [29]–[31], which shows BER gains over binary polar codes.

The existing studies on coded SCMA systems employing JDD are summarized in Table I, which facilitates a comparison of this paper's contributions with other state-of-the-art research. As can be observed, these studies have mainly focused on BER improvement and complexity reduction, while there is a scarcity of literature on latency reduction, overload impact, early termination (ET), and damping techniques. Note that none of these work jointly designed coded SCMA systems with NB-PCs.

Motivated by the impressive BER and latency performance of NB-PCs, we design a superior non-binary PC-SCMA (NB-PC-SCMA) architecture. However, there are still challenges to achieve this target. First, the existing non-binary coded SCMA system adopts a constrained order match strategy (COMS) [10], [11] to facilitate the joint design, i.e., solely considering a finite field order identical to the SCMA modulation order. COMS has an inherent weakness since it cannot trade-off system throughput, BER performance and complexity to fit the scenario. Second, if unconstrained order matching is considered, the receiver will yield the likelihood information conversion. Thus, the inner soft message exchange rules and the transmission reliability need to be resolved. Finally, although the BER performance of the NB-PC-SCMA system can be improved with increasing finite field order, the receiver suffers an increased decoding complexity. As a result, the receiver also requires a complexity reduction technique.

In this paper, we are engaged in tackling the above challenges to investigate the NB-PC-SCMA system. The main

contributions of this paper are outlined as follows.

- 1) To the best of our knowledge, it is the first time to propose an NB-PC-SCMA scheme with free order matching strategy (FOMS). Specifically, we introduce symbol-to-bit conversion at the transmitter while the receiver applies a novel information exchange rule among the bits, field symbols, and SCMA codewords. With the aid of FOMS, the proposed system can freely select the field and modulation order configurations according to the requirements without the limitation of the NB-PC alphabet size.
- 2) Furthermore, we propose a non-binary SCL (NB-SCL) and damping based JIDD (NSD-JIDD) algorithm for NB-PC-SCMA systems by combining the factor graphs of the NB-PC decoder and the SCMA detector into a joint factor graph (JFG). According to the connection in the JFG, the extrinsic soft information is exchanged and is compressed by damping techniques to improve error propagation. In addition, a cyclic redundancy check (CRC)-based ET is applied to eliminate redundant iterations, which facilitates clock cycle savings for the receiver.
- 3) To reduce the complexity of the iterative receiver and improve the convergence performance, we optimize the receiver constituents, i.e., the SCMA detector and the polar decoder. A lazy-search based NB-SCL (L-NB-SCL) decoding is proposed to avoid redundant path splitting in the decoding process. In addition, update operations that are less dominant for user nodes are removed in the SCMA detection process to make full use of the a priori information. Accordingly, the resultant optimized NSD-JIDD (OSD-JIDD) algorithm can significantly decrease the computational complexity in the receiver.

The rest of this paper is organized as follows. Section II presents the FOMS and the proposed NB-PC-SCMA system model. Following this, a multiuser iterative receiver is designed in Section III, while its optimization scheme is detailed in Section IV. Section V discusses the performance of our proposed system over BER, complexity, and latency with numerical simulation results. Finally, the conclusions are drawn in Section VI.

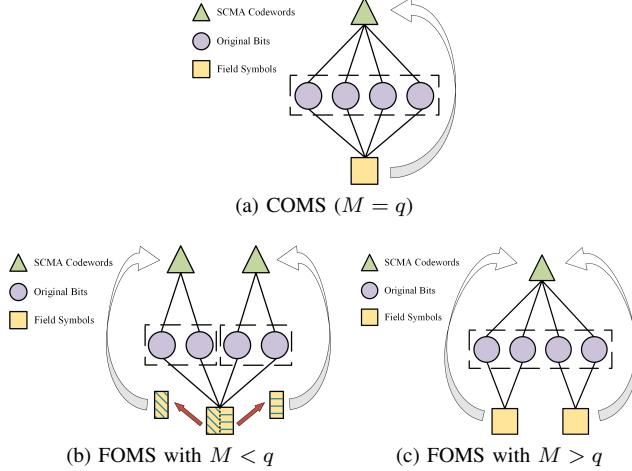


Fig. 1. Different order matching strategies for non-binary coded SCMA.

## II. SYSTEM MODEL

### A. Free Order Matching Strategy

Assume that  $q$ -ary codewords are converted into symbols by  $M$ -point SCMA modulation at the transmitter. Fig. 1 exemplifies the difference between COMS and FOMS under non-binary coded SCMA systems. For COMS of  $q = 16$  shown in Fig. 1(a), every four bits are mapped into an SCMA codeword and Galois field symbol. Thus, SCMA codewords and field symbols are a group of one-to-one mapping. When  $M = 4$  and  $q = 16$ , e.g., the case FOMS with  $M < q$  in Fig. 1(b), a field symbol and an SCMA codeword are associated with 4 bits and 2 bits, respectively. In terms of field symbols, one symbol is mapped into two codewords. FOMS splits a field symbol into two sub-symbols using bit mapping. For  $M = 16$  and  $q = 4$ , e.g., the FOMS with  $M > q$  in Fig. 1(c), which is a typical multiple-to-one mapping, two field symbols are mapped into one codeword.

For COMS, only the system with  $M = q$  is considered, which facilitates the modulation at the transmitter and the detection at the receiver. However, as the SCMA modulation order  $M$  increases with  $q$ , the receiver is more prone to misjudgment and has a worse convergence speed [32]. Moreover, the system may suffer additional overhead due to the design of a large codebook.

Intuitively, COMS can be understood as a special case of FOMS with  $M = q$ . For FOMS, we introduce bit mapping at the transmitter, which makes adopting binary codewords for modulation feasible. Since the order matching is not limited, the universality is the visible advantage of FOMS. For example, FOMS can trade-off the system throughput and reliability given the field order  $q$  according to the scene. The message exchange rule and performance analysis for FOMS-based receiver are depicted in Sections III and V-A, respectively.

### B. System Model for FOMS Based NB-PC-SCMA

The uplink NB-PC-SCMA system with FOMS is shown in Fig. 2. The data of  $J$  users are multiplexed on  $K$  orthogonal resources, giving an overload factor of  $\lambda = J/K$ . To be

more specific, the  $A$ -length information bits sent by user  $j$  ( $1 \leq j \leq J$ ) are denoted as  $\mathbf{u}_j = [u_{j,1}, u_{j,2}, \dots, u_{j,A}]$ , which is encoded as  $\mathbf{b}_j = [b_{j,1}, b_{j,2}, \dots, b_{j,D}]$  by CRC encoder. The bits in  $\mathbf{b}_j$  are converted to the field symbol over  $GF(q)$  as the  $D'$ -length uncoded vector  $\mathbf{b}'_j = [b'_{j,1}, b'_{j,2}, \dots, b'_{j,D'}]$ , where field size  $q = 2^p$  and length  $D' = D/p$ . Then,  $\mathbf{b}'_j$  are placed at  $D'$  information symbol positions of sequence  $\mathbf{a}'_j = [a'_{j,1}, a'_{j,2}, \dots, a'_{j,N'}]$ , which are determined by Monte-Carlo simulation. The remaining positions are filled with the 0-valued frozen symbols. The resultant sequence  $\mathbf{a}'_j$  is then encoded into  $\mathbf{c}'_j$  containing  $N' = 2^\omega$  symbols by the non-binary polar encoder, which can be expressed as

$$\mathbf{c}'_j = \mathbf{a}'_j \mathbf{G}_2^{\otimes \omega}, \quad (1)$$

where  $\mathbf{G}_2^{\otimes \omega}$  is the generator matrix of NB-PC and  $\otimes$  denotes the Kronecker product. According to [33], the kernel  $\mathbf{G}_2$  can be achieved by extending the Arikan kernel to the Galois field, which is written as

$$\mathbf{G}_2 = \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix}, \quad (2)$$

where  $\gamma \in GF(q) \setminus 0$  represents the non-zero element over  $GF(q)$ .

Then, the output  $\mathbf{c}'_j$  of the encoder is converted to the bit stream  $\mathbf{c}_j = [c_{j,1}, c_{j,2}, \dots, c_{j,N}]$  by the bit mapper, where  $N = pN'$ . Here, we define the code rate as  $R_c = D/N = D'/N'$ .

To mitigate interference caused by burst errors,  $\mathbf{c}_j$  is interleaved by a random interleaver, which is expressed as  $\mathbf{d}_j = \Pi(\mathbf{c}_j)$ ,  $1 \leq j \leq J$ .  $\mathbf{d}_j$  is then mapped to a  $K$ -dimensional complex codeword  $\mathbf{x}_j = [\mathbf{x}_{j,1}, \mathbf{x}_{j,2}, \dots, \mathbf{x}_{j,E}]$  by the SCMA encoder, where the  $e$ -th ( $1 \leq e \leq E$ ) codeword is a sparse column vector  $\mathbf{x}_{j,e} = [x_{j,e}^1, x_{j,e}^2, \dots, x_{j,e}^K]^T$ . Supposing the modulation order (i.e., codebook cardinality) of SCMA is  $M$ , the codeword symbol length is  $E = N/R$ , where  $R = \log_2 M$ . The resource sharing structure of SCMA can be represented by a  $K \times J$  indicator matrix. For example, a case with 4 resources and 6 users is denoted as

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}, \quad (3)$$

where the rows and columns of  $\mathbf{F}$  represent the subcarrier and user layers, respectively. The  $j$ -th user occupies the  $k$ -th subcarrier if and only if the element  $F_{kj}$  in the  $k$ -th row and  $j$ -th column of  $\mathbf{F}$  is 1. To be more specific, each user's data is assigned to  $d_u$  ( $d_u \ll K$ ) resources, while  $d_r$  ( $d_r \ll J$ ) users collide over the  $k$ -th resource. The factor graph corresponding to the indicator matrix in (3) is shown in Fig. 3. There are generally two types of nodes in a factor graph, i.e., user nodes (UNs) and resource nodes (RNs). Let UN- $j$  ( $1 \leq j \leq J$ ) and RN- $k$  ( $1 \leq k \leq K$ ) denote the  $j$ -th UN and the  $k$ -th RN, respectively.

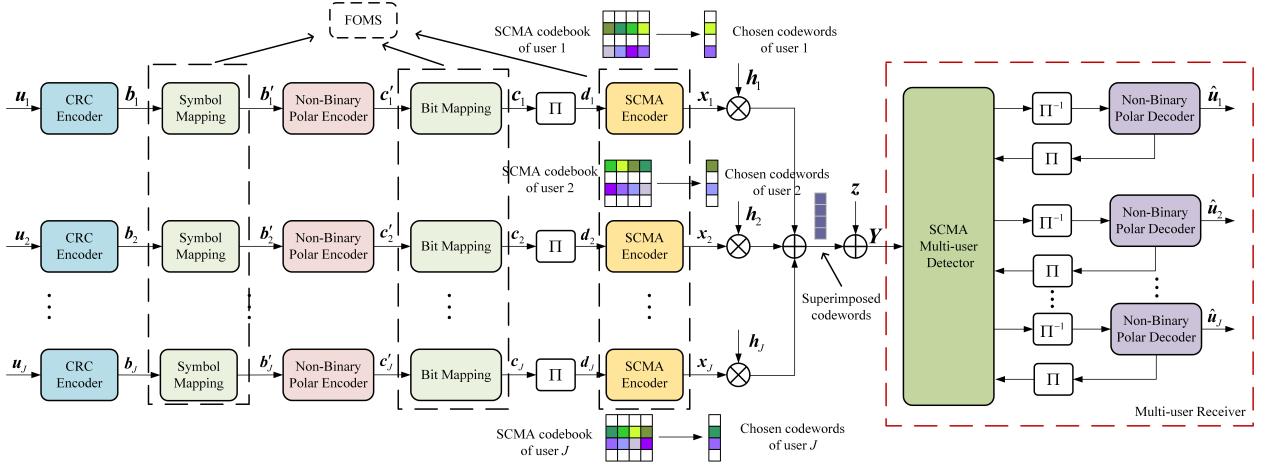


Fig. 2. FOMS based NB-PC-SCMA system.

