Gaussian and Fading Multiple Access using Linear Physical-layer Network Coding

Qiuzhuo Chen, Fangtao Yu, Tao Yang, *Member, IEEE*, and Rongke Liu, *Senior Member, IEEE*

Abstract

This paper concerns with efficient communication over Gaussian and fading multiple-access channels (MACs). Existing orthogonal multiple-access (OMA) and power-domain nonorthogonal-OMA (NOMA) cannot achieve all rate-tuples in the MAC capacity region. Meanwhile, code-domain NOMA schemes usually require big-loop receiver-iterations for multi-user decoding, which is subject to high implementation cost and latency. This paper studies a linear physical-layer network coding multiple access (LPNC-MA) scheme that is capable of achieving any rate-tuples in the MAC capacity region without receiver iterations. For deterministic Gaussian MACs with M users, we propose to utilize q-ary irregular repeat accumulate (IRA) codes over finite integer fields/rings and q-ary pulse amplitude modulation (q-PAM) as the underlying coded-modulation. The receiver sequentially computes M network coded (NC) message sequences, where the previously computed message sequence is used as side information in computing subsequent ones. All users' messages are then recovered by solving the computed M NC messages via the inverse of the NC coefficient matrix. A joint nested code construction and extrinsic information transfer (EXIT) chart based code optimization method is developed, yielding near-capacity performance (within 0.7 and 1.1 dB the capacity limits for two and three users respectively). For fading MAC, we study the symmetric rate of LPNC-MA, and propose a pragmatic method for identifying the mutual information (MI) maximizing network coding coefficient matrix. Numerical results demonstrate that the frame error rate (FER) of the optimized LPNC-MA is within a fraction of dB the outage probability of fading MAC capacity and LPNC-MA remarkably outperforms NOMA-SIC and IDMA in high spectral efficiency regime, while avoiding the big-loop receiver iteration.

Index Terms

NOMA, physical-layer network coding, compute-forward, multiple access, coded modulation, iterative decoding, successive interference cancellation

The authors are with School of electronics information engineering, Beihang University, China. This work is supported by the National Key R&D Program of China (No.2020YFB1807102) and the "Zhuoyue" program of Beihang University. Both Q. Chen and F. Yu are the first author.

I. Introduction

Multiple access (MA) is a pivotal component in wireless communication systems which distinguishes different generations of mobile networks. Beyond 5G (B5G) and 6G networks are envisaged to support vast range of throughput-hungry and latency-sensitive scenarios such as ubiquitous IoT, mobile cloud computing, digital twin, which calls for next generation multiple access techniques that provide tremendously higher system load, higher spectral-energy efficiency, lower latency and larger coverage.

From the perspective of information theory, Shannon first put forward the multiple access channel (MAC) model [1]. Ahlswede characterized the capacity region of discrete memoryless MAC [2]. Cover *et al.* showed the capacity region of the Gaussion MAC [3]. These results showed that remarkably larger rate-region can be achieved by non-orthogonal multiple access (NOMA), which allows users to transmit simultaneously in the same operating frequency band, in contrast to orthogonal multiple access (OMA) such as FDMA [4], TDMA [5], orthogonal CDMA [6], [7], OFDMA [8]. From 1G to 5G, OMA based design methods were adopted, which attempted to avoid the existence of multi-user interference. Such treatment of multi-user interference simplifies the decoding operation at the receiver, at the cost of 1) reduced spectral-efficiency, 2) poor flexibility in system load and rate allocation and 3) sophisticated algorithms required to maintain the orthogonality in time or sub-carriers. These drawbacks of OMA set the bottlenecks in B5G/6G application scenarios involving massive connectivity, low latency and high mobility requirements.

A. Power-Domain and Code-Domain NOMA

NOMA releases the orthogonality constraint imposed by OMA and can achieve the full capability of MA. The core issue becomes how to efficiently and flexibly process the multi-user interference, i.e., multi-user detection or multi-user decoding (MUD). One straightforward method is to use successive interference cancellation (SIC), which is widely referred to as power-domain NOMA-SIC [9], [10]. By decoding the stronger user first and then canceling out its interference to the weaker user, a corner point of the MAC region can be achieved, where single-user decoding is sufficient. The SIC technique was applied in the full-duplex system to handle the self-interference, which is called full-duplex NOMA (FD-NOMA) [11]. It is shown that FD-NOMA outperforms half-duplex (HD) system at moderate SNR [11], [12]. Yet, SIC only achieves the corner points and the entire "dominant face" of MAC region needs time-sharing between SIC with different ordering. This is clearly impractical in almost all typical application scenarios[1]. In terms

[1] We will provide detailed discussion on this in Section III.

of achievable symmetric rate, which is of high relevance in open-loop systems with receiver channel state information (CSI) only, NOMA-SIC may perform worse than OMA.

