A Non-Orthogonal Multiple-Access Scheme Using Reliable Physical-Layer Network Coding and Cascade-Computation Decoding

Tao Yang, *Member, IEEE*, Lei Yang, *Student Member, IEEE*, Y. Jay Guo, *Fellow, IEEE*, and Jinhong Yuan, *Fellow, IEEE*

***Abstract*— This paper studies non-orthogonal transmission over a *K-user* fading multiple access channel. We propose a new reliable physical-layer network coding and cascade-computation decoding scheme. In the proposed scheme, *K* single-antenna users encode their messages by the same practical channel code and QAM modulation, and transmit simultaneously. The receiver chooses *K* linear coefficient vectors and *computes* the associated *K* layers of finite-field linear message combinations in a *cascade* manner. Finally, the *K* users' messages are *recovered* by solving the *K* linear equations. The proposed can be regarded as a generalized onion peeling. We study the optimal network coding coefficient vectors used in the cascade computation. Numerical results show the performance of the proposed approaches that of the iterative maximum *a posteriori* probability detection and decoding scheme, but without using receiver iteration. This results in considerable complexity reduction, processing delay, and easier implementation. Our proposed scheme significantly outperforms the iterative detection and decoding scheme with a single iteration, for example, by 1.7 dB for the two user case. The proposed scheme provides a competitive solution for non-orthogonal multiple access.**

***Index Terms*— Multiuser detection, MIMO, physical-layer network coding, compute-and-forward, iterative decoding**

I. Introduction

NON-ORTHOGONAL multiple-access (NOMA) is poised to become one of the key advanced technologies in 5G cellular systems. Compared to other existing orthogonal multiple-access, NOMA can offer higher throughput, higher capacity and better energy efficiency. However, the inherit nature of the non-orthogonality poses challenges to the transceiver design, owing to the interference among users.

From an information theoretic perspective, non-orthogonal transmission is required to achieve the full capacity region of the multiple-access channel (MAC) [1], [2]. This motivates intensive research efforts on developing practical coding and signal processing techniques for MAC in the past two decades. In particular, the notion of turbo principle was used to solve the multi-user decoding problem in NOMA. Significantly improved performance for the MAC has been demonstrated for various iterative detection and decoding (IDD) schemes. Examples include the seminal paper by Wang and Poor on the iterative soft cancelation for code-division multiple-access (CDMA) [3], the pioneering paper by Hochwald and Ten Brink on IDD for multiple-input multiple-output (MIMO) [4], and the work by Li et al. on interleave-division multiple-access (IDMA) with iterative chip-by-chip detection [5]. These results demonstrated that NOMA with IDD offers significantly improved, or even capacity approaching, performance at relatively affordable decoding computational complexity. The key notion therein is to decouple the signal detection component and the channel-code decoding component by introducing *receiver iteration* and interleavers, where soft information is exchanged between the two components.

To the best of our knowledge, however, the real success of IDD in commercialized systems still remains scarce. By and large, this is because the receiver iteration in the IDD is subject to issues such as high latency, poor stability and difficulties in hardware implementation. This motivates the following question: Would it be possible to achieve the competitive performance of IDD without using receiver iteration?

It is noteworthy that the aforementioned IDD has a common feature, that is, inter-user interference is *suppressed* or *canceled*. In recent years, physical-layer network coding (PNC) or compute-and-forward (CF) revealed that *embracing interference* is beneficial in designing a number of wireless networks, such as the two-way and multi-way relay channels, multiple-access relay channel and distributed MIMO [6], [7], [8]. Not until very recently was the notion of PNC or CF considered for NOMA. In [9], the authors extended CF to Gaussian MAC (GMAC) based on nested lattice code from a theoretical perspective. It was proved that the entire capacity region of two-user GMAC can be attained with a single-user decoder without time-sharing. A network-coded multiple-access (NCMA) scheme was considered for the *two-user* MAC [10]. The PNC de-mapping and multi-user decoding (MUD) were used simultaneously to decode network-coded (NC) message and single-user messages.

Manuscript received February 24, 2016; revised July 29, 2016 and October 12, 2016; accepted December 22, 2016. Date of publication January 10, 2017; date of current version March 8, 2017. This work was supported by the Australian Research Council Discovery Early Career Researcher Award under Grant DE150100636. The associate editor coordinating the review of this paper and approving it for publication was O. O. Koyluoglu.

T. Yang and Y. J. Guo are with the Global Big Data Technologies Centre, University of Technology Sydney, Ultimo, NSW 2007, Australia (email: tao.yang@uts.edu.au; jay.guo@uts.edu.au).

L. Yang and J. Yuan are with the University of New South Wales, Sydney, NSW 2052, Australia (e-mail: lei.yang@unsw.edu.au; j.yuang@unsw.edu.au).

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

Digital Object Identifier 10.1109/TWC.2017.2650900

The above two papers only considered two-user MAC. In addition, the former employed the information theoretic tool of nested lattice code whose decoding complexity is prohibitive in practice; the latter jointly considered physical-layer and medium-access control layer with only binary XOR being used.

This paper constributes to this subject in the following aspects: We study the framework of a general *K-user* NOMA. We borrow the notion of linear PNC[1] in solving the multi-user decoding problem [9], [11]. In particular, we propose a new reliable (channel-coded) linear PNC scheme with *cascade-computation and decoding*. In the proposed scheme, K single-antenna users encode their messages by using a *same* irregular-repeat accumulate (IRA) code over GF(q) and QAM, and transmit simultaneously. The receiver takes in a noisy superposition of the K users' signals. Based on that, the receiver chooses K linear coefficient vectors, of full rank K, and *computes* the associated K layers of linear message-combinations in a *cascade* manner. Finally, the K users' messages are *decoded* by solving the K linear equations.

It is well-known that in a linear PNC or CF based scheme, the choice of the network coding coefficient vectors is crucial to the system performance. In this paper we will also study how to find the network coding coefficient vectors used in the cascade-computation and decoding that leads to the optimized performance. We will also show by numerical result that the proposed scheme achieves almost the same performance as that of the IDD, or even the single-user bound, without using receiver iteration. This results in considerable complexity reduction and easier implementation relative to IDD. The proposed scheme significantly outperforms the IDD scheme with a single iteration by as much as 1.7 dB at a similar complexity level. Therefore, our proposed scheme provides a competitive solution for the decoding problem of NOMA.

We note that our proposed new method is different from the successive-cancelation computation approach in [9] and [12], which performs signal cancelation "physically" from the received signal. Instead, our proposed cascade-computation and decoding focuses on optimized grouping and reduction in the signal constellation to enhance the decoding performance. We note that our proposed scheme is in contrast to the existing IDD schemes which employ different channel codes for various users by assigning different interleavers. Our scheme does not require receiver iteration and interleavers are not used. This paper is different from [11] where channel coding and the optimization of the network coding matrix were not considered.

II. System Model

We consider a MAC as shown in Fig. 1, where K single-antenna users communicate to a common receiver equipped with N antennas. We consider block fading channel, where the channel coefficients remain unchanged in one code block and vary independently over blocks. We assume that the channel state information (CSI) is perfectly known by the receiver but not known by the users. Here we present a real-valued size K by N system model, which can be straightforwardly extended to a complex-valued model [13].

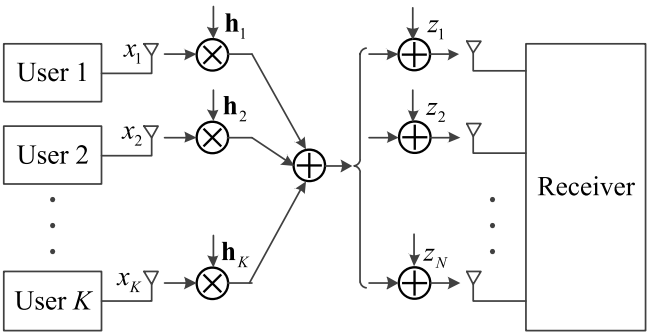


Fig. 1. Block diagram of the considered MAC with K single-antenna users and a N receive-antenna receiver.

Let $\mathbf{u}_k = [u_k[1], \cdots, u_k[m]] \in (GF(q))^m$ be a length-m *message* of user k, $k \in \{1, \cdots, K\}$. User k maps its message to a length-n *coded and modulated sequence* $\mathbf{x}_k = [x_k[1], \cdots, x_k[n]]$. The per-user rate is $\frac{m}{n}\log_2 q$ bits/channel-use. For simplicity, we consider that the average symbol energy of the entries of $\mathbf{x}_k$ are normalized, i.e.,

$$E\left(|x_k[t]|^2\right) = 1, \quad t = 1, \cdots, n, \quad k \in \{1, \cdots, K\}.$$

In a NOMA scheme, the K users' modulated signals are transmitted simultaneously. W.l.o.g, we assume that the K users transmit with equal power, i.e., the energy per symbol is given by E_s for all users. For a given channel realization, at time instant t, the received signal is given by

$$\mathbf{y}[t] = \sum_{k=1}^{K} \mathbf{h}_k \sqrt{E_s} x_k[t] + \mathbf{z}[t], \quad t = 1, \cdots, n, \qquad (1)$$

where $\mathbf{h}_k$ denotes the N-by-1 channel coefficient vector from user k to the N antennas of the receiver and $\mathbf{z}[t]$ is the additive white Gaussian noise (AWGN) vector. Note that the entries of $\mathbf{h}_k$ take values in the field of real numbers, i.e., $\mathbf{h}_k \in \mathcal{R}^N$. The entries of $\mathbf{z}$ are assumed to be i.i.d. and have zero mean and variance σ^2. The SNR is defined as $\rho \triangleq \frac{E_s}{\sigma^2}$. In the above, symbol-synchronization is assumed. We note that synchronization of the user nodes can be implemented at the base station (BS) even if the user nodes do not have transmitter-side CSI (CSI-T) [2], particularly in a relatively static environment. The task of user synchronization is primarily to compensate the different distances between the user nodes to the BS, while obtaining the CSI is related to combating the small-scale multi-path fading. There is no controversy between the assumption of no CSI-T and user synchronization. For the scenario with high mobility users where perfect user synchronization becomes extremely challenging, some recent results on asynchronous physical-layer network coding can be used on top of our propose scheme. For example, [14] suggested to use cyclic codes and guarding interval, which ensures that the addition of two codewords asynchronously is also a codeword after modulo-q operation. Another way of dealing with symbol asynchrony is using OFDM as in [15] and [16]. Specifically,

[1]Non-linear PNC will not be considered in this paper, for there lacks a explicit way for its optimization and channel coding.

the delay in symbol arrival time is translated into a distortion in frequency domain which can be addressed by implementing frequency domain equalization in the detector. The details are beyond the scope of this paper.