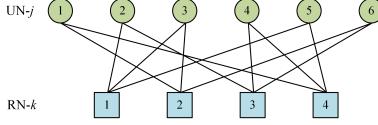


Fig. 3. SCMA factor graph for 6 users multiplexed over 4 resources.

Here, we assume that all users are time-synchronized and thus the  $e$ -th received signal is the superposition of all users' signals, which can be expressed as

$$\mathbf{y}_e = \sum_{j=1}^J \text{diag}(\mathbf{h}_{j,e}) \mathbf{x}_{j,e} + \mathbf{z}_e, \quad (4)$$

where  $\mathbf{y}_e = [y_e^1, y_e^2, \dots, y_e^K]^T$  is the received signal,  $\mathbf{h}_{j,e} = [h_{j,e}^1, h_{j,e}^2, \dots, h_{j,e}^K]$  is the channel gain vector and  $\mathbf{z}_e$  is the  $K \times 1$  additive Gaussian vector with element-wise 0 mean and covariance matrix  $N_0 \mathbf{I}_K$ . Note that  $h_{j,e}^k$  in  $\mathbf{h}_{j,e}$  denotes the channel gain of the  $e$ -th transmitted codeword between resource  $k$  and user  $j$ , while  $\mathbf{I}_K$  denotes the  $K \times K$  diagonal matrix. Finally, all received signals in a transmission block can be represented as  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_E]$ .

### III. MULTIUSER ITERATIVE RECEIVER DESIGN FOR NB-PC-SCMA SYSTEMS

In this section, the JFG and receiver for the proposed NB-PC-SCMA system are presented.

#### A. Joint Factor Graph Model for FOMS Based NB-PC-SCMA

To visualize our proposed receiver, we first introduce the JFG of NB-PC-SCMA, since the message exchange rules in the receiver solely rely on the connection of the JFG. Moreover, the designed JFG equally reflects the impact of FOMS on the receiver, which corresponds to the characteristics regarding FOMS introduced in Section II-A.

Fig. 4 exemplifies the JFG with  $M = 4$ ,  $q = 16$  and  $N' = 8$ , where 6 users and 4 resources are considered. Since FOMS allows  $M \neq q$ , the symbol-level MPA and NB-SCL are carried out over  $M$ -ary and  $q$ -ary, respectively. We build the mapping

norm between the SCMA detector and the polar decoder, as shown in the transformation level in Fig. 4. Note that we term the node on the side of the polar decoder receiving a priori information as the polar node (PN) and the node at the transformation level introduced by FOMS as the intermediate node (IN). Considering  $J$  users and  $E$  SCMA codewords, JFG consists of  $J$  polar factor graphs,  $E$  SCMA factor graphs and  $JN$  INs. Each polar factor graph contains  $N'$  PNs. Let  $\mathcal{X}_{e,k}$  and  $\mathcal{V}_{e,j}$  denote RN- $k$  ( $1 \leq k \leq K$ ) and UN- $j$  ( $1 \leq j \leq J$ ) of the  $e$ -th ( $1 \leq e \leq E$ ) codeword, respectively, and let  $\mathcal{I}_{n,j}$  ( $1 \leq n \leq N$ ) and  $\mathcal{P}_{n',j}$  ( $1 \leq n' \leq N'$ ) denote the  $n$ -th IN and  $n'$ -th PN of the  $j$ -th user, respectively.

Here, each UN and PN is associated with 2 INs and 4 INs, respectively, according to Fig. 1(b). Specifically, each  $\mathcal{V}_{e,j}$  is connected to  $\mathcal{I}_{2e-1,j}$  and  $\mathcal{I}_{2e,j}$ , while each  $\mathcal{P}_{n',j}$  is connected to  $\mathcal{I}_{4n'-3,j}$ ,  $\mathcal{I}_{4n'-2,j}$ ,  $\mathcal{I}_{4n'-1,j}$ , and  $\mathcal{I}_{4n',j}$ . Therefore, we can imagine that  $\mathcal{P}_{n',j}$  is linked to  $\mathcal{V}_{2n'-1,j}$  and  $\mathcal{V}_{2n',j}$ .

We integrate the polar factor graph and SCMA factor graph of different symbol levels into one JFG using IN, where messages are passed iteratively. Thus, the receiver for the proposed NB-PC-SCMA system only requires outer iterations, i.e., the loop of the overall operation. There is no inner iteration in the component SCMA detector or polar decoder. Note that we only give the JFG example for  $M < q$ . The model for  $M > q$  is similar except for the connection pattern of IN. The result for  $M > q$  can be easily found by referring to the interpretation for the FOMS shown in Fig. 1(c).

#### B. NB-SCL and Damping Based Joint Iterative Detection and Decoding

In this section, we introduce the proposed multiuser receiver for NB-PC-SCMA systems, which jointly performs SCMA detection and polar decoding. Specifically, the NSD-JIDD message passing process is exemplified in Fig. 5, which is elaborated in Sections III-B1 to III-B4.

##### 1) SCMA Detection:

The SCMA detection process can be interpreted as message passing on the factor graph. Therefore, the connection of the

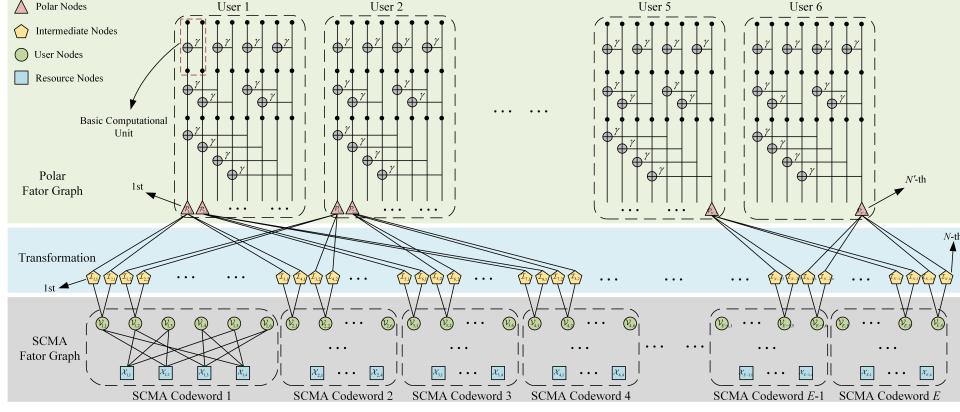


Fig. 4. The JFG for NB-PC-SCMA system.

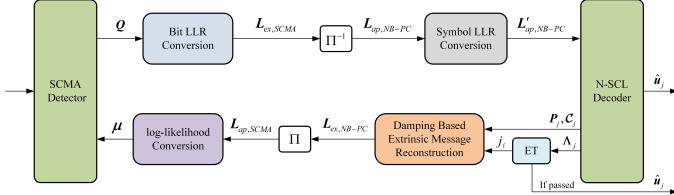


Fig. 5. Message passing in the NSD-JIDD receiver.

factor graph can be uniquely defined by a pair of sets  $\mathcal{U}(j)$  and  $\mathcal{R}(k)$  as follows,

$$\mathcal{R}(k) = \{j \mid \mathbf{F}_{k,j} = 1, 1 \leq j \leq J\}, \quad (5.a)$$

$$\mathcal{U}(j) = \{k \mid \mathbf{F}_{k,j} = 1, 1 \leq k \leq K\}. \quad (5.b)$$

According to the max-log-MPA method [34], in the  $t$ -th iteration, the messages are passed back and forth through the links in the factor graph and can be updated by (6-7). For the sake of convenience, we remove all subscripts  $e$  indicating the codeword positions when describing SCMA detector since the SCMA detection operation is identical for all codewords.

For UN- $j$  ( $1 \leq j \leq J$ ), we have

$$\begin{aligned} \xi_{j \rightarrow k}^{(t)}(x_j^k = w_{j,k}^m) &= \mathcal{N}\left(\mu^{(t-1)}(x_j^k = w_{j,k}^m)\right. \\ &\quad \left. + \sum_{i \in \mathcal{U}(j) \setminus k} \xi_{i \rightarrow j}^{(t-1)}(x_j^i = w_{j,i}^m)\right), \end{aligned} \quad (6)$$

while for RN- $k$  ( $1 \leq k \leq K$ ), we have

$$\begin{aligned} \xi_{k \rightarrow j}^{(t)}(x_j^k = w_{j,k}^m) &= \max_{\substack{x_i^k \in \mathcal{W}_{j,k}, i \in \mathcal{R}(k) \setminus j \\ x_j^k = w_{j,k}^m}} \left\{ \psi(\mathbf{x}_{[k]}) + \sum_{x_i^k \in \mathbf{x}_{[k]}, i \in \mathcal{R}(k) \setminus j} \xi_{i \rightarrow k}^{(t)}(x_i^k) \right\}, \end{aligned} \quad (7)$$

where  $\xi_{k \rightarrow j}^{(t)}(x_j^k = w_{j,k}^m)$  and  $\xi_{j \rightarrow k}^{(t)}(x_j^k = w_{j,k}^m)$  are the messages sent from RN- $k$  to UN- $j$  and from UN- $j$  to RN- $k$  in the  $t$ -th ( $1 \leq t \leq T$ ) iteration given the codeword  $x_j^k$ , respectively.  $w_{j,k}^m \in \mathcal{W}_{j,k}$  denotes the  $m$ -th ( $1 \leq m \leq M$ ) SCMA codeword of user  $j$  transmitted by subcarrier  $k$  in codebook  $\mathcal{W}$ . Here,  $\mu^{(t-1)}(x_j^k = w_{j,k}^m)$  denotes the a priori symbol log-likelihood

information input in the  $(t-1)$ -th iteration.  $\mathcal{R}(k) \setminus j$  and  $\mathcal{U}(j) \setminus k$  denote the set  $\mathcal{R}(k)$  excluding  $j$  and the set  $\mathcal{U}(j)$  excluding  $k$ , respectively. In (6),  $\mathcal{N}(\cdot)$  refers to the normalization function which ensures  $\sum_{m=1}^M \exp[\xi_{j \rightarrow k}^{(t)}(x_j^k = w_{j,k}^m)] = 1$ . Assume that

$$\begin{aligned} \xi_{j \rightarrow k}^{(t)}(x_j^k = w_{j,k}^m) &= \mu^{(t-1)}(x_j^k = w_{j,k}^m) \\ &\quad + \sum_{i \in \mathcal{U}(j) \setminus k} \xi_{i \rightarrow j}^{(t-1)}(x_j^i = w_{j,i}^m). \end{aligned} \quad (8)$$

Then  $\mathcal{N}(\cdot)$  can be further expressed as

$$\begin{aligned} \xi_{j \rightarrow k}^{(t)}(x_j^k = w_{j,k}^m) &= \xi_{j \rightarrow k}^{(t-1)}(x_j^k = w_{j,k}^m) \\ &\quad - \max_{w_{j,k}^o \in \mathcal{W}_{j,k}} \left\{ \xi_{j \rightarrow k}^{(t-1)}(x_j^k = w_{j,k}^o) \right\}. \end{aligned} \quad (9)$$

In addition, the vector function  $\psi(\mathbf{x}_{[k]})$  in (7) can be defined as

$$\psi(\mathbf{x}_{[k]}) = -\frac{1}{N_0} \left\| \mathbf{y}^k - \sum_{i \in \mathcal{R}(k)} h_i^k x_i^k \right\|^2, \quad (10)$$

where  $\mathbf{x}_{[k]} = [x_i^k]_{i \in \mathcal{R}(k)}$  is a vector comprising all symbols transmitted over the  $k$ -th subcarrier.