Beyond NOMA-SIC, advanced coding and signal processing techniques have been introduced to address the MA problem. Those schemes are widely referred to as code-domain NOMA. Thanks to the invention and renewed understanding of turbo and low-density parity-check (LDPC) code principle, turbo-like iterative detection and decoding are exploited. Essentially, the goal is to approximate the joint maximum likelihood (ML) decoding that is capacity region achieving, with realistic computational complexity. To achieve the ultimate performance of code-domain NOMA, iterative multiuser detection and decoding is required. Specifically, the inner multiuser detector calculates soft a posteriori probabilities (APPs) symbol-wisely for all users. The APPs sequence w.r.t. each user is forwarded to that user's channel code decoder, which could be iterative by itself. The outputs of a bank of M channel-code decoders are fed back to the inner detector. With the updated a priori information, the inner detector refines its output APPs and forwards them again to the outer channel-code decoders. Such process continues until convergence is achieved. Here, the iteration within the outer channel code decoder itself is referred to as the "small-loop iteration". The iteration between the multi-user detection and a bank of M channel-code decoders is referred to as the "big-loop iteration". Note that each big-loop iteration consists of a (large) number of small-loop iterations. Along this line, turbo-CDMA receivers were first proposed by Wang and Poor in late 90s [13]. Li *et al.* introduced a chip-level interleaved CDMA, named after interleave division multiple-access (IDMA) [14]. The chip interleaver enables uncorrelated chip interference and thus simple matched filter optimally combines the chip-level signal to yield the bit-level soft information.

Sparse code multiple-access (SCMA) differs from IDMA in that each bit-level information is spread only to a small number of chips, which forms a sparse matrix in the representation of the multi-user signal that can be depicted using a bi-partite factor graph [15], [16]. SCMA can also be upgraded to cater for grant-free random access, which is an excellent fit to the massive-connectivity IoT scenario. For both IDMA and SCMA, irregular design of the spreading/sparse codes have been investigated including the work of ourselves [17], which yields improved convergence behavior of the multi-user decoding. There are other code-domain NOMA techniques proposed such as PDMA, MUSA and etc. [18]. Albeit all the benefits and potential performance enhancement promised by NOMA, there are quite some challenging issues for them to be utilized for next generation MA. Empirically the required number of big-loop receiver iteration is 4 to 10. If the outer decoder is sum-product based LDPC decoder, the number of small-loop iterations is hundreds per big-loop iteration, and could be thousands for all 10 big-loop iterations. Note that the big-loop iterations have to be done in serial, which can bring about significant processing delay and huge expense in computation.

In short, power-domain NOMA-SIC may have low implementation cost but performs quite far from the MAC capacity region, whereas code-domain NOMA schemes may perform close to the MAC capacity region but the implementation is costly due to the big-loop receiver iteration. This motivates the following question: can we achieve the entire MAC region without using the big-loop receiver iteration?

B. Compute-Forward and Physical-Layer Network Coding

Compute-forward (CF) is an information theoretic notion, which exploits the property that the linear combination with integer coefficients of any two lattice points belong to the lattice to efficiently recover linear combinations of M users' messages. Existence of good nested lattice codes are used to prove the achievable rates of CF. Linear physical-layer network coding (LPNC) is a practical technique to realize the theoretical idea of CF. Specifically, LPNC implements soft APPs calculation, practical linear codes and modulation, as well as practical BP decoding algorithms that replaces the theoretical notion of nested lattice codes. It is widely known that LPNC is a practical embodiment of the CF notion. In this paper, we use CF and LPNC interchangeably whichever is appropriate within the context.

The superposition of M users' signals in real field is linked to the linear combination with integer coefficients of their codewords in finite field/ring. Then, single-user decoding can be used to directly compute their integer-sum in finite field/ring, referred to as *network coded (NC) message*, without need to completely decoding of all users' individual messages. For the two-user setup where the receiver is set to compute one NC message, up to doubled throughput relative to conventional complete-decoding based method is achieved using LPNC [19]. Many prior research findings showed the effectiveness of using CF or LPNC in solving communication over wireless channels with side information [20], [21]. The notion of CF is also generalized to the multiple-input multiple-output (MIMO) setup, e.g., the integer-forcing (IF) framework [22], [23].