A. Problem at a Glance

The receiver wants to recover all users' messages $\mathbf{u}_1, \cdots, \mathbf{u}_K$. Denoted the decisions on the messages by $\widehat{\mathbf{u}}_1, \cdots, \widehat{\mathbf{u}}_K$. A decoding error is declared if $\widehat{\mathbf{u}}_k \neq \mathbf{u}_k$ for any $k \in \{1, \cdots, K\}$. Generally speaking, a NOMA scheme that yields "good" error probability performance with low computation complexity, as well as low implementation cost, is desirable. This is the objective of this paper.

B. Existing Schemes in the Literature

In fact, the multi-user decoding problem has been extensively investigated in the literature. Since the 90's, there has been great efforts devoted to solve the multi-user decoding problem in randomly coded CDMA systems. The difficulty lies in the huge dimension in performing the optimal *joint maximum likelihood sequence decoding*, which quickly becomes prohibitive even for a moderate number of users and a moderate block length. The most famous and promising approach is the IDD approach. The concept of this approach is similar to that of turbo codes, in which the detection and decoding operations are decoupled. Soft information (extrinsic information or a posteriori probability) is exchanged between the symbol-by-symbol detection component and a bank of channel code decoding components in an iterative fashion. Therein, interleavers and deinterleavers are employed to minimize the correlation among the soft information.

In the seminal work by Wang and Poor, an iterative soft interference cancelation and decoding is presented for randomly coded asynchronous CDMA [3]. The receiver performs two successive soft output decisions, achieved by a soft-input soft-output (SISO) multiuser detector and a bank of single-user SISO convolutional code decoders, through an iterative process. The SISO multiuser detector performs soft cancelation using the soft information from the decoders, and linear minimum mean square error (MMSE) filter is used to suppress the remaining interference. Li et al. proposed a chip-level interleaved multi-user scheme called interleave division multiple-access (IDMA) [5]. Thanks to the chip-level interleaver, the correlation over chip signals becomes negligible, which enables a high-performance chip-by-chip iterative elementary signal estimation detector. Such a detector does not require the MMSE filter. The IDD for MIMO was pioneered by Hochwald and Ten Brink, where a list sphere detector and a bank of turbo code decoders exchange extrinsic information iteratively [4]. Near capacity performance of MIMO channel is demonstrated therein. Later, the convergence behavior analysis using extrinsic information transfer (EXIT) chart technique was considered for MIMO, and the curve fitting between the EXIT of the detector and LDPC code decoders leads to further reduced gap to the capacity limit [17], [18]. A few groups, including us, have worked on combining the soft information between adjacent iterations to improve the performance or reduce the complexity [19], [20].

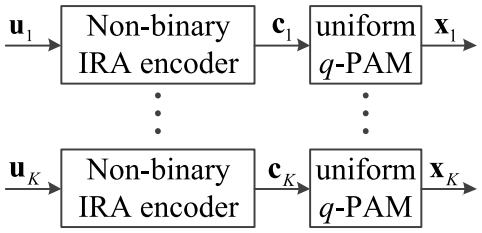


Fig. 2. Block diagram of the encoding of the proposed scheme. The same channel-code encoder is used for all users.

Despite the promising performance of IDD, it is yet to see a successful implementation of IDD in a well-established system from a commercial point of view. The primary reason would be that the receiver iteration may result in long processing delays, poor stability as well as burdens to hardware implementation. Therefore, it is desirable to find a practical scheme that yields the promising performance without using receiver iteration.

III. The Proposed Scheme

A. Encoding

Fig. 2 shows the block diagram of the encoding process of K user. User k maps its message sequence $\mathbf{u}_k$ to a length-n *channel-coded sequence* $\mathbf{c}_k = [c_k[1], \cdots, c_k[n]] \in (GF(q))^n$ by using a linear code over GF(q). This encoding process is written as

$$\mathbf{c}_k = \mathbf{u}_k \otimes \mathbf{G}_C, k \in \{1, \cdots, K\}. \tag{2}$$

Here $\mathbf{G}_C$ denotes the m-by-n *channel coding generator matrix* whose entries take values from GF(q).

We note that the same channel coding generator matrix $\mathbf{G}_C$ is used for all users, following the spirit of PNC or CF [12], and no interleaver is needed. This is in contrast to the existing IDD schemes which employ different codes for users by assigning different interleavers.

In this work, the underlying channel code under consideration will be a practical *random-coset IRA code* over GF(q) [21], [22]. The encoding process is briefly described below. The message symbols of every user are repeated according to a certain repeat-node degree distribution. The repeated messages sequence undergoes an interleaving process, and the output is forwarded to a bank of check-accumulator nodes of a certain degree distribution. Next, a random-coset vector is added to the coded sequence, yielding the coded sequence $\mathbf{c}_k$, $k \in \{1, \cdots, K\}$. In such a way, the encoding process in Eq. (2) is completed. The full detail of the encoding algorithm with the random-coset IRA code, as well as the optimization of the node degree distribution using a generalized EXIT curve-fitting technique, can be found in [22].

By using the uniform q-PAM modulation, the resultant *modulated signals* are written as

$$\mathbf{x}_k = \frac{1}{\gamma}\left(\mathbf{c}_k - \frac{q-1}{2}\right), \quad k \in \{1, \cdots, K\}, \tag{3}$$

where γ is a power normalization factor to ensure that $\frac{1}{n}E\left(\|\mathbf{x}_k\|^2\right)=1$. Here, as in [12] and [23], we assume that q is a prime number.

B. Reliable Linear PNC

1) Preliminaries: Recall the received signal in Eq. (1), which is a noisy superposition of the K users' modulation-coded signal sequences. For convenience, denote $\mathbf{X}=\left[\mathbf{x}_1^T,\cdots,\mathbf{x}_K^T\right]^T$. That is,

$$\mathbf{X}=\begin{bmatrix} x_1[1] & x_1[2] & \cdots & x_1[n] \\ x_2[1] & \ddots & & x_2[n] \\ \vdots & \cdots & \ddots & \vdots \\ x_K[1] & x_K[2] & \cdots & x_K[n] \end{bmatrix}. \tag{4}$$

Further, denote $\mathbf{H}=[\mathbf{h}_1,\cdots,\mathbf{h}_K]$, $\mathbf{Y}=\left[\mathbf{y}[1],\cdots,\mathbf{y}[n]\right]$ and $\mathbf{Z}=[\mathbf{z}[1],\cdots,\mathbf{z}[n]]$. Then, (1) can be written as

$$\mathbf{Y}=\mathbf{H}\sqrt{E_s}\mathbf{X}+\mathbf{Z}. \tag{5}$$

Denote $\mathbf{U}=\left[\mathbf{u}_1^T,\cdots,\mathbf{u}_K^T\right]^T$ the matrix format of all K users' message sequences, a linear combination of the K users' messages can be written as

$$\mathbf{w}=\mathbf{g}\otimes\mathbf{U} \tag{6}$$

where $\mathbf{g}$ is a size 1 by K row vector with entries taken from GF(q), and "$\otimes$" represents the modulo-q matrix multiplication. We refer to $\mathbf{g}$ as a *network coding (NC) coefficient vector*, and $\mathbf{w}$ as a *message-combination*, or a *NC message sequence* [24]. For the prime q under consideration, the finite integer set $\{0,...,q-1\}$ forms a finite field under modular addition and multiplication.

In a reliable linear PNC scheme, the receiver reconstructs the NC message directly from its received signal, without complete decoding of all users' messages [23], [25], [26]. This is in contrast to non-PNC schemes where complete decoding of all users' messages is performed, followed by (manually) combining the decoded messages.

The receiver can choose to compute multiple NC message sequences. In the NOMA scenario under consideration, the receiver aims to compute K NC message sequences. Let us denote them by

$$\mathbf{w}_l=\mathbf{g}_l\otimes\mathbf{U},\quad l=1,\cdots,K. \tag{7}$$

where $\mathbf{g}_1,\mathbf{g}_2,\cdots,\mathbf{g}_K$ are K NC coefficient vectors, each associated with a NC message sequence.

Let $\mathbf{W}=\left[\mathbf{w}_1^T,\cdots,\mathbf{w}_K^T\right]^T$ and $\mathbf{G}_N=\left[\mathbf{g}_1^T,\cdots,\mathbf{g}_K^T\right]^T$. Then, (7) can be written as

$$\mathbf{W}=\mathbf{G}_N\otimes\mathbf{U}.$$

We refer to $\mathbf{G}_N$ as an *NC generator matrix*. The subscript "N" is used to distinguish the NC generator matrix from the channel-code generator matrix $\mathbf{G}_C$ in Eq. (2). It is obvious that $\mathbf{G}_N$ must have full rank K in GF(q). This ensures the K users' messages can be recovered by

$$\mathbf{U}=\mathbf{G}_N^{-1}\otimes\mathbf{W}. \tag{8}$$

Let $\widehat{\mathbf{W}}$ be the decision on the NC messages computed by the receiver. Given $\mathbf{Y}$, the optimal joint maximum a posteriori probability (MAP) rule for computing the K NC message sequences is

$$\begin{aligned}\widehat{\mathbf{W}}&=\left[\widehat{\mathbf{w}}_1^T,\cdots,\widehat{\mathbf{w}}_K^T\right]^T\\&=\arg\max_{\mathbf{w}_l\in\{0,\cdots,q-1\}^m,l=1,\cdots,K}p\left(\mathbf{w}_1,\cdots,\mathbf{w}_K|\mathbf{Y},\mathbf{G}_N\right).\end{aligned} \tag{9}$$

Unfortunately, this task of finding the solution to (9) using joint MAP rule is very challenging. First, there lacks an approach that can apply the believe propagation decoding for the IRA code to obtain the solution. Second, even if such an approach is developed, the complexity will be very high due the multi-levels of the NC messages. To overcome the difficulties, we propose a new cascade-compution and decoding approach in the following.