Therefore, the soft information output by each user can be calculated as

$$Q^{(t)}(\mathbf{x}_j = \mathbf{w}_j^m) = \sum_{i \in \mathcal{U}(j)} \xi_{i \rightarrow j}^{(t)}(x_j^i = w_{j,i}^m), \quad (11)$$

where  $\mathbf{w}_j^m \in \mathcal{W}_j$  ( $1 \leq m \leq M$ ) represents the  $m$ -th SCMA codeword of user  $j$  in codebook  $\mathcal{W}$ .

## 2) Conversion and Calculation of LLRs under FOMS:

The soft message output from the SCMA detector will be converted to bit level and further to field symbol level, which follows the JFG connection rigidly. Note that hereinafter we remove the superscript  $t$  as all operations are performed within the same iteration.

To be more specific, the exchange messages are first converted to the available log-likelihood ratio (LLR) form for the

polar decoder. The extrinsic bit LLR of the  $e$ -th ( $1 \leq e \leq E$ ) detected SCMA codeword for user  $j$  can be expressed as

$$L_{ex,scma}(\hat{d}_{j,(e-1)R+r}) = \ln \frac{\sum_{s_{j,e}^i \in \mathcal{S}_{j,e}, s_{j,e}^r=0} \exp(Q(\mathbf{x}_{j,e}))}{\sum_{s_{j,e}^i \in \mathcal{S}_{j,e}, s_{j,e}^r=1} \exp(Q(\mathbf{x}_{j,e}))}, \quad (12)$$

where  $\mathcal{S}_{j,e} = \{s_{j,e}^1, s_{j,e}^2, \dots, s_{j,e}^R\}$  ( $1 \leq r, i \leq R$  and  $i \neq r$ ) represents a bit set that can be mapped to SCMA codeword  $\mathbf{x}_{j,e}$  by codebook  $\mathcal{W}_j$  of user  $j$ . In other words, each SCMA codeword message corresponds to  $R$  bit messages, which conforms to the mapping relationship at the transmitter. Using the Jacobi approximation, we can further derive

$$\begin{aligned} L_{ex,scma}(\hat{d}_{j,(e-1)R+r}) &= \max_{s_{j,e}^i \in \mathcal{S}_{j,e}, s_{j,e}^r=0} Q(\mathbf{x}_{j,e}) \\ &\quad - \max_{s_{j,e}^i \in \mathcal{S}_{j,e}, s_{j,e}^r=1} Q(\mathbf{x}_{j,e}). \end{aligned} \quad (13)$$

After the extrinsic bit LLRs  $\mathbf{L}_{ex,scma}(\hat{d}_j) = [L_{ex,scma}(\hat{d}_{j,1}), L_{ex,scma}(\hat{d}_{j,2}), \dots, L_{ex,scma}(\hat{d}_{j,N})]$  of all  $E$  codewords for user  $j$  are obtained, de-interleaving is performed to yield the a priori bit LLRs of the NB-SCL decoder, which can be expressed as

$$\mathbf{L}_{ap,nb-pc}(\hat{\mathbf{c}}_j) = \Pi^{-1}(\mathbf{L}_{ex,scma}(\hat{d}_j)). \quad (14)$$

The de-interleaved bit LLRs will be converted into a priori symbol LLRs of NB-PCs and then input to the NB-SCL decoder. Before giving the conversion operation, we explain the symbol LLRs of NB-PCs. Unlike bit-level LLRs, the symbol LLRs with more than two likelihood information cannot be defined by a single ratio. We define the  $n'$ -th ( $1 \leq n' \leq N'$ ) a priori symbol LLR vector of the non-binary polar decoder as

$$\mathbf{L}'_{ap,nb-pc}(\hat{c}'_{j,n'}) = [L'_{ap,nb-pc}(\hat{c}'_{j,n'}=0), L'_{ap,nb-pc}(\hat{c}'_{j,n'}=1), \dots, L'_{ap,nb-pc}(\hat{c}'_{j,n'}=\alpha^{q-2})]^T, \quad (15)$$

where  $\alpha$  is the primitive element of the finite field and a certain LLR is defined as

$$L'_{ap,nb-pc}(\hat{c}'_{j,n'}=\theta) = \ln \frac{\Pr(\hat{c}'_{j,n'}=0)}{\Pr(\hat{c}'_{j,n'}=\theta)}, \theta \in \mathbb{F}_q, \quad (16)$$

where  $\mathbb{F}_q$  denotes the set  $\{0, 1, \alpha, \alpha^2, \dots, \alpha^{q-2}\}$  of all elements over  $GF(q)$ .

**Theorem 1:** For user  $j$ , the  $n'$ -th NB-PC symbol LLR with estimate  $\theta$  can be defined as

$$L'_{ap,nb-pc}(\hat{c}'_{j,n'}=\theta) = \mathcal{X}_\theta \mathcal{L}_{j,n'}, \quad (17)$$

where  $\mathcal{X}_\theta = [v_1, v_2, \dots, v_p]$  is a binary row vector satisfying the mapping relationship  $\{f : \mathcal{X}_\theta \rightarrow \theta\}$  over  $GF(q)$  and  $\mathcal{L}_{j,n'}$  is a column vector of a priori bit LLRs, which can be written as

$$\mathcal{L}_{j,n'} = [L_{ap,nb-pc}(\hat{c}_{j,(n'-1)p+1}), L_{ap,nb-pc}(\hat{c}_{j,(n'-1)p+2}), \dots, L_{ap,nb-pc}(\hat{c}_{j,n'p})]^T. \quad (18)$$

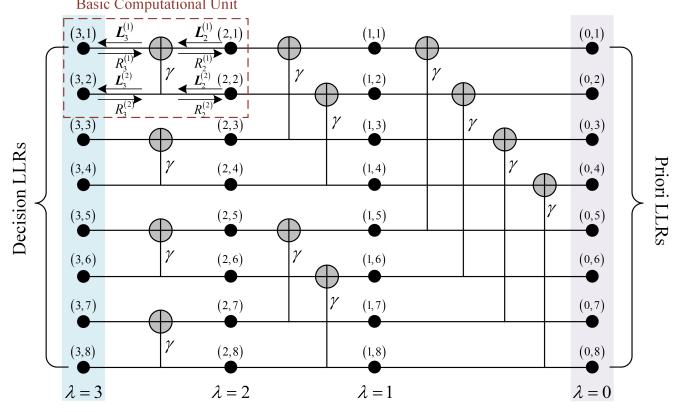


Fig. 6. The example of polar factor graph with  $N' = 8$ .

**Proof:** Typically, the  $n$ -th ( $1 \leq n \leq N$ ) a priori bit LLR of user  $j$  can be defined as

$$L_{ap,nb-pc}(\hat{c}_{j,n}) = \ln \frac{\Pr(\hat{c}_{j,n}=0)}{\Pr(\hat{c}_{j,n}=1)}. \quad (19)$$

According to the connection of JFG described in Section III-A, we can conclude that  $\mathcal{P}_{n',j}$  is associated with  $\mathcal{I}_{(n'-1)p+1,j}, \mathcal{I}_{(n'-1)p+2,j}, \dots, \mathcal{I}_{n'p,j}$ , which implies that  $\Pr(\hat{c}'_{j,n'})$  can be expressed as a joint probability mass function with respect to  $\Pr(\hat{c}_{j,(n'-1)p+i})$ . When  $\hat{c}'_{j,n'} = \theta$ , the function can be written as

$$\Pr(\hat{c}'_{j,n'} = \theta) = \prod_{i=1}^p \Pr(\hat{c}_{j,(n'-1)p+i} = v_i), v_i \in \mathcal{X}_\theta. \quad (20)$$

Then, by substituting (20) into (16), we can derive

$$\begin{aligned} L'_{ap,nb-pc}(\hat{c}'_{j,n'} = \theta) &= \sum_{i=1}^p \ln \frac{\Pr(\hat{c}_{j,(n'-1)p+i} = 0)}{\Pr(\hat{c}_{j,(n'-1)p+i} = v_i)} \\ &= \sum_{1 \leq i \leq p, v_i \neq 0} L_{ap,nb-pc}(\hat{c}_{j,(n'-1)p+i}). \end{aligned} \quad (21)$$

We can find that the result of (21) is equivalent to (17). ■

Note that  $L'_{ap,nb-pc}(\hat{c}'_{j,n'} = 0)$  is always equal to 0 and thus does not require calculation via (17). Finally, we can obtain the all a priori symbol LLRs of user  $j$ , which can be expressed as  $\mathbf{L}'_{ap,nb-pc}(\hat{\mathbf{c}}'_j) = [\mathbf{L}'_{ap,nb-pc}(\hat{c}'_{j,1}), \mathbf{L}'_{ap,nb-pc}(\hat{c}'_{j,2}), \dots, \mathbf{L}'_{ap,nb-pc}(\hat{c}'_{j,N'})]$

### 3) N-SCL Decoding and Early Termination Mechanism:

Since the polar decoding process is equivalent for each user, we remove the subscript  $j$  indicating users for brevity when describing polar decoder. After the transmitted soft information reaches the input of the non-binary polar decoder, i.e., the lower edge of the polar factor graph in the JFG shown in Fig. 4, the receiver will perform NB-SCL decoding.

To be more specific, as shown in Fig. 6, the non-binary polar factor graph with  $N' = 8$  contains  $\omega + 1$  columns indexed by  $\lambda$  and  $N'$  rows indexed by  $n'$ . Each node  $(\lambda, n')$  stores a pair of information, i.e., soft LLR information passed to the left and hard estimated information passed to the right, denoted as  $\mathbf{L}_\lambda^{(n')}$  and  $\mathbf{R}_\lambda^{(n')}$ , respectively, where  $1 \leq n' \leq N'$  and  $0 \leq \lambda \leq w$ . Moreover, let  $\mathbf{L}_\lambda = [\mathbf{L}_\lambda^{(1)}, \mathbf{L}_\lambda^{(2)}, \dots, \mathbf{L}_\lambda^{(N')}]$

and  $\mathbf{R}_\lambda = [R_\lambda^{(1)}, R_\lambda^{(2)}, \dots, R_\lambda^{(N')}]$  denote all left and right information of the  $\lambda$ -th column, respectively.