Recently, Zhu and Gastpar showed that any rate-tuple of the entire Gaussian MAC capacity region can be achieved using CF, and the scheme was named CFMA [24]. Almost at the same time, we investigated using practical LPNC, that borrows the notion of CF, for fading multiple access with multi-antenna at the receiver [20], [25]. Later, Sula *et al.* studied practical design of CF for the Gaussian MAC with practical quadrature amplitude modulation (QAM) and binary LDPC codes [26]. These initial works demonstrated that without receiver iteration, CF/LPNC can potentially achieve the entire capacity-region with *sequential computation and decoding* (SCD). This translated into improved energy-spectral efficiency, reduced computation complexity and processing delay. The concept of SCD is: 1) sequentially computes $L = M$ NC messages in a finite integer field/ring; 2) the previously computed NC messages are used as side information to compute the subsequential NC messages. Finally, all users' messages can be recovered by multiplying the inverse of the network coding coefficient matrix (with full rank M). It was understood that NOMA-SIC is a special case of LPNC-MA with the network coding coefficient matrix set to a diagonal or permutation matrix.

Another MAC system that applies the notion of PNC is the network-coded MA (NCMA) [27]. NCMA differs from CFMA in a) it considered both the medium-access layer and physical-layer, and b) successive computation decoding, which can yield a greater coding gain, was not exploited. Also, the information theoretic characterization is yet to be elaborated for NCMA.

In the current literature, CFMA or LPNC-MA with practical channel code and PAM/QAM modulation remains insufficiently researched. The primary challenging issues involves: 1) practical capacity-approaching nested code construction and decoding for Gaussian MAC; 2) code optimization under SCD; 3) identifying the optimal network coding coefficient matrix for fading MAC with receiver CSI only. Following the spirit of nested lattice codes, [28] developed low density lattice codes (LDLC), but the optimization of the code degree distribution is formidable. The practical aspect of CFMA was reported in [26], which only considered two-user Gaussian MAC, binary CF, and with fixed NC coefficients. Our work generalizes it to M users, q-ary CF, and any NC coefficients, with nested code optimization.

C. Contribution of this Paper

- We present a LPNC-MA framework, which is regarded as a practical embodiment of the CFMA notion. We propose to employ q-ary irregular repeat accumulate (IRA) codes over finite integer fields/rings and q-PAM as the underlying coding-modulation for LPNC-MA, as well as a practical SCD algorithm at the receiver. This is in contrast to previous works that are based on off-the-shelf binary channel codes with many-to-one mapping to QAM symbol, such as those in bit-interleaved coded modulation (BICM) [29], superposition coded modulation (SCM) [30] and the work on CFMA in [26]. The achievable mutual information of the proposed method with q-PAM is presented and is shown to be able to achieve the entire MAC region in practice.
- We propose a joint nested construction and degree distribution optimization method for the q-ary IRA codes in LPNC-MA. Our treatment is based on the EXIT chart curve fitting technique tailored for the nested IRA codes and SCD. For a number of typical rate-tuples, we demonstrate that the designed LPNC-MA is within 0.7 dB and 1.1 dB the capacity limit for two and three users, respectively, and overwhelmingly outperforms NOMA-SIC.

- For fading MAC where the channel state information (CSI) is only available at the receiver, we analyze the achievable symmetric rate and present a new efficient method for identifying the symmetric mutual information maximizing NC coefficient matrix of LPNC-MA. It is demonstrated that the outage probability (OP) of our designed LPNC-MA scheme performs within a fraction of dB the lower bound of the fading MAC, and easily outperforms NOMA-SIC by more than 5 dB. At a practical frame error rate (FER) level, LPNC-MA's FER is within 1dB to the lower bound, using practical IRA codes. For a large number of users, LPNC-MA can cooperate with orthogonal spreading codes and extend to any number of users. Simulation results show that for 16-user case, LPNC-MA significantly outperforms IDMA in high spectral efficiency regime, while avoiding the big-loop receiver iteration.

D. Scopes

This paper focuses on multiple-access using practical channel coding and q-PAM modulation. By Gaussian MAC we mean that the receiver is subject to additive white Gaussian noise. No efforts are devoted to the Gaussian shaping of the coded-modulation symbols. All the achievable rates or mutual information in this paper are with respect to the q-PAM channel input, not Gaussian inputs. This paper focus on single-antenna setup and uplink MAC. Our developed LPNA-MA can also be applied to the multi-antenna and downlink broadcast channel scenarios, e.g., in conjunction with the integer-forcing and reverse integer-forcing for MIMO receiver or precoder [31]. This paper focus on narrow band system (e.g., a segment of sub-carriers within the coherent bandwidth) and the model involves no inter-symbol-interference. In addition, massive MIMO and grant-free random-access [32] are out of the scope of this paper.