Remark 1: For a given NC generator matrix $\mathbf{G}_N$, there is an exact one-to-one mapping between $\mathbf{U}$ and $\mathbf{W}$. That is, $\widehat{\mathbf{U}}=\mathbf{U}$ if $\widehat{\mathbf{W}}=\mathbf{W}$, and $\widehat{\mathbf{U}}\neq\mathbf{U}$ if $\widehat{\mathbf{W}}\neq\mathbf{W}$. In other words, an error in computing the NC messages will definitely result in an error in the final decoding of the K users' messages. In the proposed multiuser decoding scheme, the problem of minimizing $\Pr(\widehat{\mathbf{U}}\neq\mathbf{U})$ now becomes that of minimizing $\Pr(\widehat{\mathbf{W}}\neq\mathbf{W})$.

Remark 2: We will see later that $\Pr(\widehat{\mathbf{W}}\neq\mathbf{W})$ is dependent on the choice of the NC generator matrix $\mathbf{G}_N$. In this section, we exclusively consider that $\mathbf{G}_N$ is given. Section IV will study how to choose $\mathbf{G}_N$ that yields the best performance for the proposed scheme.

2) An Existing Parallel-Computation: Before presenting the new cascade-compution and decoding scheme, we quickly sketch a parallel-computation approach [12]. Consider one block of transmission. In the first step, the receiver selects a NC generator matrix $\mathbf{G}_N$ based on the knowledge of channel state information $\mathbf{H}$. Next, the receiver computes the associated NC messages sequences $\mathbf{w}_1,\cdots,\mathbf{w}_K$ in parallel. That is, the computation operation of $\mathbf{w}_i$ is independent of that of $\mathbf{w}_j$, for $j\neq i$. The MAP rule associated with such a parallel-computation is written as

$$\widehat{\mathbf{w}}_l=\arg\max_{\mathbf{w}_l\in\{0,\cdots,q-1\}^m}p\left(\mathbf{w}_l|\mathbf{Y},\mathbf{g}_l\right),\quad l=1,\cdots,K. \tag{10}$$

After finishing the parallel computation, the receiver obtains $\widehat{\mathbf{W}}$. It then recovers K users' messages by $\widehat{\mathbf{U}}=\mathbf{G}_N^{-1}\otimes\widehat{\mathbf{W}}$.

C. A New Cascade-Computation and Decoding Scheme

We now propose a new cascade-computation and decoding (CCD) for computing the NC message sequences $\mathbf{W}=\left[\mathbf{w}_1^T,\cdots,\mathbf{w}_K^T\right]^T$. The block diagram is depicted in Fig. 3. Here we focus on presenting the CCD for a given NC generator matrix $\mathbf{G}_N$.

Layer 1: The receiver first computes the first NC message $\mathbf{w}_1=\mathbf{g}_1\otimes\mathbf{U}$ entirely based on the received signal $\mathbf{Y}$. The decision made is denoted by $\widehat{\mathbf{w}}_1$, calculated by the MAP rule:

$$\widehat{\mathbf{w}}_1=\arg\max_{\mathbf{w}_1\in\{0,\cdots,q-1\}^m}p\left(\mathbf{w}_1|\mathbf{Y}\right). \tag{11}$$

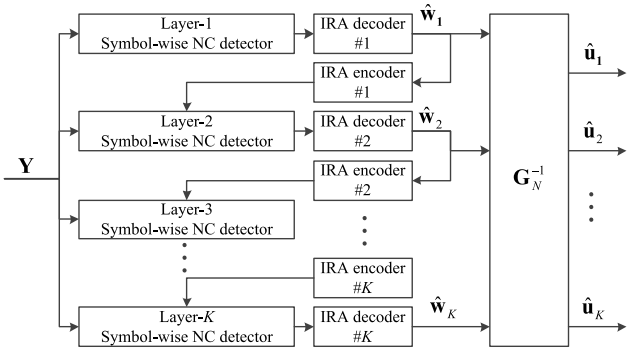


Fig. 3. Block diagram of the proposed CCD scheme for K-user MAC. Note that if the links between adjacent layers are removed, the scheme becomes the parallel-computation and decoding.

For the ease of presentation, we have removed the condition on $\mathbf{G}_N$ in the a posteriori probability above as $\mathbf{G}_N$ has been given.

Layer 2: The receiver computes the second NC message w.r.t. $\mathbf{w}_2 = \mathbf{g}_2 \otimes \mathbf{U}$ not only based on $\mathbf{Y}$, but also take into account $\widehat{\mathbf{w}}_1$. We refer to $\widehat{\mathbf{w}}_1$ as a *side information* which assists the computation of $\mathbf{w}_2$. The decision made is denoted by $\widehat{\mathbf{w}}_2$, calculated by

$$\widehat{\mathbf{w}}_2 = \arg \max_{\mathbf{w}_2 \in \{0,\cdots,q-1\}^m} p\left(\mathbf{w}_2 \,|\mathbf{Y}, \widehat{\mathbf{w}}_1\right). \tag{12}$$

$\vdots$

Layer l: The receiver computes the lth NC message based on $(\mathbf{Y}, \widehat{\mathbf{w}}_1, \cdots, \widehat{\mathbf{w}}_{l-1})$, given by

$$\widehat{\mathbf{w}}_l = \arg \max_{\mathbf{w}_l \in \{0,\cdots,q-1\}^m} p\left(\mathbf{w}_l \,|\mathbf{Y}, \widehat{\mathbf{w}}_1, \cdots, \widehat{\mathbf{w}}_{l-1}\right). \tag{13}$$

$\vdots$

Layer K: The receiver computes the last NC message given by

$$\widehat{\mathbf{w}}_K = \arg \max_{\mathbf{w}_K \in \{0,\cdots,q-1\}^m} p\left(\mathbf{w}_l \,|\mathbf{Y}, \widehat{\mathbf{w}}_1, \cdots, \widehat{\mathbf{w}}_{K-1}\right). \tag{14}$$

Final decoding: the receiver decodes the K users' messages by $\widehat{\mathbf{U}} = \mathbf{G}_N^{-1} \otimes \widehat{\mathbf{W}}$.

Here we see that the K NC messages are computed layer after layer in a cascaded manner, hence the name *cascade-computation*. In the proposed scheme, the receiver exploits the notion of linear PNC or CF for computing $\widehat{\mathbf{w}}_1$ in the first layer. In the subsequent layers, the receiver exploits *information-combining* on top of PNC, using the side-information of the computed NC messages from previous layers. Our proposed CCD advances the conventional PNC for the TWRC.

Remark 3: In the proposed scheme, only K (single-user) channel code decoding operations are required in total. The order of complexity of channel code decoding is identical to that of the IDD scheme with only one iteration. Thus, the proposed scheme may have far less decoding complexity compared to the IDD scheme with many iterations. Note also that the proposed scheme does not require interleavers and deinterleavers, whereas the IDD scheme requires K interleaving and deinterleaving operations per iteration. We will provide detailed complexity evaluation in Section V.

1) Why Cascade-Computation Helps?: In the parallel computation, the K NC messages are computed independently. From an information theoretic point of view, the overall performance is dictated by the specific NC message that has the smallest computation rate [12]. In particular, if the NC coefficient vectors are arranged in the descending order with their computation rates, the performance of the parallel computation is exactly determined by the computation rate w.r.t. the last NC message.

In fact, it is possible to do better than using this independent computation operation, by taking into account the relation between the NC messages. The best way to show this would be to take a closer look at the signal constellation. Suppose that a q-PAM mapping is used by each user. The superimposed signals seen by the receiver has a cardinality of q^K for $\mathbf{h}_1, \cdots, \mathbf{h}_K \in \mathcal{R}^N$. Let us consider that the first NC messages is correctly computed, i.e., $\widehat{\mathbf{w}}_1 = \mathbf{w}_1$. In Layer 2, taking $\widehat{\mathbf{w}}_1$ as a side information in computing $\mathbf{w}_2$, the constellation points

$$\mathbf{U} : \mathbf{g}_1 \otimes \mathbf{U} \neq \widehat{\mathbf{w}}_1$$

that do not satisfy the side information constraint becomes irrelevant, and are thus expurgated from the constellation. Then, the remaining effective constellation has a cardinality of only q^{K-1}. This will make the computation of $\mathbf{w}_2$ be subjected to less constellation points compared to that in the parallel computation, where the cardinality of the constellation is still q^K.

Likewise, the relevant constellation has a cardinality of only q^{K-l+1} in computing $\mathbf{w}_l$ in Layer l. This is far less than that in the parallel computation where the cardinality of the constellation is q^K. It is apparent that the CCD deals with a much sparser constellation as compared to that of the parallel computation. For a channel coded system, this will lead to a improved distance spectrum, i.e., with larger minimum distance as well as reduced multiplicities, yielding a lower error probability or a higher achievable rate.

Let us look at a simple example to show that by considering the relation between NC messages, the computation error probability of NC messages will be decreased.

Example 1: Consider a MAC with two users ($K = 2$) and a two receive-antenna receiver ($N = 2$). The channel coefficients are $\mathbf{H} = \left[[0.23, 0.98]^T, [1.15, 0.99]^T\right]$. The receiver intends to compute two NC messages $\mathbf{w}_1$ and $\mathbf{w}_2$ w.r.t NC coefficient vectors $\mathbf{g}_1 = [1, 1]^T$ and $\mathbf{g}_2 = [1, 0]^T$, respectively. For illustration purposed, consider an un-channel-coded system here.