To start with, the left message  $\mathbf{L}_0$  of each user's polar decoder is initialized as the received prior information  $\mathbf{L}'_{ap,nb-pc}(\hat{\mathbf{c}}')$ , representing the rightmost input in Fig. 6. Then, the message will be passed through the basic computational unit in the factor graph, as shown in the red dashed box. We can obtain the update rules (22-25) for the  $G_2$ -based basic computational unit by converting the probability-domain based recursive function [28] to the LLR form.

$$\begin{aligned} L_\lambda^{(n')}[\theta] &= \max_{\varphi \in \mathbb{F}_q} \left\{ -\sum_{i=1}^2 L_{\lambda-1}^{(n')}[\varpi_{(0,\varphi)}^i] \right\} \\ &\quad - \max_{\varphi \in \mathbb{F}_q} \left\{ -\sum_{i=1}^2 L_{\lambda-1}^{(n'+2^{w-\lambda})}[\varpi_{(\theta,\varphi)}^i] \right\}, \end{aligned} \quad (22)$$

$$\begin{aligned} L_\lambda^{(n'+2^{w-\lambda})}[\theta] &= \sum_{i=1}^2 L_{\lambda-1}^{(n'+2^{w-\lambda})}[\varpi_{(R_\lambda^{(n')},\theta)}^i] \\ &\quad - \sum_{i=1}^2 L_{\lambda-1}^{(n')}[\varpi_{(R_\lambda^{(n')},0)}^i], \end{aligned} \quad (23)$$

$$R_{\lambda-1}^{(n')} = R_\lambda^{(n')} + \gamma \cdot R_\lambda^{(n'+2^{w-\lambda})}, \quad (24)$$

$$R_{\lambda-1}^{(n'+2^{w-\lambda})} = R_\lambda^{(n'+2^{w-\lambda})}, \quad (25)$$

where  $L_\lambda^{(n')}[\theta]$  ( $\theta \in \mathbb{F}_q$ ) in  $\mathbf{L}_\lambda^{(n')}$  denotes the LLR with the estimated value  $\theta$  and  $\varpi_\alpha^i$  can be calculated as

$$\varpi_\alpha = [\varpi_\alpha^1, \varpi_\alpha^2] = \alpha \cdot \mathbf{G}_2, \quad (26)$$

where  $\alpha \in \mathbb{F}_q^2$ .

For the path metric of the NB-SCL decoder, we modify the hardware-friendly function of the NB-PC path metric in [35] to the compatible form with the LLR defined in (16). Hence, for any path  $\ell$  and level  $n'$ , the path metric  $\rho_\ell^{(n')}$  of the NB-SCL decoder can be calculated recursively as

$$\rho_\ell^{(n')} = \rho_\ell^{(n'-1)} + L_\omega^{(n')}[\eta]_{(\ell)} - \min_{\theta \in \mathbb{F}_q} L_\omega^{(n')}[\theta]_{(\ell)}, \quad (27)$$

where  $L_\omega^{(n')}[\cdot]_{(\ell)}$  is the decision LLR for a given path  $\ell$ , representing the LLR at the decision layer shown in Fig. 6, and  $\eta \in \mathbb{F}_q$  is the  $n'$ -th estimate for the  $\ell$ -th path.

As a result, the decoded sequences with smaller path metrics can survive as candidates. Given a list size  $l$  for the NB-SCL decoder, the  $l$  most reliable paths and corresponding metrics of user  $j$  are expressed as  $\Lambda'_j = [\hat{a}'_{j1}, \hat{a}'_{j2}, \dots, \hat{a}'_{jl}]$  and  $\mathbf{P}_j = [\rho_{j1}^{(N')}, \rho_{j2}^{(N')}, \dots, \rho_{jl}^{(N')}]$ , respectively, where  $\hat{a}'_{je}$  and  $\rho_{je}^{(N')}$  ( $1 \leq \ell \leq l$ ) represent the  $\ell$ -th candidate symbol sequence and path metric, respectively. The codeword symbol sequences of the surviving paths are denoted as  $\mathcal{C}'_j = [\hat{c}'_{j1}, \hat{c}'_{j2}, \dots, \hat{c}'_{jl}]$ , i.e., the hard estimate information  $\mathbf{R}_0$  reaching the rightmost side, where  $\hat{c}'_{je}$  is the  $\ell$ -th codeword symbol sequence. After bit mapping, we can get  $l$  candidate bit paths  $\Lambda_j = [\hat{a}_{j1}, \hat{a}_{j2}, \dots, \hat{a}_{jl}]$  and codeword bits  $\mathcal{C}_j = [\hat{c}_{j1}, \hat{c}_{j2}, \dots, \hat{c}_{jl}]$ .

Assume that the  $\ell$ -th candidate path  $\hat{a}_{je}$  can pass the CRC and show the smallest path metric. Here, if all paths fail the

CRC,  $\hat{a}_{je}$  only means the path with the smallest metric. Then the NB-SCL decoder outputs the parameter  $j_\ell$  for the next stage and performs the CRC check for the next user. In the last iteration, the information bits corresponding to path  $\hat{a}_{je}$  are directly output as the estimated sequence  $\hat{u}_j$  of user  $j$ . The receiver performs ET and directly outputs the estimated information bits  $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_J$  for all users if and only if the optimal path selected by each user passes the CRC.

#### 4) Damping based Extrinsic Message Reconstruction and Prior Information Update:

Damping technique is an effective scheme to mitigate the error propagation problem and accelerate the convergence [36]. In this paper, we introduce a damping factor  $\varepsilon \in (0, 1]$  to compress the extrinsic message output by the polar decoder.

After NB-SCL decoding, the codeword bits  $\mathcal{C}_j$ , the corresponding path metric  $\mathbf{P}_j$ , and the index  $j_\ell$  of the selected path will be used for extrinsic message reconstruction. In this paper, we use the Bayes rule to calculate the likelihood information of codeword bits. The path metric  $\rho_{je}^{(N')}$  of each candidate path is first normalized to

$$\delta_{j_\ell} = \exp(-\rho_{je}^{(N')}) / \sum_{1 \leq i \leq l} \exp(-\rho_{ji}^{(N')}). \quad (28)$$

Then, the probability that the  $n$ -th ( $1 \leq n \leq N$ ) bit takes  $\phi$  ( $\phi \in \{0, 1\}$ ) can be obtained by

$$\Pr(\hat{c}_{j,n} = \phi) = \sum_{1 \leq \ell \leq l, \hat{c}_{j_\ell} = \phi} \delta_{j_\ell}. \quad (29)$$

Therefore, the extrinsic bit LLR of the NB-SCL decoder can be written as

$$L_{ex,nb-pc}(\hat{c}_{j,n}) = \begin{cases} -\infty & \Pr(\hat{c}_{j,n} = 0) = 0 \\ \ln \frac{\Pr(\hat{c}_{j,n}=0)}{\Pr(\hat{c}_{j,n}=1)} & \Pr(\hat{c}_{j,n} = 0) \neq 0 \& \Pr(\hat{c}_{j,n} = 1) \neq 0 \\ +\infty & \Pr(\hat{c}_{j,n} = 1) = 0 \end{cases}. \quad (30)$$

We can correct the extrinsic bit LLR of user  $j$  with the selected codeword bit path  $\hat{c}_{j_\ell}$ , which can be expressed as

$$L_{ex,nb-pc}(\hat{c}_{j,n}) = (1 - 2\hat{c}_{j_\ell,n}) |L_{ex,nb-pc}(\hat{c}_{j,n})|. \quad (31)$$

Before sent to the SCMA detector, the messages in the current iteration are directly damped by

$$\mathbf{L}_{ex,nb-pc}^{(t)}(\hat{\mathbf{c}}_j) = \varepsilon \mathbf{L}_{ex,nb-pc}^{(t)}(\hat{\mathbf{c}}_j), \quad (32)$$

where  $\mathbf{L}_{ex,nb-pc}^{(t)}$  denotes the extrinsic message of NB-PCs in the  $t$ -th iteration.

Note that the extrinsic messages of SCMA are not moderated since the damping of the extrinsic messages output by MPA detectors and polar decoder lead to a similar a priori information behavior.

After the damping-aided extrinsic message reconstruction, we can obtain the extrinsic information of the polar decoder, which is then be interleaved as a priori bit LLR of the SCMA detector as

$$\mathbf{L}_{ap,scma}(\hat{\mathbf{d}}_j) = \Pi(\mathbf{L}_{ex,nb-pc}(\hat{\mathbf{c}}_j)). \quad (33)$$

Finally, the interleaved bit LLRs are remapped to the symbol log-likelihood of SCMA and are sent into the SCMA detector

**Algorithm 1:** NB-SCL and Damping Based Joint Iterative Detection and Decoding

---

**Input:** received signals  $\mathbf{Y}$ , maximum number of iterations  $T$  and damping factor  $\varepsilon$

**Output:** hard decisions of decoded bits  $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_J$

- 1 **Initialization:**  $\xi_{j \rightarrow k}^{(0)}(x_j^k = w_{j,k}^m) = 0$  and  $\mu^{(0)}(x_j^k = w_{j,k}^m) = \log \frac{1}{M}$  for  $\forall w_{j,k}^m \in \mathcal{W}_{j,k}$ ,  $j = 1, 2, \dots, J$  and  $k = 1, 2, \dots, K$
- 2 **for**  $t = 1, 2, \dots, T$  **do**
  - 3 Perform SCMA detection using (6-7) and (11);
  - 4 **for**  $e = 1, 2, \dots, E$  **do**
    - 5 **for**  $j = 1, 2, \dots, J$  **do**
      - 6 Calculate extrinsic bit LLRs  $\mathbf{L}_{ex,scma}(\hat{d}_j)$  of SCMA using (12);
    - 7 **for**  $j = 1, 2, \dots, J$  **do**
      - 8 De-interleave  $\mathbf{L}_{ex,scma}(\hat{d}_j)$  to obtain a priori bit LLRs  $\mathbf{L}_{ap,nb-pc}(\hat{c}_j)$  of NB-PCs;
      - 9 **for**  $n' = 1, 2, \dots, N'$  **do**
        - 10 Calculate the a priori symbol LLRs  $\mathbf{L}'_{ap,nb-pc}(\hat{c}'_j)$  of NB-PCs for  $\forall \theta \in \mathbb{F}_q$  using (17) to give the input of the NB-SCL decoder;
      - 11 Perform NB-SCL decoding for user  $j$  using (22-27) to obtain  $\Lambda_j$ ,  $\mathcal{C}_j$  and  $\mathbf{P}_j$ ;
    - 12 **if** CRC passes **then**
      - 13 Output the decoded sequences  $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_J$ ;
      - 14 Break; // Activate ET
    - 15 **for**  $j = 1, 2, \dots, J$  **do**
      - 16 Perform damping based extrinsic message reconstruction using (28-32);
      - 17 Interleave  $\mathbf{L}_{ex,nb-pc}(\hat{c}_j)$  to get a priori bit LLRs  $\mathbf{L}_{ap,scma}(\hat{d}_j)$  of SCMA;
    - 18 **for**  $e = 1, 2, \dots, E$  **do**
      - 19 **for**  $j = 1, 2, \dots, J$  **do**
        - 20 Calculate the a priori symbol log-likelihood of SCMA employing (34);
  - 21 Get the decoded sequences  $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_J$ ;

---

subsequently as the a priori information for the next iteration, which can be expressed as (34) by employing the Jacobi approximation.

$$\mu(x_{j,e}^k) = \sum_{s_{j,e}^r \in \mathcal{S}_{j,e}} \left\{ (1 - s_{j,e}^r) L_{ap,scma}(\hat{d}_{j,(e-1)R+r}) - \max \left[ 0, L_{ap,scma}(\hat{d}_{j,(e-1)R+r}) \right] \right\} \quad (34)$$

Note that the calculation of  $\mu(x_{j,e}^k)$  in (34) is identical for  $\forall k \in [1, K]$ , which only depends on the  $e$ -th codeword mapped by the codebook  $\mathcal{W}_j$ . The proposed NSD-JIDD algorithm for the NB-PC-SCMA system is summarized in Algorithm 1.

#### IV. OPTIMIZATION SCHEME OF NSD-JIDD ALGORITHM

In this section, we conceive an optimized implementation of NSD-JIDD, namely the OSD-JIDD algorithm. To be more specific, the L-NB-SCL decoding algorithm is proposed in Section IV-A to optimize the path search pattern in NB-SCL decoding, leading to a complexity reduction. Then, in Section IV-B, we optimize the UN update for SCMA detection to mitigate the convergence errors during the iterations.

##### A. Lazy-Search Based NB-SCL Decoder

Since NB-SCL decoding adopts a global optimal search strategy over  $GF(q)$ , the decoder searches  $q$  paths at each information position, leading to redundant computation. The Monte Carlo simulation shows that the symbols at certain positions are completely decoded correctly. Therefore, we consider these high-reliable positions can be decided directly during decoding. In other words, we introduce a local lazy search for NB-SCL decoding. The resultant decoding algorithm, i.e., L-NB-SCL decoding, only seeks paths at unreliable positions evaluated by Monte Carlo simulations.

After the Monte Carlo simulation, let set  $\mathcal{B}$  denote the position where the decision can be achieved directly, i.e., the high-reliable position. To be more specific,  $g_i$  denotes the error rate of the  $i$ -th ( $1 \leq i \leq N'$ ) symbol in the Monte Carlo simulation. Then, we define the set of lazy positions as  $\mathcal{B} = \{i | g_i \leq g_{th}\}$  with cardinality  $\chi = |\mathcal{B}|$ , where  $g_{th}$  is the selected threshold with the constraint  $0 \leq g_{th} \leq 10^{-3}$  (assuming 10,000 Monte Carlo simulations are performed). Naturally, set  $\mathcal{B}$  is a subset of information position set  $\mathcal{A}$ . The lazy search rate is then defined as  $\beta = \chi / D'$ . In particular, the lazy search rates at  $g_{th} = 0$  and  $g_{th} = 10^{-3}$  are denoted as  $\beta_{\min}^*$  and  $\beta_{\max}^*$ , respectively. Therefore, we have the bound  $\beta \in [\beta_{\min}^*, \beta_{\max}^*]$ .

Taking the  $GF(4)$  and  $GF(16)$  based NB-PC as an example, Table II gives the thresholds of  $\beta$  for different code lengths  $N'$  and code rates  $R_c$ , where Monte Carlo simulations are carried out at a signal-to-noise ratio (SNR) of 2 dB. To obtain complexity reduction while maintaining performance, an appropriate  $\beta$  is required to balance the trade-off between complexity and BER performance.

Different from the NB-SCL decoder, the L-NB-SCL decoder performs a lazy search, i.e., hard decision to reduce the split paths if  $n' \in \mathcal{B}$ . The hard decision function for estimate  $\hat{a}'_{\ell,n'}$  in the  $\ell$ -th path can be expressed as

$$\hat{a}'_{\ell,n'} = \arg \min_{\theta \in \mathbb{F}_q} L_{\omega}^{(n')}[\theta]_{(\ell)}. \quad (35)$$

Note that if  $n' \in \mathcal{B}$ , an update for the path metric is no longer required, i.e.,  $\rho_{\ell}^{(n')} = \rho_{\ell}^{(n'-1)}$ , since the selected path complies with the SC decision, leading to a burden value of 0 in (27).

The path search process in SCL decoding can be characterized by a decoding tree. Fig. 7(a) and Fig. 7(b) show the  $GF(4)$ -based NB-SCL decoding tree and L-NB-SCL decoding tree, respectively. Typically, the  $N'$ -length  $GF(q)$ -based NB-PC decoding tree is a  $q$ -ary tree with depth  $N'$ . Here, the topmost root node denotes a null state and the number adjacent

TABLE II  
THE AVAILABLE RANGE OF LAZY SEARCH RATE WITH DIFFERENT CODE LENGTHS AND CODE RATES.

$q$	$\beta$	$N' = 64$			$N' = 128$			$N' = 256$			$N' = 512$		
		$R_c = \frac{1}{3}$	$R_c = \frac{1}{2}$	$R_c = \frac{2}{3}$	$R_c = \frac{1}{3}$	$R_c = \frac{1}{2}$	$R_c = \frac{2}{3}$	$R_c = \frac{1}{3}$	$R_c = \frac{1}{2}$	$R_c = \frac{2}{3}$	$R_c = \frac{1}{3}$	$R_c = \frac{1}{2}$	$R_c = \frac{2}{3}$
4	$\beta_{\min}^*$	0.524	0.536	0.512	0.581	0.578	0.565	0.718	0.711	0.684	0.825	0.816	0.763
	$\beta_{\max}^*$	0.667	0.656	0.628	0.767	0.734	0.729	0.859	0.836	0.801	0.942	0.918	0.850
16	$\beta_{\min}^*$	0.571	0.563	0.558	0.721	0.719	0.706	0.871	0.859	0.825	0.924	0.910	0.889
	$\beta_{\max}^*$	0.762	0.750	0.721	0.837	0.828	0.812	0.941	0.930	0.895	0.959	0.941	0.936

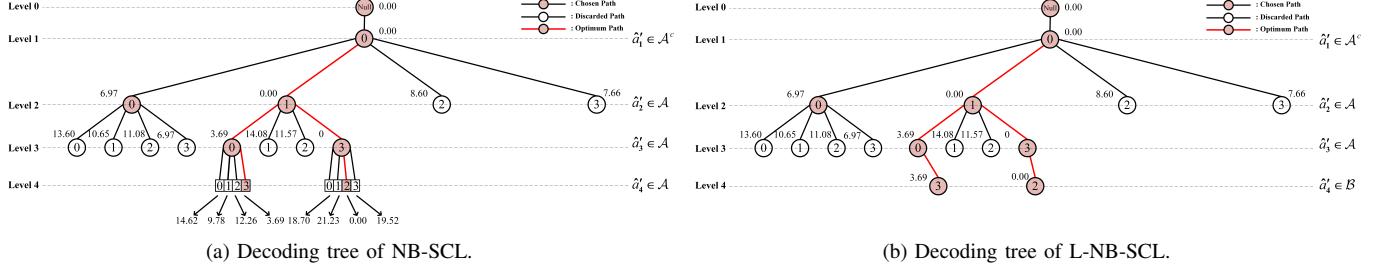


Fig. 7. Example of a non-binary polarized decoding tree, where  $N' = 4$  and  $l = 2$ .

to each node denotes the corresponding path metric. From Fig. 7, compared with NB-SCL decoding, L-NB-SCL decoding splits only one search path when the level  $n' \in \mathcal{B}$ . Generally, L-NB-SCL decoding can reach list saturation slower and avoid redundant global search.

Assume that the list is unfull with  $l_{pre}$  paths surviving at the previous level. If the current stage belongs to  $\mathcal{B}$ , the search paths for L-NB-SCL and NB-SCL are  $l_{now} = l_{pre}$  and  $l_{now} = ql_{pre}$ , respectively. Then, at the next information level belonging to  $\mathcal{B}^c$  (the complement of the set  $\mathcal{B}$ ), the search paths for both are  $l_{now} = ql_{pre}$  and  $l_{now} = q^2l_{pre}$ , respectively. Thus, the calculation of  $(q^2 - q)l_{pre}$  LLRs and path metrics are saved by L-NB-SCL decoding.

The L-NB-SCL decoding is summarized in Algorithm 2, where  $\mathbf{R}_{\lambda(\ell)} = [R_{\lambda(\ell)}^{(1)}, R_{\lambda(\ell)}^{(2)}, \dots, R_{\lambda(\ell)}^{(N')}]$  ( $0 \leq \lambda \leq w$ ) denotes the right message of the  $\lambda$ -th column in the  $\ell$ -th factor graph. Note that for simplicity, the subscript  $j$  representing user is omitted in Algorithm 2.

### B. MPA Detection Optimization

During the UN update of SCMA detection, the conventional JIDD scheme employs the SCMA a priori information from the decoder and the information passed by RN in the prior iteration. This process leads to heavy storage of intermediate variables and the unavailability of update messages in time. Especially, the results of SCMA detection at low  $E_b/N_0$  are not reliable. Thus, for the UN update process, the information passed by RN may lead to convergence errors.

Explicitly, the dominant belief information passed by the UN should be the a priori information  $\mu^{(t-1)}$  in (6) since the latest a priori information from the non-binary polar decoder provides high reliability. By contrast, the passed information  $\xi_{i \rightarrow j}^{(t-1)}$  in the prior iteration is not an instant message and performs weakly at low  $E_b/N_0$ . After  $\xi_{i \rightarrow j}^{(t-1)}$  is employed to

update  $\mathbf{L}_{ex,scma}$  and  $\mathbf{L}_{ap,nb-pc}$ , the information  $\mu^{(t-1)}$  will contain the main messages in  $\xi_{i \rightarrow j}^{(t-1)}$  and will be updated again.

As such, we optimize the MPA detection in NSD-JIDD by removing the second term  $\xi_{i \rightarrow j}^{(t-1)}$  in (6). In addition, the a priori information  $\mu^{(t-1)}$  is strictly normalized since  $\mathbf{L}_{ex,nb-pc}$  is constructed by the Bayes rule. Then, the  $\mathcal{N}(\cdot)$  operation can be discarded and thus save calculations. The modified UN update can be expressed as

$$\xi_{j \rightarrow k}^{(t)} (x_j^k = w_{j,k}^m) = \mu^{(t-1)} (x_j^k = w_{j,k}^m). \quad (36)$$

Then, we can obtain the modified RN update rule by substituting (36) into (7) as follows,

$$\begin{aligned} & \xi_{k \rightarrow j}^{(t)} (x_j^k = w_{j,k}^m) \\ &= \max_{\substack{x_i^k \in \mathcal{W}_{i,k}, i \in \mathcal{R}(k) \setminus j \\ x_j^k = w_{j,k}^m}} \left\{ \psi(x_{[k]}) + \sum_{x_i^k \in \mathcal{X}_{[k]}, i \in \mathcal{R}(k) \setminus j} \mu^{(t-1)} (x_i^k) \right\}. \end{aligned} \quad (37)$$

Therefore, the UN update in the optimized MPA can be merged into the a priori information update. In this way, the OSD-JIDD does not comprise the UN update process, where the MPA detection stage can be represented by (37).

The partial message passing for NSD-JIDD and OSD-JIDD are described in Fig. 8(a) and Fig. 8(b), respectively, where the same parameters as Fig. 4 are considered. For NSD-JIDD, the messages passed by each RN are stored for the UN update in the next iteration. Whereas for OSD-JIDD, it can be interpreted that each UN integrates with  $R$  INs, instead of being stored for previous messages, which saves the system memory.