Fig. 4 depicts the constellation at the receiver for the given channel coefficients $\mathbf{H}$ and $q = 3$. The horizontal and vertical axes stand for the signal received by the first and second antennas, respectively, as specified in (1). For illustration purpose, the additive noise in not included. There are $q^K = 9$ constellation points in total. In Fig. 4 (a) and Fig. 4 (b), we partition all constellation points into $q = 3$ sets w.r.t the NC coefficient vector $\mathbf{g}_1$ and $\mathbf{g}_2$, respectively. In each sub-figure, the constellation points with different underlying

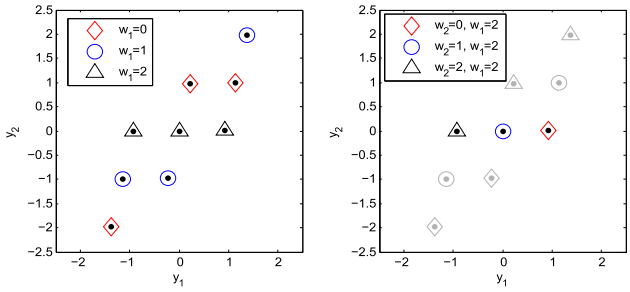


Fig. 4. Illustration of the constellation for computing the first NC message $\mathbf{v}_1$ in the left subfigure, and the second NC message $\mathbf{v}_2$ in the right subfigure, where $K = N = 2$. In this example, $\mathbf{H} = \left[[0.23, 0.98]^T, [1.15, 0.99]^T\right]$, $\mathbf{g}_1 = [1, 1]^T$, $\mathbf{g}_2 = [1, 0]^T$ and $q = 3$. Here, y_1 and y_2 represent the signals received on two antennas of the receiver.

NC messages are labeled with different icons. In the first layer, where Fig. 4 (a) is relevant, the receiver decides $\mathbf{w}_1$ by finding the most likely set. Assume that $\widehat{\mathbf{w}}_1 = 2$. In the second layer, the receiver decides $\mathbf{w}_2$ by taking into account the side information of $\widehat{\mathbf{w}}_1$. Specifically, those constellation points whose underlying NC message $\mathbf{w}_1$ is not equal to 2 are irrelevant, and are thus expurgated. These expurgated points are grayed in Fig. 4 (b). It is clear that the cardinality of the effective constellation for computing $\mathbf{w}_2$ is *reduced* from $q^K = 9$ to $q^{K-1} = 3$. This will make the computation of $\mathbf{w}_2$ "easier" compared to that in the parallel computation, where the cardinality of the constellation is still $q^K = 9$. We note that for a channel-coded system, an improved distance spectrum can be obtained using the cascade-computation over parallel computation, leading to higher mutual information or lower decoding error probability.

2) Difference to the Successive-Cancelation Computation: We note that our CCD scheme is different from the successive-cancelation computation method in [9] and [12]. In existing successive cancellation based schemes, the strongest user's signal is first decoded and then physically cancelled from the received signal. The resulting signal is used for decoding of the subsequent layer of user's signal. In contrast, there is no signal cancellation in our proposed scheme. To be more precise, recall that the successive cancellation based scheme is performed in a user-by-user manner. Once a user's signal is decoded, that user's transmitted signal is known and can thus be cancelled from the received signal. Yet, in our proposed CCD scheme, the receiver's operation is performed in a NC message by NC message manner. Once a NC message is known, the transmitted signal is not necessarily known, and thus signal cancellation is not possible. In our proposed scheme, the NC messages computed in previous layers are used as a priori information, which expurgate the possibilities of irrelevant constellations, in the computing the NC message of the next layer in the MAP algorithm. We also note that successive-cancelation computation based on nested lattice codes is an information theoretic notion, while our CCD is a practical scheme whose error probability performance will be numerically shown in a later section.

D. Cascade-Computation Algorithm in Detail

Here we present the detailed algorithm used in CCD where practical detector and channel-code are considered. As the computation and decoding will be based on the channel-coded vectors $\mathbf{c}_1, ..., \mathbf{c}_K$, the following notations will be used.

Consider the channel-coded codewords. Denote the codebook set w.r.t. the channel code by

$$\mathcal{C} = \left\{\mathbf{c} : \mathbf{c} = \mathbf{u} \otimes \mathbf{G}_C, \quad \forall \mathbf{u} \in \{0, \cdots, q-1\}^m\right\}. \tag{15}$$

Let

$$\mathbf{C} = \left[\mathbf{c}_1^T, \cdots, \mathbf{c}_K^T\right]^T \tag{16}$$

denote the channel-coded codewords of K users.

Let the l-th linear combination of the codewords w.r.t. the NC coefficient vector $\mathbf{g}_l$ be

$$\mathbf{v}_l = \mathbf{g}_l \otimes \mathbf{C}. \tag{17}$$

We refer to $\mathbf{v}_l$ as a *NC codeword*. Note that there is a one-to-one mapping between the NC codeword $\mathbf{v}_l$ and the NC message sequence $\mathbf{w}_l$. This can be seen as follows.

Recall that the same channel code generator matrix $\mathbf{G}_C$ is used for all users, we have

$$\mathbf{C} = \mathbf{U} \otimes \mathbf{G}_C$$

and thus

$$\begin{aligned} \mathbf{v}_l &= \mathbf{g}_l \otimes \mathbf{U} \otimes \mathbf{G}_C \\ &= \mathbf{w}_l \otimes \mathbf{G}_C. \end{aligned} \tag{18}$$

Recall the MAP rule for the computation in Eq. (13), repeated below

$$\widehat{\mathbf{w}}_l = \arg \max_{\mathbf{w}_l \in \{0,\cdots,q-1\}^m} p\left(\mathbf{w}_l \,|\mathbf{Y}, \widehat{\mathbf{w}}_1, \cdots, \widehat{\mathbf{w}}_{l-1}\right), l = 1, \cdots, K.$$

Owing to the one-to-one mapping between $\mathbf{v}_l$ and $\mathbf{w}_l$, finding the solution to (13) is equivalent to finding

$$\widehat{\mathbf{v}}_l = \arg \max_{\mathbf{v}_l \in \mathcal{C}} p\left(\mathbf{v}_l | \mathbf{Y}, \widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_{l-1}\right). \tag{19}$$

A brute-force search of the solution is subject to a complexity of order at least $O\left(q^m\right)$. We propose to employ a practical soft-input soft-output *symbol-wise* detector followed by an iterative believe propagation (BP) decoder to approximate the MAP solution. The algorithm is presented below.

1) Symbol-Wise NC Detector for $l = 1$: Based on $\mathbf{Y}$, the receiver calculates the symbol-wise a posteriori probabilities (APPs). Let $\mathbf{c}[t]$ be a column vector that collects all K users' coded digits at time instant t, i.e., $\mathbf{c}[t]$ being the tth column of $\mathbf{C}$. Denote by $p_i^{(l)}[t]$ the APP w.r.t. $v_l[t] = i$, $i \in \{0, \cdots, q-1\}, t = 1, \cdots, n$.

For the first layer of NC codeword, i.e., $l = 1$, the APP is calculated as

$$\begin{aligned} p_i^{(1)}[t] &= p\left(v_1[t] = i | \mathbf{y}[t]\right) = p\left(\mathbf{g}_1 \otimes \mathbf{c}[t] = i | \mathbf{y}[t]\right) \\ &= \frac{p\left(\mathbf{y}[t] | \mathbf{g}_1 \otimes \mathbf{c}[t] = i\right) p\left(\mathbf{g}_1 \otimes \mathbf{c}[t] = i\right)}{p\left(\mathbf{y}[t]\right)} \\ &= \frac{1}{\eta} \sum_{\substack{\mathbf{c}[t] \in \{0,\cdots,q-1\}^K \\ \mathbf{g}_1 \otimes \mathbf{c}[t] = i}} \exp\left(\frac{\left\|\mathbf{y}[t] - \mathbf{H}\left(\mathbf{c}[t] - \frac{q-1}{2}\right)\frac{\sqrt{E_S}}{\gamma}\right\|^2}{2\sigma^2}\right), \end{aligned} \tag{20}$$

where η is a normalization factor to ensure $\sum_{i=0}^{q-1} p_i^{(1)}[t] = 1$. In the derivation above, we have utilized the fact that the NC codeword symbols $\mathbf{v}_1[t] = \mathbf{g}_1 \otimes \mathbf{c}[t]$ are uniformly distributed and that the conditional probability of the received signal vector follows the Gaussian distribution.

Denote by $\mathbf{p}^{(1)}[t] = [p_0^{(1)}[t], \cdots, p_{q-1}^{(1)}[t]]$. The calculated APPs $\mathbf{p}^{(1)}[t], l = 1, \cdots, K$, are fed to an iterative BP decoder to decode the NC codeword. We will provide details on the iterative BP decoder momentarily. The iterative BP decoder yields the (hard) decision on the NC codeword $\widehat{\mathbf{v}}_1$.

2) Symbol-Wise NC Detector for $l > 1$: For layer $l = 2$, the symbol-wise APP is calculated as

$$p_i^{(2)}[t] = p\left(\mathbf{g}_2 \otimes \mathbf{c}[t] = i | \mathbf{y}[t], \mathbf{g}_1 \otimes \mathbf{c}[t] = \widehat{\mathbf{v}}_1\right) = \frac{1}{\eta} \sum_{\Omega^{(2)}[t]} \exp\left(-\frac{\left\|\mathbf{y}[t] - \mathbf{H}\left(\mathbf{c}[t] - \frac{q-1}{2}\right)\frac{\sqrt{E_S}}{\gamma}\right\|^2}{2\sigma^2}\right), \tag{21}$$

where $\Omega^{(2)}[t] = \{\mathbf{c}[t] : \mathbf{g}_2 \otimes \mathbf{c}[t] = i, \mathbf{g}_1 \otimes \mathbf{c}[t] = \widehat{\mathbf{v}}_1\}$ and $\mathbf{c}[t] \in \{0, ..., q-1\}^K$. Here, the set $\Omega^{(2)}[t]$ yields all possible candidates of $\mathbf{c}[t]$ that satisfy $\mathbf{g}_2 \otimes \mathbf{c}[t] = i$ and the side information constraint $\mathbf{g}_1 \otimes \mathbf{c}[t] = \widehat{\mathbf{v}}_1$ given by layer 1. The calculated APPs $\mathbf{p}^{(2)}[t], t = 1, \cdots, n$, are fed to the BP decoder to decode the NC codeword, yielding the hard decision on the NC codeword $\widehat{\mathbf{v}}_2$.