The overall steps of the OSD-JIDD algorithm are identical to the NSD-JIDD algorithm, only with the modifications of SCMA detection and polar decoding. Specifically, if we replace step 3 in Algorithm 1 by (37) and (11) to update the

**Algorithm 2:** Lazy-Search Based NB-SCL Decoding

**Input:** maximum search width  $l$ , information position set  $\mathcal{A}$ , high-reliable position set  $\mathcal{B}$  and LLR initial value  $\mathbf{L}_0$

**Output:** path metrics  $\mathbf{P} = [\rho_1^{(N')}, \rho_2^{(N')}, \dots, \rho_l^{(N')}]$ ,  $l$  bit paths  $\hat{\mathbf{a}}_\ell = [\hat{a}_{\ell,1}, \hat{a}_{\ell,2}, \dots, \hat{a}_{\ell,N'}]$  and codeword bits  $\hat{\mathbf{c}}_\ell = [\hat{c}_{\ell,1}, \hat{c}_{\ell,2}, \dots, \hat{c}_{\ell,N'}]$

---

**1 Initialization:**  $\mathbb{L} = \{1\}$ ,  $\rho_1^{(0)} = 0$

**2 for**  $n' = 1, 2, \dots, N'$  **do**

**3**   Update left message  $L_\omega^{(n')}[\eta]_{\langle\ell\rangle}$  for  $\forall \ell \in \mathbb{L}$  and  $\eta \in \mathbb{F}_q$  using (22-23);

**4**   **if**  $n' \notin \mathcal{A}$  **then**

**5**      $\hat{a}'_{\ell,n'} \leftarrow 0$  for  $\forall \ell \in \mathbb{L}$ ;

**6**     Calculate the path metric  $\rho_\ell^{(n')}$  for  $\forall \ell \in \mathbb{L}$  using (27);

**7**   **else**

**8**     **if**  $n' \in \mathcal{B}$  **then**

**9**       Estimate  $\hat{a}'_{\ell,n'}$  for  $\forall \ell \in \mathbb{L}$  using (35);

**10**       $\rho_\ell^{(n')} \leftarrow \rho_\ell^{(n'-1)}$  for  $\forall \ell \in \mathbb{L}$ ;

**11**   **else**

**12**     Calculate temporary metric  $\rho_{\ell,\eta}^{temp}$  for  $\forall \ell \in \mathbb{L}$  and  $\eta \in \mathbb{F}_q$  using (27);

**13**     Select  $\min\{l, q|\mathbb{L}|\}$  smallest  $\rho_{\ell,\eta}^{temp}$  and get survived symbol set  $\mathbb{Y}_\ell$  of each path;

**14**   **for**  $\ell \in \mathbb{L}$  **do**

**15**     **if**  $\rho_{\ell,\eta}^{temp}$  is selected **then**

**16**        $(\hat{a}'_{\ell,n'}, \rho_\ell^{(n')}) \leftarrow (\eta, \rho_{\ell,\eta}^{temp})$ ;

**17**       **if**  $|\mathbb{Y}_\ell| > 1$  **then**

**18**         **for**  $\theta \in \mathbb{Y}_\ell \setminus \eta$  **do**

**19**           Clone the path  $\ell$  to a new path  $\ell'$ ,  $\mathbb{L} \leftarrow \mathbb{L} \cup \ell'$ ;

**20**            $(\hat{a}'_{\ell',n'}, \rho_{\ell'}^{(n')}) \leftarrow (\theta, \rho_{\ell,\theta}^{temp})$ ;

**21**       **else**

**22**         Kill the path  $\ell$  as  $\mathbb{L} \leftarrow \mathbb{L} \setminus \ell$ ;

**23**     $R_0^{(n')}_{\langle\ell\rangle} \leftarrow \hat{a}_{\ell,n'}$ ;

**24**    Update right message employing (24-25);

**25** **for**  $\ell = 1, 2, \dots, l$  **do**

**26**     $\hat{\mathbf{c}}'_\ell \leftarrow \mathbf{R}_0_{\langle\ell\rangle}$ ;

**27**    Map  $\hat{\mathbf{a}}'_\ell$  and  $\hat{\mathbf{c}}'_\ell$  to  $\hat{\mathbf{a}}_\ell$  and  $\hat{\mathbf{c}}_\ell$ , respectively

---

UNs and calculate soft messages, respectively, and step 11 performs the L-NB-SCL decoding in Algorithm 2, then the procedures complete the OSD-JIDD algorithm.

## V. RESULTS AND DISCUSSIONS

Simulation results of the proposed NB-PC-SCMA system over AWGN and Rayleigh fading channels are given and analyzed in this section. Particularly, the BER performance, computational complexity and latency are characterized in Sections V-A to V-C, respectively.

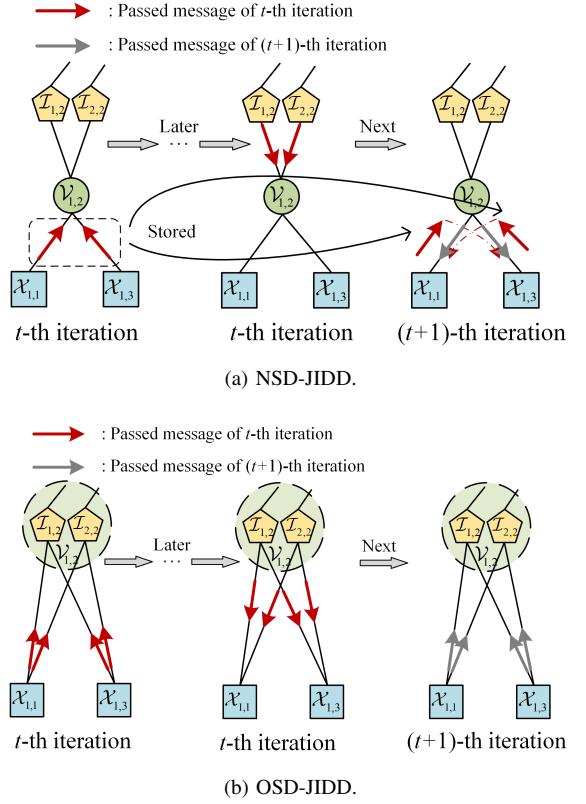


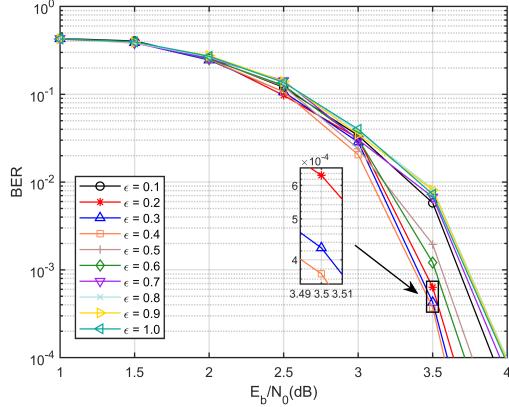
Fig. 8. Partial message passing on the factor graph.

### A. BER Performance

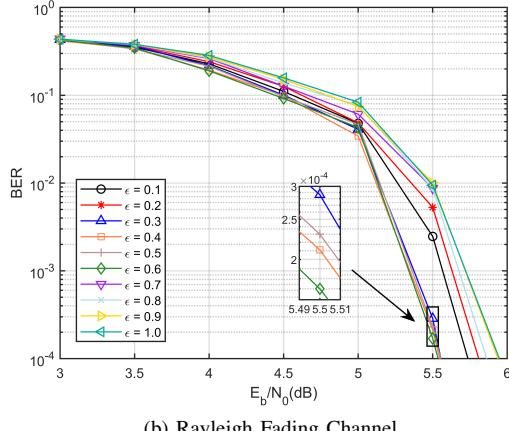
In this section, the error performance of the proposed NB-PC-SCMA system is evaluated. The SCMA codebook used in the simulation is designed according to [37]. Here, we define the  $M$ -dimension SCMA codebook with  $J$  users and  $K$  resources as  $(J, K, M)$ . For the receiver, the number of inner iterations for max-log-MPA detection is set to 1. NSD-JIDD and OSD-JIDD algorithms with 5 outer iterations are employed for multi-user detection. The NB-PCs transmitted over AWGN and Rayleigh fading channels are constructed by the Monte Carlo method at  $E_b/N_0 = 2$  dB and  $E_b/N_0 = 4$  dB, respectively, where the kernel parameter  $\gamma$  is set according to [30]. Unless otherwise specified, the field order  $q$  of the NB-PCs is 16. In addition, 16-CRC and 24-CRC are employed for the NB-PCs with  $N = 256$  and  $N = 1024$ , respectively. To limit the computational complexity, the list size  $l$  is set to 8.

Fig. 9(a) and Fig. 9(b) characterize the BER performance of NSD-JIDD with different  $\varepsilon$  over AWGN and Rayleigh fading channels, respectively, where the codebook  $(6, 4, 4)$  is considered. It can be seen that  $\varepsilon = 0.4$  and  $\varepsilon = 0.6$  achieve the best performance over AWGN and Rayleigh fading channels, respectively, and hence they are adopted for the following simulations.

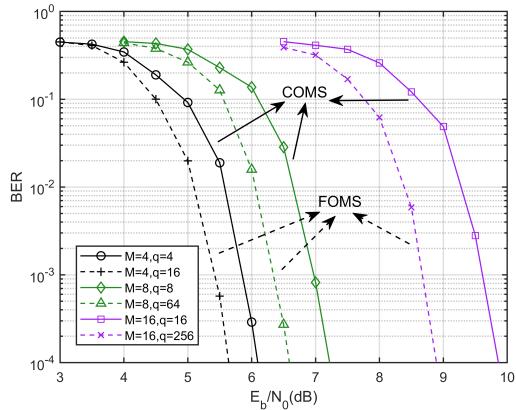
Fig. 10 shows the BER comparison between FOMS and COMS over the Rayleigh fading channel, where 4-point, 8-point, and 16-point SCMA modulation are considered. Here, the throughput in terms of bits per symbol (BPS) [7] is defined as  $\lambda R_c \log_2 M$ . It can be observed that the BER performance of both COMS and FOMS degrades with the increase of



(a) AWGN Channel.



(b) Rayleigh Fading Channel.

Fig. 9. BER performance with different  $\epsilon$ , where  $N = 256$  and  $R_c = 1/2$ .Fig. 10. The BER of NB-PC-SCMA systems with COMS and FOMS, where  $J = 6$ ,  $K = 4$ ,  $N = 256$  and  $R_c = 1/2$ .

throughput. Particularly, FOMS obtains 0.46 dB, 0.64 dB, and 0.97 dB gain versus COMS at the BER of  $10^{-4}$  for BPS of 1.5, 2.25, and 3, respectively. When  $q = 16$ , FOMS at  $M = 4$  outperforms COMS at  $M = 16$  about 4.23 dB at the cost of reduced throughput. Thus, we can adjust the order matching to trade-off the throughput and reliability, enabling the NB-PC-SCMA system with flexibility.

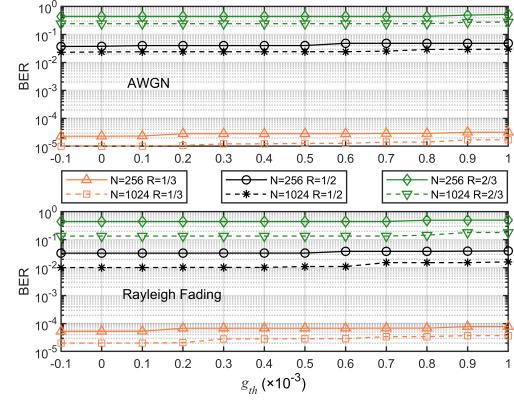
Fig. 11. The relationship between the threshold  $g_{th}$  and BER of the OSD-JIDD algorithm with different  $N$  and  $R_c$ , where  $E_b/N_0$  is set to 3 dB and 5 dB for AWGN and Rayleigh fading channels when  $N = 256$ , respectively and where  $E_b/N_0$  is set to 2.5 dB and 4.3 dB for AWGN and Rayleigh fading channels when  $N = 1024$ , respectively.

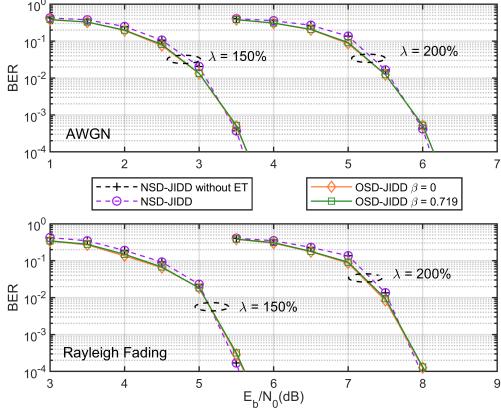
TABLE III  
SIMULATION PARAMETERS OF OSD-JIDD ALGORITHM FOR DIFFERENT CONFIGURATIONS.

Channel	Parameters	$N = 256$			$N = 1024$		
		$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
AWGN	$g_{th}$	0.1	0.5	0.8	0.2	0.7	0.8
	$\beta$	0.714	0.719	0.698	0.871	0.906	0.871
Rayleigh Fading	$g_{th}$	0.1	0.5	0.7	0.2	0.6	0.8
	$\beta$	0.714	0.719	0.674	0.859	0.898	0.871

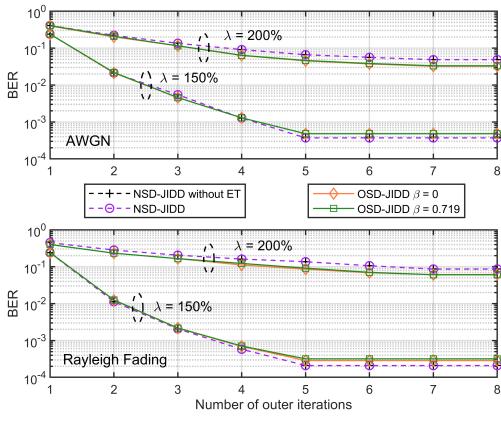
Before quantifying the performance of the OSD-JIDD algorithm for NB-PC-SCMA systems, we investigate the effect of  $g_{th}$  in L-NB-SCL decoding on the BER performance in Fig. 11, where the codebook (6, 4, 4) is considered. Note that  $g_{th} = -10^{-4}$  represents the case when NB-SCL decoding is employed. We can see that within the limit of the threshold, especially at high  $R_c$ , there is no performance degradation when L-NB-SCL decoding is adopted at the receiver. Based on Fig. 11, we select the optimal  $g_{th}$  for different configurations and calculate the corresponding  $\beta$  from Monte Carlo simulation results, as shown in Table III.

The performance comparison of OSD-JIDD and NSD-JIDD with different overloads in terms of BER and convergence are given in Fig. 12(a) and (b), respectively. For  $\lambda = 150\%$  and  $\lambda = 200\%$ , the codebooks are considered as (6, 4, 4) and (12, 6, 4), respectively. Here, the NSD-JIDD without ET and the OSD-JIDD with  $\beta = 0$  are also plotted as the benchmark.

As seen in Fig. 12(a), a higher user overload leads to a poorer BER performance. ET has no influence on the detection results at all. Moreover, OSD-JIDD shows even better performance than NSD-JIDD at low  $E_b/N_0$  since OSD-JIDD avoids some convergence errors with the aid of optimized MPA. Overall, OSD-JIDD lags NSD-JIDD only about 0.06 to 0.09 dB at the BER of  $10^{-4}$ . Fig. 12(b) shows that both NSD-JIDD and OSD-JIDD require two extra iterations to converge when  $\lambda = 200\%$ . Apparently, the convergence performance of



(a) BER performance.



(b) convergence performance.

Fig. 12. Performance comparison of OSD-JIDD and NSD-JIDD algorithms with different overloads, where  $N = 256$  and  $R_c = 1/2$

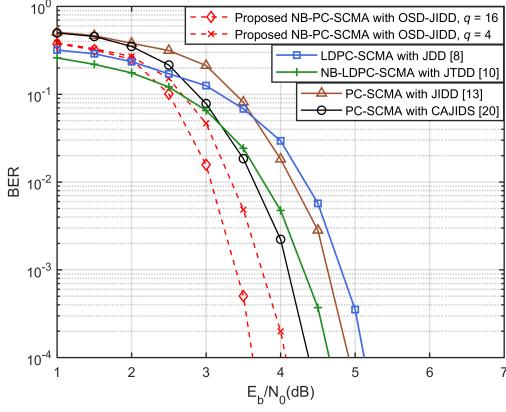


Fig. 13. BER comparison of the proposed NB-PC-SCMA to other state-of-the-art coded SCMA schemes over AWGN channels, where  $N = 256$  and  $R_c = 1/2$ .

OSD-JIDD is not suffered, which is attributed to the ET and damping mechanisms.

Finally, we compare the BER performance of the proposed NB-PC-SCMA with OSD-JIDD to other state-of-the-art schemes, where the codebook  $(6, 4, 4)$  is considered. To be more specific, LDPC-SCMA with JDD [8], non-binary

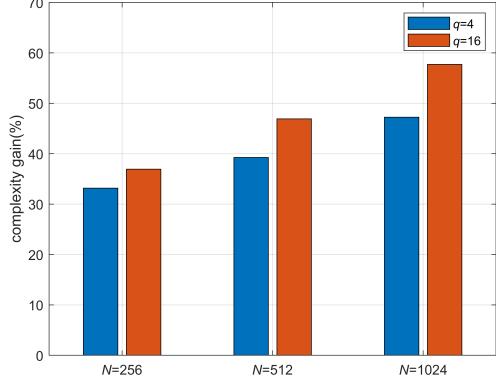


Fig. 14. Performance of complexity gain over AWGN channels for OSD-JIDD with different  $N$  and  $q$ , where  $R_c$  is set to 1/2.

LDPC (NB-LDPC) coded SCMA (NB-LDPC-SCMA) with joint trellis based joint decoding and detection (JTDD) [10], PC-SCMA with JIDD [13], and PC-SCMA with CRC aided joint iterative detection and SCL decoding (CAJIDS) [20] are simulated as counterparts. The LDPC and NB-LDPC codes in the simulations are constructed in [38] and [39], respectively, where the min-sum and extended min-sum decoders with 30 inner iterations are employed in the receiver, respectively.

As demonstrated in Fig. 13, the NB-PC-SCMA system with OSD-JIDD outperforms PC-SCMA with CAJIDS and PC-SCMA with JIDD about 0.32 dB and 0.86 dB at the BER of  $10^{-4}$  when  $q = 4$ . If the field order  $q$  is set to 16 with FOMS, the performance gains increase to 0.80 dB and 1.34 dB, respectively. Besides, a performance gain of 0.59 dB is attained by NB-PC-SCMA with  $q = 4$  compared to NB-LDPC-SCMA. In contrast to the COMS based NB-LDPC-SCMA in [10], our proposed FOMS based NB-PC-SCMA can achieve up to 1.07 dB gain when  $q = 16$ . Overall, based on the above results, we can find that the proposed NB-PC-SCMA system with OSD-JIDD shows the best error performance.

### B. Complexity

This section evaluates the decoding complexity gain for the OSD-JIDD algorithm with respect to the NSD-JIDD algorithm, which is mainly attributed to the L-NB-SCL decoding. The complexity of the list decoding is proportional to  $l$ , which determines the search width. Therefore, we apply the number  $\eta_{path}^{N'}$  of search paths in the decoding tree to reflect the decoding complexity, which can be calculated recursively as

$$\eta_{path}^i = \eta_{path}^{i-1} + \varsigma \eta_{node}^i, \quad (38)$$

where  $\eta_{path}^0 = 0$ ,  $1 \leq i \leq N'$ , and  $\eta_{node}^i$  denotes the number of split nodes at the  $i$ -th level, which can be calculated recursively as

$$\eta_{node}^i = \begin{cases} \varsigma \eta_{node}^{i-1} & \eta_{node}^{i-1} \leq l \\ \varsigma l & \eta_{node}^{i-1} > l \end{cases}, \quad (39)$$

where  $\varsigma$  in (38) and (39) denotes the the search factor. For OSD-JIDD, when  $i \in \mathcal{A}$  and  $i \notin \mathcal{B}$ , we have  $\varsigma = q$ ; otherwise,

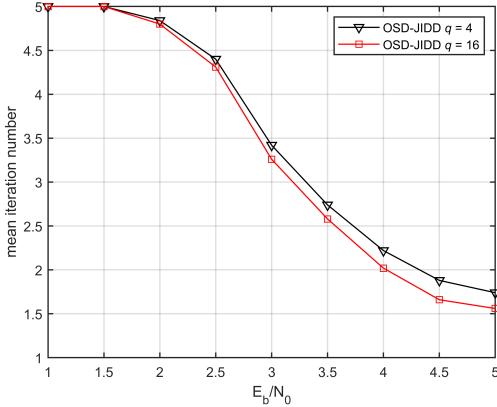


Fig. 15. The average number of iterations for OSD-JIDD over AWGN channels, where  $N = 256$  and  $R_c = 1/2$ .

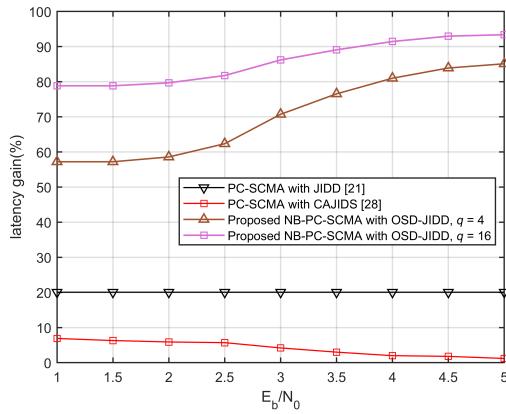


Fig. 16. Comparison of latency gain over AWGN channels, where  $N = 256$  and  $R_c = 1/2$ .

$\varsigma = 1$ . While for NSD-JIDD, when  $i \in \mathcal{A}$ , we have  $\varsigma = q$ ; otherwise,  $\varsigma = 1$ . Then, the complexity gain is characterized by the relative reduction rate of  $\eta_{path}^{N'}$ , which can be denoted as

$$\mathcal{G}_{complexity} = \frac{\eta_{path,NSD}^{N'} - \eta_{path,OSD}^{N'}}{\eta_{path,NSD}^{N'}}, \quad (40)$$

where  $\eta_{path,NSD}^{N'}$  and  $\eta_{path,OSD}^{N'}$  denote the number of search paths in the NSD-JIDD and OSD-JIDD algorithms, respectively.

Fig. 14 describes the complexity gain of OSD-JIDD with different  $N$  and  $q$  over AWGN channels, where the SCMA codebook is considered as (6, 4, 4). It can be seen that OSD-JIDD obtains higher complexity gain when  $N$  or  $q$  increases. When  $N = 1024$ , the GF(16)-based OSD-JIDD yields about 57.71% complexity gain compared with the NSD-JIDD.