In general for layer l, the symbol-wise APP is

$$p_i^{(l)}[t] = p\left(\mathbf{g}_l \otimes \mathbf{c}[t] = i | \mathbf{y}[t], \mathbf{g}_{l'} \otimes \mathbf{c}[t] = \widehat{\mathbf{v}}_{l'} \forall l' = 1, \cdots, l-1\right) = \frac{1}{\eta} \sum_{\Omega^{(l)}[t]} \exp\left(-\frac{\left\|\mathbf{y}[t] - \mathbf{H}\left(\mathbf{c}[t] - \frac{q-1}{2}\right)\frac{\sqrt{E_S}}{\gamma}\right\|^2}{2\sigma^2}\right), \tag{22}$$

where

$$\Omega^{(l)}[t] = \left\{\mathbf{c}[t] : \mathbf{g}_l \otimes \mathbf{c}[t] = i, \mathbf{g}_{l'} \otimes \mathbf{c}[t] = \widehat{\mathbf{v}}_{l'} \forall l' = 1, \cdots, l-1, \right\}.$$

Here, the set $\Omega^{(l)}[t]$ counts all possible candidates that satisfy $\mathbf{g}_l \otimes \mathbf{c}[t] = i$ and the side information constraints $\mathbf{g}_{l'} \otimes \mathbf{c}[t] = \widehat{\mathbf{v}}_{l'}$, $l' = 1, \cdots, l-1$, given by the previous layers. Likewise, the calculated APPs $\mathbf{p}^{(l)}[t], t = 1, \cdots, n$, are fed to the BP decoder to decode the NC codeword, yielding the hard decision on the NC codeword $\widehat{\mathbf{v}}_l$. This layer-by-layer cascade-computation process continues until $\widehat{\mathbf{v}}_K$ is obtained.

3) BP Decoder of IRA Code: Here we briefly describe the iterative BP decoder used in the CCD scheme. The full details of the decoding as well as the optimized design of the random-coset non-binary IRA code can be found in our previous work in [22].

The random-coset non-binary IRA code can be viewed as a serially-concatenated code over GF(q) with repetition code as outer code and check-accumulator as inner code. Interleaver is inserted between the constituent codes according to the structure of the IRA code. Each output symbol of the inner code is added by a random-coset over GF(q) to create a symmetric non-binary input channel, which is useful for the optimization of the non-binary IRA code [22]. At the receiver side, a q-dimension APP vector w.r.t each NC message symbol is computed by (20) (21) or (22) regarding to the computation decoding layer. The APP vectors are fed to a coset remover to obtain the channel information for the non-binary BP decoder. The BP decoder then attempts to recover the NC message sequence in an iterative manner. The iteration starts from the check-accumulator decoder, which takes in the channel information from the coset remover and the a prior information from the repetition code decoder to update the a priori information for the repetition coder decoder. Then the repetition code decoder takes in the updated a priori information from the check-accumulator decoder to compute the a priori information for the check-accumulator decoder. In the first iteration the a prior information from the repetition code decoder is initialized to $[1/q, \cdots, 1/q]_{1\times q}$. Note that the a priori information is exchanged through interleaver/deinterleaver between the constituent decoders. If all the check constrains in the check-accumulator decoder are satisfied, or a predefined maximum number of iterations is reached, the iterative decoding process stops. In such a manner, the hard decision on the NC message sequence $\widehat{\mathbf{w}}_l$ or the NC codeword $\widehat{\mathbf{v}}_l$, $l = 1, \cdots, K$, is made, which is used in the cascade-computation described previously.

In this paper, the underlying channel codes under consideration are the non-binary IRA codes. The ensemble of IRA codes is a subset of LDPC code with simple encoding structure. IRA code can be easily designed using the EXIT chart technique for a capacity-approaching performance [21]. Adopting a capacity-approaching channel code in the proposed scheme will have more theoretical interests. It is noteworthy that if low system delay is of the paramount importance, short channel codes with low decoding complexities, such as Hamming codes and BCH codes, can be straightforwardly adopted in our proposed scheme to reduce the system delay.

It is noteworthy that the work in [9] applies to only two-user scenario and single-antenna case, while our proposed scheme applies to arbitrary number of users and arbitrary number of receiver antennas. Our work considers wireless fading channel and exploit the notion of CF with operation over GF(q), whereas the work in [9] jointly considered physical-layer and medium-access control layer with only binary XOR being used. Note also that the CCD and parallel computation decoding are irrelevant in [9], but are important new notions in our works. Our proposed new channel-code decoding algorithms for the CCD plays an essential role in the new NOMA transceiver structure.

IV. Design of CCD Scheme and Complexity Analysis

A. CCD is a Generalized Onion-Peeling Approach

Our proposed CCD can be thought of as a generalized onion-peeling. To see this, consider that the NC generator matrix is set to

$$\mathbf{G}_N = \mathbf{I}. \tag{23}$$

In this case, the CCD reduces to a conventional onion-peeling scheme: The first user's message is decoded by treating other users' signals as interference. Then, the first user's signal can

be removed and the second user's messages is to be decoded. The operation continues until the last layer of decoding is finished. We note that when the NC generator matrix is set to be an identity matrix, our proposed scheme boils down to a successive interference cancellation (SIC) scheme, where the strongest user's signal is decoded first and peeled off from the received signal. Therefore the performance of SIC can be viewed as a performance lower bound of our proposed scheme. By using the optimized NC generator matrix, improved performance over the SIC is achieved.

Also, consider that the NC generator matrix is set to be a permutation matrix

$$\mathbf{G}_N = \mathbf{P}, \tag{24}$$

which is arranged according to the signal-to-interference plus noise ratio (SINR) of the users. In this case, the CCD reduces to an onion-peeling scheme with user-ordering, typically the strongest user is decoded first while the weakest user is decoded last. Obviously, the scheme by setting $\mathbf{G}_N$ to a permutation matrix according the the SINR of the users is likely to outperform that by setting $\mathbf{G}_N$ to an identity matrix.

Now we can see that our proposed CCD scheme generalizes the aforementioned onion-peeling schemes. From the above two examples, it is apparent that a conventional onion-peeling scheme is strictly a subset (special case) of the CCD scheme. The CCD scheme has the flexibility to peel off not only the users' messages, but also the combinations of the users' messages. Specifically, by optimizing the NC generator matrix $\mathbf{G}_N$, the best way for carrying out this generalized onion-peeling is obtained.

We note that there are fundamental differences between our proposed scheme and the integer-forcing (IF) linear receiver [27]. Our proposed work does not perform "integer forcing" and is not a "linear receiver", although both of our work and [27] exploit the notion of CF (or PNC) in solving the MIMO detection problem. Note also that our proposed scheme employs a non-linear MAP decoding algorithm to calculate the soft probabilities of the network coded symbols, rather than the linear filtering algorithm used in [27]. Our proposed CCD algorithm is different from the successive IF linear receiver [28]. In [28], the computed superposition of the signals from a previous layer is physically cancelled from the received signal to compute the next layer, which yields new residual interference and quantization errors. In contrast, in our proposed CCD scheme, the NC message computed in the previous layer is not physically cancelled. Instead, it is used as a priori information in computing the NC message of the next layer in the MAP algorithm.

B. Design of $\mathbf{G}_N$

So far we have presented a new NOMA scheme using reliable linear PNC and CCD, for a given choice of NC generator matrix $\mathbf{G}_N$. This section is devoted to finding the $\mathbf{G}_N$ that yields the optimal performance of the proposed scheme. Since the CSI is available only at the receiver and a fixed date rate is employed, the performance metric under consideration is the decoding error probability.

1) Design Problem Formulation: Consider one block of transmission with channel coefficient $\mathbf{H}$. The optimal NC generator matrix that minimizes the decoding error probability is formulated as

$$\mathbf{G}_N^* = \arg \min_{\mathbf{G}_N:\, \mathrm{Rank}(\mathbf{G}_N)=K} \Pr(\widehat{\mathbf{W}} \neq \mathbf{W} | \mathbf{G}_N). \tag{25}$$

For a channel-coded system, finding the exact optimal NC generator matrix is well-known to be very difficult.

In this section, we develop practical design solutions to (25). We first characterize the effective minimum distances of the proposed scheme w.r.t. an un-channel-coded system. Then, a choice of $\mathbf{G}_N$ that maximizes the effective minimum distance is obtained. Notably, the minimum distance dictates the error probability performance of an un-channel-coded system as SNR becomes sufficiently high.

Instead of (25), we consider solving

$$\mathbf{g}_l^* = \arg \min_{\mathbf{g}_l \in \{0,\cdots,q-1\}^K, \mathbf{g}_l \neq \mathbf{0}} \Pr(\widehat{\mathbf{w}}_l \neq \mathbf{w}_l | \mathbf{g}_l) \tag{26}$$

subject to

$$\mathrm{Rank}\left(\left[\mathbf{g}_1^T, \cdots, \mathbf{g}_l^T\right]\right) = l$$

where $l = 1, \cdots, K$. Here, the NC generator matrix is obtained in a component-wise (row-by-row) manner.

2) An Approximate Solution: Let us denote $\mathbf{x}[t] = [x_1[t], \cdots, x_K[t]]^T$ and $\widetilde{\mathbf{x}}[t] = [\widetilde{x}_1[t], \cdots, \widetilde{x}_K[t]]^T$ be two different signal vectors at time instant t. For notational simplicity, we will omit the time instant t in the following. Define

$$\delta \triangleq (\mathbf{x} - \widetilde{\mathbf{x}})\, \gamma \tag{27}$$

as the *difference vector* (DV) w.r.t. to the pair of $(\mathbf{x}, \widetilde{\mathbf{x}})$. Consider all possible transmitted signal vector pairs, the entries of δ belong to $\{1-q, \cdots, q-1\}$. The squared Euclidean distance w.r.t. $(\mathbf{x}, \widetilde{\mathbf{x}})$ is

$$E_s \left\|\mathbf{H}\left(\mathbf{x} - \widetilde{\mathbf{x}}\right)\right\|^2 = \frac{E_s}{\gamma^2} \left\|\mathbf{H}\delta\right\|^2. \tag{28}$$

Denote by

$$d_1 = \min_{\delta_1 \in \{1-q,\ldots,q-1\}^K, \|\delta_1\| \neq 0} \frac{E_s}{\gamma^2} \left\|\mathbf{H}\delta_1\right\|^2 \tag{29}$$

the minimum squared Euclidean distance for all possible $(\mathbf{x}, \widetilde{\mathbf{x}})$ pairs, and

$$\Delta_1 = \arg\min_{\delta_1 \in \{1-q,\ldots,q-1\}^K, \|\delta_1\| \neq 0} \frac{E_s}{\gamma^2} \left\|\mathbf{H}\delta_1\right\|^2 \tag{30}$$

the associated DV. Similarly, let d_l and Δ_l respectively be the lth smallest squared Euclidean distance for all possible $(\mathbf{x}, \widetilde{\mathbf{x}})$ pairs, subject to

$$\mathrm{Rank}(\mathrm{mod}([\Delta_1, \cdots, \Delta_l], q)) = l, \text{ for } l = 1, \cdots, K. \tag{31}$$

Let

$$\mathbf{r}_l \triangleq \mathrm{mod}(\Delta_l, q), \quad l = 1, 2, \cdots K. \tag{32}$$

and

$$\mathbf{R} = \left[\mathbf{r}_K, \mathbf{r}_{K-1}, \cdots, \mathbf{r}_1\right]. \tag{33}$$

Recall that at a high SNR regime, i.e., $\rho \to \infty$, a necessary condition to the solution to (26) is the maximization of the minimum squared Euclidean distance.

Now, we are in the position to present our design solution, given by

$$\left[\mathbf{g}_1^T, \cdots, \mathbf{g}_l^T\right]^T : \left[\mathbf{g}_1^T, \cdots, \mathbf{g}_l^T\right]^T \otimes \mathbf{R} = \Psi, \tag{34}$$

where Ψ is a lower-triangular matrix with non-zeros diagonal entries, i.e,

$$\Psi = \begin{bmatrix} \psi_{1,1} & 0 & \cdots & 0 \\ \psi_{2,1} & \psi_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{K,1} & \psi_{K,2} & \cdots & \psi_{K,K} \end{bmatrix} \tag{35}$$

with $\psi_{i,i} \neq 0$ for $i = 1, \cdots, K$. With such a choice, the effective minimum distance w.r.t the lth NC coefficient vector $\mathbf{g}_l$ is no less than d_{K-l+1}, and this leads to minimized error probability as SNR becomes sufficiently large. In Appendix I, we include a detailed explanation of (34) and (35).

3) Comments on the Presented Solution: The solution presented in (34) and (35) is only an approximation to the original problem (25) in the following aspects:

1. The presented solution minimizes the error probability only as SNR becomes sufficiently high. At a medium-to-low SNR, the effective distance spectrum related to the choice of $\mathbf{G}_N$ must be characterized, which is a challenging issue in the literature.

2. The presented solution is obtained based on analyzing the effective minimum distance of the un-channel-coded system. To find the exact solution for a channel-coded system, a full characterization of the effective distance spectrum, that jointly takes into account the algebraic structure of channel code generator matrix $\mathbf{G}_C$ and NC generator matrix $\mathbf{G}_N$, is required. This is a difficult task in the current literature.

It is noteworthy that another approximate solution to $\mathbf{G}_N$ based on optimizing the computation rate, by calculating the mutual information between the received signal and the NC messages, may be obtained. According to intensive numerical results, it has been observed that if the channel code is indeed capacity-approaching and the block length n is sufficiently long, such a choice of $\mathbf{G}_N$ will yield very good approximation to the exact optimal solution to (25). However, in a practical scenario where the channel code is not capacity-approaching (or the codeword length n is not long enough), such a choice of $\mathbf{G}_N$ may not be an appropriate option.

3. The presented solution is obtained where parallel computation was assumed. The NC generator matrix optimized for the parallel computation is not necessarily optimized for the cascade-computation. This is because with the side information from the previous layers, the effective distance associated with the constellation is likely to change. Unfortunately, there lacks a systematic way for characterizing the effective minimum distance under cascade-computation. In our future works, we will attempt to address these challenges in order to obtain a further improved solution to the original problem (25).

In the next section, numerical results will show that our presented approximate solution yields competitive performance for the proposed NOMA system. We note that if the exact optimal solution, which addresses all of the above three aspects, can be obtained, our proposed NOMA scheme will have further improved performance.

C. On the Complexity of CCD

In this section, the complexity of the proposed CCD scheme will be evaluated and compared to that of IDD scheme. Let us first look at the complexity of the symbol-wise detection. In the CCD scheme, the symbol-wise detection complexity for the first computation layer in a K-user MAC is $O(q^K)$ as the detection is based on the entire constellation, which has a number of q^K constellation points. For the second computation layer, the symbol-wise detection complexity is $O(q^{K-1})$ as the number of constellation points relevant to the detection is reduced from q^K to q^{K-1}. Likewise, for a general lth, $1 \leq l \leq K$, computation layer, the symbol-wise detection complexity is $O(q^{K-l+1})$. There are K computation layers in total for the proposed CCD scheme. Hence the symbol-wise detection complexity for all K computation layers is

$$\sum_{l=1}^{K} O(q^{K-l+1}). \tag{36}$$

In the benchmark IDD scheme with APP detection, the symbol-wise detection is always based on the entire constellation. Therefore the symbol-wise detection complexity for each iteration is $\sum_{l=1}^{K} O(q^K)$. Let I be the number of iterations, the symbol-wise detection complexity in total is

$$I \sum_{l=1}^{K} O\left(q^K\right). \tag{37}$$

It can be seen that the symbol-wise detection complexity of the proposed CCD scheme is strictly less than that of the existing IDD scheme.

Now, let us look at the channel code decoding complexity of the two schemes. For a length-n IRA code over GF(q), the iterative BP decoding complexity is in the order of $O(nq \log q)$. In the CCD scheme, only one NC message is decoded in a computation layer. Hence the channel code decoding complexity for all K computation layers in total is

$$O(Knq \log q). \tag{38}$$

In the IDD scheme, all K users' messages are decoded in each iteration. The channel code decoding complexity for each receiver iteration is $O(Knq \log q)$, which is the same as that of the CCD scheme. For I receiver iterations, the channel code decoding complexity in total is

$$O(IKnq \log q). \tag{39}$$

It is obvious that if more than one receiver iteration is used in the IDD scheme, the channel code decoding complexity of the proposed CCD scheme is strictly lower than that of the IDD scheme.

It is noteworthy that the CCD scheme introduces some overheads on the complexity. First, the users' messages are recovered from the computed NC messages by solving linear equations, i.e., by $\widehat{\mathbf{U}} = \mathbf{G}_N^{-1} \otimes \widehat{\mathbf{W}}$. The complexity for obtaining $\mathbf{G}_N^{-1}$ is of order $O\left(2K^{2.376}\right)$ if Coppersmith–Winograd algorithm is used. Second, the complexity for finding the design solution to $\mathbf{G}_N$, as presented in (34) and (35), is of order $O\left(K\,(2q-1)^K\right)$. We emphasize that both of the two additional operations only need to be done once per-block. Thus, the introduced complexity overheads become negligible as the block length n increases.

Compared to the existing receiver, our proposed scheme requires to find the optimized NC generator matrix, which will introduce additional system delay. We note that the NC generator matrix is required to be updated only once for each block of fading channel realization. Therefore, the additional system delay may be negligible when the channel coherence time is sufficiently large, i.e., the block length is greater than several hundreds. Compared to the IDD receiver, our proposed scheme can considerably reduce the system delay as there is no receiver iteration, while achieving a similar near-optimal performance. For applications that can tolerate moderate to large system delay, our proposed scheme is more competitive than the existing IDD scheme and the parallel computation scheme. For application where the system delay is of paramount importance, parallel computation may be a good choice due to its low system delay at the price of degraded error probability performance.

In summary, the total computation complexity of CCD is lower than that of the existing IDD scheme. In particular, as the number of iterations in the IDD scheme increases, the complexity of CCD becomes much lower than the of the IDD scheme. In addition, the proposed CCD scheme does not consist of interleavers and deinterleavers, which are used in the IDD scheme to randomize the information exchanged between the detector and the channel decoders. This may save implementation cost and processing latency.

V. Numerical Results

This section presents numerical results of our proposed NOMA scheme using reliable PNC and CCD for a block fading MAC. We consider that the fading coefficients follow i.i.d. Rayleigh distribution. We consider the average frame error rate (FER) at the receiver. In all simulations, the block length of the IRA channel codes is set to 1000, and the code rate is set to $\frac{3}{4}$. In the simulations, we convert a K by N complex-valued model to a $2K$ by $2N$ real-valued model by the means as shown in Appendix II. The real and imaginary parts of the transmitted signal convey two independent q-PAM signals respectively, resulting in a q^2-QAM signal which is the Cartesian product of two q-PAM signal sets. The real part and the imaginary part of the complex channel are then generated i.i.d and randomly in the simulations.