### C. Latency

Here, clock cycles are used to measure the latency gain of the NB-PC-SCMA system compared to the binary counterpart. Suppose that all parallelizable instructions are carried out in one clock cycle. SCMA detection takes only three cycles, i.e.,

each of which corresponds RN update, UN update, and calculation of soft information. The latency of SCMA detection is negligible compared to polar decoding. Therefore, we will focus on the impact of polar decoding on the receiver latency.

The cycles of the binary SCL decoder can be calculated as  $\Gamma_{SCL} = 2N - 2 + D$  [40] with  $D$  cycles for path sorting and  $2N - 2$  cycles for SC decoding. The latency of the SCAN decoder can be expressed as  $\Gamma_{SCAN} = T_{SCAN}(2N - 2)$ , where  $T_{SCAN}$  is the number of SCAN inner iterations. The latency of the L-NB-SCL decoder is written as  $\Gamma_{LN-SCL} = 2N' - 2 + (1 - \beta)D'$ . Thus, the latency for the JIDD in [13] can be expressed as

$$\Gamma_{JIDD} = T \cdot T_{SCAN}(2N - 2) = T(2N - 2). \quad (41)$$

According to [20], the latency for the CAJIDS can be calculated as

$$\Gamma_{CAJIDS} = \sum_{t=1}^T 2N_t - 2 + D_t, \quad (42)$$

where  $N_t$  and  $D_t$  denote the number of decoded bits and information bits before ET in the  $t$ -th iteration for the DCA-SCL decoding, respectively. The latency for our proposed OSD-JIDD can be expressed as

$$\begin{aligned} \Gamma_{OSD-JIDD} &= T_a(2N' - 2 + (1 - \beta)D') \\ &= \frac{T_a}{p}[2N - 2p + (1 - \beta)D], \end{aligned} \quad (43)$$

where  $T_a$  denotes the average number of iterations for the OSD-JIDD. To make a fair comparison, we employ the latency of a multiuser receiver with conventional SCL decoding as the benchmark. Then, the latency gain for algorithm  $\mathbf{X}$  can be written as

$$\mathcal{G}_{latency,\mathbf{X}} = \frac{T \cdot \Gamma_{SCL} - \Gamma_{\mathbf{X}}}{T \cdot \Gamma_{SCL}}, \quad (44)$$

where  $\Gamma_{\mathbf{X}}$  denotes the latency for a given algorithm  $\mathbf{X}$ , i.e.,  $\Gamma_{\mathbf{X}} \in \{\Gamma_{JIDD}, \Gamma_{CAJIDS}, \Gamma_{OSD-JIDD}\}$ .

Before characterizing the latency performance,  $T_a$  versus  $E_b/N_0$  over AWGN channels is presented in Fig. 15. It can be observed that with the improvement of  $E_b/N_0$ , the ET mechanism leads to a decrease of  $T_a$ . Especially at  $E_b/N_0 = 5$  dB, the OSD-JIDD with  $q = 16$  converges and activates ET after 1.55 average iterations, which contrasts significantly with the maximum number of iterations  $T = 5$ .

The latency gain performance for different schemes is presented in Fig. 16, where the same parameters as Fig. 13 are considered. It can be found that for the PC-SCMA system, JIDD has a 20% latency gain regardless of  $E_b/N_0$ , while CAJIDS exhibits a weak latency gain at high  $E_b/N_0$ . By contrast, for the proposed NB-PC-SCMA system, OSD-JIDD obtains about 57% to 92% latency gain, attributed to the characteristics of NB-PC, ET mechanism and L-NB-SCL decoding. Benefiting from FOMS, this gain becomes larger as  $q$  increases. Note that when  $E_b/N_0 = 5$  dB, a latency saving of 92% can be observed by OSD-JIDD with  $q = 16$ .

## VI. CONCLUSION

In this paper, we have presented the design of the FOMS-based NB-PC-SCMA system for the first time. Specifically, an NSD-JIDD receiver have been proposed to guarantee the BER performance. Moreover, the OSD-JIDD algorithm with L-NB-SCL decoding and optimized MPA have been proposed to reduce computational complexity and convergence error. Simulation results show that the proposed NB-PC-SCMA system leads to the best BER performance with up to 92% latency gain compared to other state-of-the-art architecture. Furthermore, OSD-JIDD achieve about 33% to 57% complexity reduction over NSD-JIDD without significant BER loss.

## REFERENCES

- [1] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, Sep. 2015.
- [2] H. Nikopour and H. Baligh, "Sparse code multiple access," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2013, pp. 332–336.
- [3] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, Oct 2017.
- [4] M. Rebhi, K. Hassan, K. Raouf, and P. Charg, "Sparse code multiple access: Potentials and challenges," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1205–1238, 2021.
- [5] Y. Wu, S. Zhang, and Y. Chen, "Iterative multiuser receiver in sparse code multiple access systems," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 2918–2923.
- [6] Z. Zhang, K. Han, J. Hu, and J. Chen, "Joint detection and decoding schemes for turbo coded SCMA systems," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–6.
- [7] Y. Liu, L. Xiang, R. G. Maunder, L.-L. Yang, and L. Hanzo, "Hybrid iterative detection and decoding of near-instantaneously adaptive turbo-coded sparse code multiple access," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.
- [8] K. Han, Z. Zhang, J. Hu, and J. Chen, "A high performance joint detection and decoding scheme for LDPC coded SCMA system," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–6.
- [9] W.-C. Sun, Y.-C. Su, Y.-L. Ueng, and C.-H. Yang, "An LDPC-coded SCMA receiver with multi-user iterative detection and decoding," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 9, pp. 3571–3584, Sep. 2019.
- [10] K. Lai, L. Wen, J. Lei, P. Xiao, A. Maaref, and M. A. Imran, "Sub-graph based joint sparse graph for sparse code multiple access systems," *IEEE Access*, vol. 6, pp. 25 066–25 080, 2018.
- [11] Q. He, B. Bai, D. Feng, H. Xu, and M. Zhu, "A nonbinary LDPC-coded SCMA system with optimized codebook design," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–6.
- [12] S. Jing, C. Yang, J. Yang, X. You, and C. Zhang, "Joint detection and decoding of polar-coded scma systems," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2017, pp. 1–6.
- [13] Z. Pan, E. Li, L. Zhang, J. Lei, and C. Tang, "Design and optimization of joint iterative detection and decoding receiver for uplink polar coded SCMA system," *IEEE Access*, vol. 6, pp. 52 014–52 026, 2018.
- [14] J. Miao, X. Hu, W. Li, and H. Li, "A low complexity joint iterative multi-user detection decoding receiver based on verified message," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, Aug 2020, pp. 699–704.
- [15] Y. Zhang, W. Ge, P. Zhang, and M. Gao, "The optimization scheme for joint iterative detection and decoding of polar coded SCMA system," *Optical Fiber Technology*, vol. 58, p. 102283, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S106852002030273X>
- [16] Y. Zhang, W. Ge, P. Zhang, M. Gao, and G. Zhang, "A joint detection and decoding scheme for PC-SCMA system based on pruning iteration," *Symmetry*, vol. 12, no. 10, 2020. [Online]. Available: <https://www.mdpi.com/2073-8994/12/10/1624>
- [17] J. Dai, K. Niu, Z. Si, C. Dong, and J. Lin, "Polar-coded non-orthogonal multiple access," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1374–1389, March 2018.
- [18] L. Karakchieva and P. Trifonov, "Joint list multistage decoding with sphere detection for polar coded SCMA systems," in *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, Feb 2019, pp. 1–6.
- [19] H. Mu, Y. Tang, L. Li, Z. Ma, P. Fan, and W. Xu, "Polar coded iterative multiuser detection for sparse code multiple access system," *China Communications*, vol. 15, no. 11, pp. 51–61, Nov 2018.
- [20] X. Wu, Y. Wang, and C. Li, "Low-complexity CRC aided joint iterative detection and SCL decoding receiver of polar coded SCMA system," *IEEE Access*, vol. 8, pp. 220 108–220 120, 2020.
- [21] X. Wu and Y. Wang, "Improving polar-coded SCMA system by information coupling and parity check," *Sensors*, vol. 20, no. 23, 2020. [Online]. Available: <https://www.mdpi.com/1424-8220/20/23/6740>
- [22] J. Jiao, K. Liang, B. Feng, Y. Wang, S. Wu, and Q. Zhang, "Joint channel estimation and decoding for polar coded SCMA system over fading channels," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 210–221, March 2021.
- [23] L. Xiang, Y. Liu, C. Xu, R. G. Maunder, L.-L. Yang, and L. Hanzo, "Iterative receiver design for polar-coded SCMA systems," *IEEE Transactions on Communications*, pp. 1–1, 2021.
- [24] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Multiplexing and channel coding (Release 15), vol. 3GPP TS 38.212 V1.2.1, Dec. 2017.
- [25] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [26] P. Popovski, . Stefanovi, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (urllc)," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, Aug 2019.
- [27] W. Park and A. Barg, "Polar codes for  $q$ -ary channels,  $q = 2^r$ ," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 955–969, 2013.
- [28] P. Chen, B. Bai, and X. Ma, "Two-stage polarization-based nonbinary polar codes for 5G URLLC," *arXiv*, Feb. 2019.
- [29] N. Cheng, R. Zhang, Y. Ge, W. Shi, Q. Zhang, and X. S. Shen, "Encoder and list decoder of Reed-Solomon kernel based polar codes," in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, 2016, pp. 1–6.
- [30] P. Yuan and F. Steiner, "Construction and decoding algorithms for polar codes based on  $2 \times 2$  non-binary kernels," in *2018 IEEE 10th International Symposium on Turbo Codes Iterative Information Processing (ISTC)*, 2018, pp. 1–5.
- [31] M. Falk, G. Bauch, and I. Nissen, "Analysis of non-binary polar codes over  $gf(3)$  and  $gf(5)$  with phase shift keying for short messages," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, Nov 2020, pp. 1–5.
- [32] M. Gao, W. Ge, P. Zhang, and Y. Zhang, "An efficient codebook design for uplink SCMA," *IEEE Access*, vol. 8, pp. 211 665–211 675, 2020.
- [33] R. Mori and T. Tanaka, "Channel polarization on  $q$ -ary discrete memoryless channels by arbitrary kernels," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 894–898.
- [34] W. B. Ameur, P. Mary, M. Dumay, J.-F. Huard, and J. Schwoerer, "Performance study of MPA, log-MPA and max-log-MPA for an uplink SCMA scenario," in *2019 26th International Conference on Telecommunications (ICT)*, April 2019, pp. 411–416.
- [35] B. Feng, R. Liu, and H. Sun, "Simplified successive-cancellation list decoding of non-binary polar codes with rate-1 node," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, May 2020, pp. 1–6.
- [36] J. Zhang, Y. He, Y.-W. Li, C.-K. Wen, and S. Jin, "Meta learning-based MIMO detectors: Design, simulation, and experimental test," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1122–1137, Feb 2021.
- [37] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sep. 2014, pp. 1–5.
- [38] H. Zhang, J. Zhu, H. Shi, and D. Wang, "Layered approx-regular LDPC: Code construction and encoder/decoder design," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 2, pp. 572–585, March 2008.
- [39] D. Declercq and M. Fossorier, "Decoding algorithms for nonbinary LDPC codes over  $GF(q)$ ," *IEEE Transactions on Communications*, vol. 55, no. 4, pp. 633–643, April 2007.

- [40] S. A. Hashemi, C. Condo, and W. J. Gross, “A fast polar code list decoder architecture based on sphere decoding,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 12, pp. 2368–2380, Dec 2016.