Three benchmark schemes are considered in this work. The first one is the FER performance of the interference-free bound, also called the single-user bound, where it is assumed that different users' signals are non-interfering (as if there is only a single-user in the system), and standard maximum ratio combining is used in the 1-by-N channel [3], [5]. We first obtain the FER of the single-user ($K = 1$) N receive-antennas system, denoted by FER_1, by using a symbol-wise ML detection algorithm across the N antennas and the non-binary IRA BP decoder. Then, the interference-free bound for a K-user N receive-antennas system is calculated by $\mathrm{FER}_K = 1 - (1-\mathrm{FER}_1)^K$. This provides a lower bound on the error probability that any MUD scheme can achieve. The second benchmark scheme is the FER performance of the IDD scheme, which operates by iterating between the optimal joint MAP detector and the BP decoder. In this scheme, each user employs a random interleaver, which is used to minimize the correlation among the soft information fed back from the soft-input soft-output channel code decoders, and hence reduces the estimation bias in the iterative process. A sufficient number of receiver iterations is set for the IDD scheme to converge. The third benchmark scheme is the FER performance of the parallel-computation and decoding scheme, which first obtains the NC generator matrix $\mathbf{G}_N$ by using (34) and (35) presented in Section IV. B and then decodes the NC messages in parallel. That is the computation of one NC message is independent from any other NC messages. For decoding each NC message, a symbol-wise NC message APP detector and a non-binary IRA BP decoder are used therein.

For our proposed scheme, the receiver first obtains the NC generator matrix $\mathbf{G}_N$ by using (34) and (35) based on the CSI and then computes the NC messages in a cascade manner, which is described in Section III. C. For computing each NC message, a symbol-wise NC message APP detector, as specified in (20) (21) and (22) for different computation layers, and a non-binary IRA BP decoder are used therein. In contrast to the IDD scheme, no interleaver/deinterleaver is used in our proposed scheme.

Fig. 5 shows the numerical results of a NOMA system with two users ($K = 2$) and two receiver-antennas ($N = 2$). It is observed that the performance of our proposed scheme is almost identical to that of the IDD scheme with 8 receiver iterations, for various values of q. Also, we observe that both the IDD scheme with 8 receiver iterations and our proposed scheme are within a small fraction of dB away from the single-user bound. It is noted that our proposed scheme outperforms the IDD scheme with one iteration by about 1.0 dB and 1.7 dB for $q = 3$ and $q = 5$, respectively, at a practical value of FER of 10^{-3}.

Fig. 6 shows the numerical results of a NOMA system with three users ($K = 3$) and three receiver-antennas ($N = 3$). Again, it is observed that the performance of our proposed scheme is almost identical to that of the IDD scheme with 8 receiver iterations. It is noted that our proposed scheme considerably outperforms the IDD scheme with one iteration. For example, at the FER of 10^{-3}, the propose scheme outperforms the IDD scheme with one iteration by about 0.7 dB and 1.6 dB for $q = 3$ and $q = 5$, respectively. Note that the complexity of our proposed scheme is similar to the IDD scheme with one iteration, as discussed in Section IV. C. Our proposed scheme approaches single-user bound within a fraction of dB at FER of 10^{-3}.

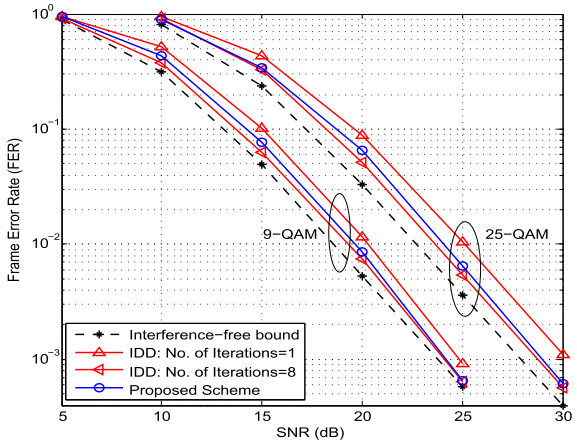


Fig. 5. FER performance of the proposed CCD scheme in a two-user ($K = 2$) two receive-antenna ($N = 2$) NOMA system.

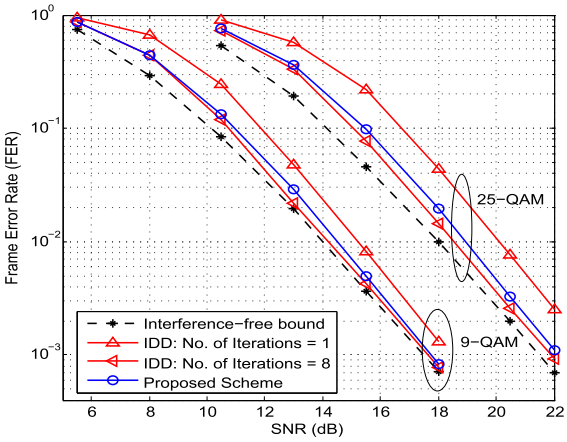


Fig. 6. FER performance of our proposed CCD scheme in a three-user ($K = 3$) three receive-antenna ($N = 3$) NOMA system.

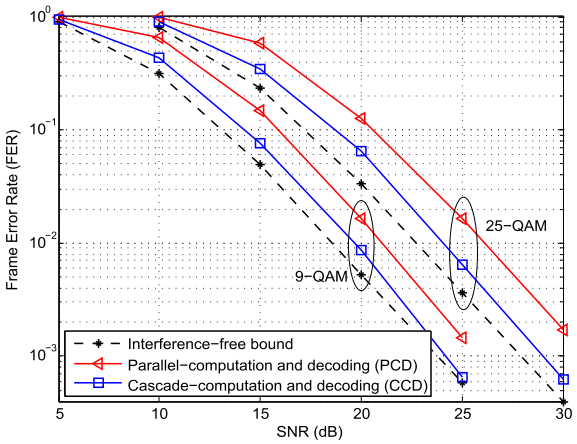


Fig. 7. Comparison of FER performance between the proposed cascade-computation scheme and the parallel computation scheme in a two-user ($K = 2$) two receive-antenna ($N = 2$) NOMA system.

Fig. 7 and Fig. 8 show a comparison between the parallel-computation decoding and cascade-computation decoding in a two-user ($K = 2$) two receive-antenna ($N = 2$) and three-user ($K = 3$) three receive-antenna ($N = 3$) NOMA models, respectively. It is observed that cascade-computation significantly outperforms the parallel-computation. This is in line with our illustration in Section III. C, that is, the side

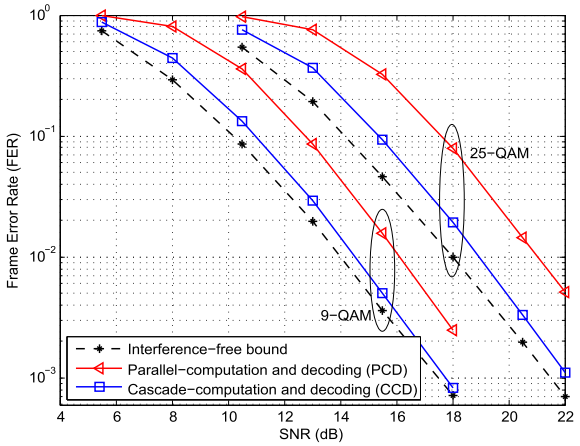


Fig. 8. Comparison of FER performance between the proposed cascade-computation scheme and the parallel computation scheme in a three-user ($K = 3$) three receive-antenna ($N = 3$) NOMA system.

information provided by the previous layers offers significant benefit in the computation of the NC messages.

In summary, numerical results have demonstrated that our proposed reliable PNC and CCD scheme (almost) achieves the performance of the IDD scheme, without using receiver iterations. This leads to significantly reduced computational complexity and easier implementation. Also, our proposed scheme with the approximate solution on $\mathbf{G}_N$ is within a fraction of dB away from the single-user bound, which may suggest that the scheme is near-optimal. We expect that the performance gap to the single-user bound can be further closed if the exact optimal $\mathbf{G}_N$ is used.

VI. Summary

A new reliable PNC and CCD scheme is proposed for the K-user NOMA. The proposed cascaded-computation was found to be a generalized onion peeling approach, where the NC messages were computed layer-by-layer with the help of side information provided by previous layers. The choice of the NC generator matrix was studied. The proposed scheme was shown to achieve the performance of iterative MAP detection and decoding and almost achieve the single-user bound, but without using receiver iteration. Thus, the complexity was significantly reduced and the implementation became easier. The proposed scheme significantly outperforms the IDD scheme with a single iteration.

This work can be further enriched from the following aspects. First, the exact optimal NC generator matrix is yet to be found. We expect that if the exact optimal solution can be obtained, our proposed NOMA scheme will have further improved performance. Second, a prime-size Galois field was considered as in linear PNC and CF, and it is yet to generalize the proposed scheme to the case where q is non-prime. Also, how to exploit the benefits of linear PNC in the downlink scenario remains largely unclear [29], [30].

Appendix I

Here we present detailed explanation of (34) and (35). Without loss of generality, we order the NC coefficient vectors in the ascending order according to their error probabilities

(i.e., the error probability w.r.t. $\mathbf{g}_1$ is smallest and that w.r.t. $\mathbf{g}_K$ is highest). Let $\mathbf{r}_l$ be the lth column of $\mathbf{R}$.

The NC coefficient vector $\mathbf{g}_1$ satisfies

$$\mathbf{g}_1 \otimes \mathbf{r}_l = 0, \quad \forall l = 1, 2, \cdots, K-1. \tag{40}$$

The effective squared Euclidean distance w.r.t. $\mathbf{g}_1$ is then d_K [13]. Since there does not exist a non-zero vector $\mathbf{g}_1$ which is perpendicular to all $\mathbf{r}_1, \cdots, \mathbf{r}_K$ in a K-dimension space, we have $\psi_{1,1} \neq 0$.

The NC coefficient vector $\mathbf{g}_2$ satisfies

$$\mathbf{g}_2 \otimes \mathbf{r}_l = 0, \quad \forall l = 1, 2, \cdots, K-2 \tag{41}$$

and $\text{Rank}([\mathbf{g}_1, \mathbf{g}_2]) = 2$. The effective squared Euclidean distance w.r.t. $\mathbf{g}_2$ is then d_{K-1}. Since $\mathbf{g}_2$ and $\mathbf{g}_1$ span the two-dimension subspace that is orthogonal to the subspace spanned by $\mathbf{r}_1, \cdots, \mathbf{r}_{K-2}$ and $\mathbf{g}_1 \otimes \mathbf{r}_{K-1} = 0$, we have $\mathbf{g}_2 \otimes \mathbf{r}_{K-1} \neq 0$. This leads to $\psi_{2,2} \neq 0$.

Similarly, the NC coefficient vector $\mathbf{g}_l$ satisfies

$$\mathbf{g}_l \otimes \mathbf{r}_l = 0, \text{for } l = 1, 2, \cdots, K-l. \tag{42}$$

and $\text{Rank}([\mathbf{g}_1, \cdots, \mathbf{g}_l]) = l$. Since $\mathbf{g}_l, \cdots, \mathbf{g}_1$ span the l-dimension subspace that is orthogonal to the subspace spanned by $\mathbf{r}_1, \cdots, \mathbf{r}_{K-l}$ and $\mathbf{g}_{l-1} \otimes \mathbf{r}_{K-l+1} = 0$, we have $\mathbf{g}_l \otimes \mathbf{r}_{K-l+1} \neq 0$. This leads to $\psi_{l,l} \neq 0$. Note that the squared Euclidean distance w.r.t. $\mathbf{g}_l$ is then d_{K-l+1}.

Other choice of $\mathbf{G}_N$ (except the matrices that permute the columns of $\mathbf{G}_N$) results in an effective squared Euclidean distance that is strictly no greater than the one given by $\mathbf{G}_N$. Thus the $\mathbf{G}_N$ obtained by (34) and (35) minimizes the error probability as SNR becomes sufficiently large.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[3] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

[4] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[5] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.

[6] T. Yang, X. Yuan, P. Li, I. B. Collings, and J. Yuan, "A new physical-layer network coding scheme with eigen-direction alignment precoding for MIMO two-way relaying," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 973–986, Mar. 2013.

[7] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Distributed MIMO in multi-cell wireless systems via finite-capacity links," in *Proc. ISCCSP*, Mar. 2008, pp. 203–206.

[8] T. Yang, Q. T. Sun, A. Zhang, and J. Yuan, "A linear network coding approach for uplink distributed MIMO systems: Protocol and outage behavior," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 2, pp. 250–263, Feb. 2015.

[9] J. Zhu and M. Gastpar. (2014). "Gaussian multiple access via compute-and-forward." [Online]. Available: https://arxiv.org/abs/1407.8463

[10] L. Lu, L. You, and S. C. Liew, "Network-coded multiple access," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2853–2869, Dec. 2014.

[11] L. Yang, T. Yang, Y. Xie, J. Yuan, and P. An, "Multiuser decoding scheme for K-user fading multiple-access channel based on physical-layer network coding," *IEEE Commun. Lett.*, vol. 20, no. 52, pp. 1046–1049, May 2016.

[12] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[13] J. Guo, T. Yang, J. Yuan, and J. A. Zhang, "Linear vector physical-layer network coding for MIMO two-way relay channels: Design and performance analysis," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2591–2604, Jul. 2015.

[14] P.-C. Wang, Y.-C. Huang, and K. R. Narayanan, "Asynchronous physical-layer network coding with quasi-cyclic codes," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 2, pp. 309–322, Feb. 2013.

[15] F. Rossetto and M. Zorzi, "A practical architecture for OFDM-based decode-and-forward physical layer network coding," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4747–4757, Sep. 2012.

[16] L. Lu, T. Wang, S. C. Liew, and S. Zhang, "Implementation of physical-layer network coding," *Phys. Commun.*, vol. 6, pp. 74–87, Mar. 2013.

[17] S. T. Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2764–2772, Nov. 2003.

[18] X. Yuan, Q. Guo, X. Wang, and L. Ping, "Evolution analysis of low-cost iterative equalization in coded linear systems with cyclic prefixes," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 301–310, Feb. 2008.

[19] R. Milner, L. K. Rasmussen, and F. Brännström, "Recursive LLR combining in iterative multiuser decoding of coded CDMA," in *Proc. Austral. Commun. Theory Workshop*, Feb. 2007, pp. 1–6.

[20] T. Yang, J. Yuan, and Z. Shi, "Jointly Gaussian approximation and multi-stage LLR combining in the iterative receiver for MIMO-BICM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5250–5256, Dec. 2008.

[21] M.-C. Chiu, "Bandwidth-efficient modulation codes based on nonbinary irregular repeat–accumulate codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 152–167, Jan. 2010.

[22] L. Yang, T. Yang, J. Yuan, and P. An, "Achieving the near-capacity of two-way relay channels with modulation-coded physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5225–5239, Sep. 2015.

[23] T. Yang and I. B. Collings, "On the optimal design and performance of linear physical-layer network coding for fading two-way relay channels," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 956–967, Feb. 2014.

[24] S.-Y. R. Li, Q. T. Sun, and Z. Shao, "Linear network coding: Theory and algorithms," *Proc. IEEE*, vol. 99, no. 3, pp. 372–387, Mar. 2011.

[25] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. 12th Annu. Int. Conf. Mobile Comput. Netw.*, 2006, pp. 358–365.

[26] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proc. IEEE*, vol. 99, no. 3, pp. 438–460, Mar. 2011.

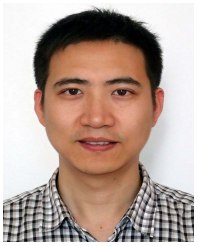
[27] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[28] O. Ordentlich, U. Erez, and B. Nazer, "Successive integer-forcing and its sum-rate optimality," in *Proc. Allerton Conf. Commun., Control, Comput.*, Oct. 2013, pp. 282–292.

[29] T. Yang, X. Yuan, and Q. T. Sun, "A signal-space aligned network coding approach to distributed MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 27–40, Jan. 2017.

[30] T. Yang, "Distributed MIMO broadcasting: Reverse compute-and-forward and signal-space alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 581–593, Jan. 2017.

Tao Yang (S'07–M'10) received the B.Sc. degree in electronic engineering from Beihang University, Beijing, China, in 2003, and the master's and Ph.D. degrees in electrical engineering from the University of New South Wales (UNSW), Sydney, Australia, in 2006 and 2010, respectively. He was an OCE Post-Doctoral Fellow with the Commonwealth Scientific and Industrial Research Organization (CSIRO) ICT Centre, Australia. He was with UNSW as a Research Fellow. He is currently a Lecturer with the Global Big Data Technologies Centre and the School of Computing and Communications, University of Technology Sydney. He has authored over 50 research articles in IEEE journals and conferences. His research expertise and interests include physical-layer network coding, multi-user and MIMO, error-control coding, iterative signal processing, and network information theory. He served as a TPC member of the IEEE ICC and WCNC. He holds an Australian Research Council Discovery Early Career Research Award Fellowship. He was a recipient of the Australian Postgraduate Award, the NICTA Research Project Award, the Supplementary Engineering Award from UNSW, and the Publication Award from the CSIRO ICT Centre.



Lei Yang (S'13) received the B.Sc. and M.Sc. degrees in electronics engineering from Beihang University, Beijing, China, in 2004 and 2007, respectively, and the Ph.D. degree in communications and information system from the Beijing Institute of Technology, Beijing, in 2015. He is currently a Post-Doctoral Research Fellow with the University of New South Wales, Sydney, Australia. His current research interests include error control coding, physical-layer network coding, and iterative signal processing.

Y. Jay Guo (F'14) received the bachelor's and master's degrees from Xidian University, China, in 1982 and 1984, respectively, and the Ph.D. degree from Xian Jiaotong University, China, in 1987.

He held various senior leadership positions in Fujitsu, Siemens, and NEC in the U.K. He served as a Research Director in CSIRO for over nine years, directing a number of ICT research portfolios. He is currently a Distinguished Professor and the Director of Global Big Data Technologies Centre, University of Technology Sydney, Australia. His research interests include antennas, mm-wave, and THz communications and sensing systems.

Dr. Guo is a Fellow of the Australian Academy of Engineering and Technology, a Fellow of IET, and a member of the College of Experts of Australian Research Council. He received a number of most prestigious Australian national awards, and was named one of the top 100 most influential engineers in Australia, in 2014 and 2015. He has chaired numerous international conferences. He is the International Advisory Committee Chair of the IEEE VTC2017, the General Chair of ISAP2015, the iWAT2014, and the WPMC'2014, and the TPC Chair of the 2010 IEEE WCNC, and the 2012 and 2007 IEEE ISCIT. He serves as a Guest Editor of special issues on antennas for satellite communications, and antennas and propagation aspects of 60–90 GHz wireless communications, in the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, special issue on communications challenges and dynamics for unmanned autonomous vehicles, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and special issue on 5G for mission critical machine communications, and the *IEEE Network Magazine*.

Jinhong Yuan (M'02–SM'11–F'16) received the B.E. and Ph.D. degrees in electronics engineering from the Beijing Institute of Technology, Beijing, China, in 1991 and 1997, respectively. From 1997 to 1999, he was a Research Fellow with the School of Electrical Engineering, University of Sydney, Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, where he is currently a Telecommunications Professor with the School. He has authored two books, three book chapters, over 200 papers in telecommunications journals and conference proceedings, and 40 industrial reports. He is a Co-Inventor of one patent on MIMO systems and two patents on low-density-parity-check codes. His current research interests include error control coding and information theory, communication theory, and wireless communications. He has co-authored three Best Paper Awards and one Best Poster Award, including the Best Paper Award from the IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011, and the Best Paper Award from the IEEE International Symposium on Wireless Communications Systems, Trondheim, Norway, in 2007. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He served as the IEEE NSW Chair of Joint Communications/Signal Processions/Ocean Engineering Chapter from 2011 to 2014.