II. LPNC-MA FOR GAUSSIAN MAC

A. Preliminary

Consider an uplink MAC where M users transmit simultaneously in the same frequency band. The baseband equivalent signal model is given by

$$\mathbf{y} = \sum_{m=1}^{M} h_m \sqrt{P_m} \mathbf{x}_m + \mathbf{z}, \tag{1}$$

where $\mathbf{x}_m,\ m = 1, 2, \cdots, M$, denotes user m's coded-modulation symbol sequence, P_m denotes user m's transmitting power, h_m denotes the channel gain, $\mathbf{y}$ denotes the received signal sequence at the receiver, and $\mathbf{z}$ denotes the additive white Gaussian noise (AWGN) with zero mean and variance σ^2. We consider that the average symbol energy of $\mathbf{x}_m$ is normalized. The power of the weakest user is normalized to 1, and the signal to noise ratio (SNR) is given by $\triangleq \frac{1}{\sigma^2}$ for simplicity. For Gaussian MAC, $h_m = 1,\ m = 1, \cdots, M$. For fading MAC, h_m denotes the fading channel coefficient of user i, which varies over different users and code blocks. For clarity of presentation, real-valued model is used throughout this paper. A M-user complex-valued model can be converted to a $2M$-user real-valued model [33], and the results developed in this paper apply.

In this section, we consider Gaussian MAC channel model with perfect receiver-side CSI. This setup applies to close-loop systems where the channel conditions remain constant for a long time and can be fed back to the transmitters, or the adaptive modulation and coding (AMC) scheme with exact channel estimation (including amplitude and phase) where an index for choosing (from a discrete set of) the coding rates and modulation-levels is sent to the transmitters. For open-loop systems where feeding back can not be done, the fading MAC model applies which will be separately studied in the next section.

1) Capacity Region of MAC: The capacity region of a Gaussian MAC is a polygon with a certain number of corner points. The entire MAC region can be achieved using joint ML decoding, where the proof was based on the random coding and jointly typicality decoding argument [3]. These corner points collectively determine the "Dominant Face" (DF) of the capacity region [3]. The rate-tuples that are on the DF are of the greatest interests in the research. In general, OMA cannot achieve the DF except one special point. In particular when the received powers of different users vary significantly, the special point that OMA touches with the DF is subject to huge gap in the rates among the users, which leads to fairness issue. With NOMA-SIC, only the corner points can be achieved. Different SIC ordering yield different corner points. The DF may be achieved by time-sharing of the SIC with different ordering, but this is very difficult to realize in practice. For example, there are far too many segments for allocating time for sharing among different SIC orders even when M is moderate, which leads to very short effective block length.

The code-domain NOMA schemes, such as iterative APP detection and decoding, CDMA/IDMA and SCMA etc., attempt to approximate the jointly ML decoding solution by iteratively exchanging the extrinsic information between the soft multi-user detector and a bank of M decoders. Typically, five to ten iterations between the inner multi-user detector and outer channel code decoders are required for convergence. If the channel code decoders are turbo or LDPC codes, which are iteratively decoded by themselves, the receiver consists of big-loop iteration, each involves a large number of small-loop iterations for channel code decoding. This may cause implementation difficulties and instability issues.

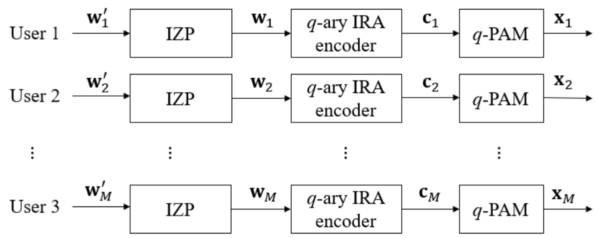


Fig. 1. Block diagram of the encoders of M-user LPNC-MA. Here IZR stands for irregular zero-padding.

B. LPNC-MA Encoder

We next present a LPNC-MA which realizes the CFMA notion and can be readily used for reliable communication over Gaussian and fading MACs.

The LPNC-MA encoders are shown in Fig. 1. The messages of the M users are denoted by $\mathbf{w}'_m \in \mathbb{Z}_q^{k_m}$, where $\mathbb{Z}_q^{k_m} \triangleq \{0, 1, \cdots, q-1\}^{k_m}, m = 1, 2, \cdots, M$, with sequence length given by k_m. Let $k = \max_{m=1,2,\cdots,M} k_m$ which denotes the common length of the M message sequences after irregular zero padding (IZP). The zero-padded messages of user m is denoted as $\mathbf{w}_m$. With common length k for any user, LPNC-MA uses q-ary linear codes over finite integer fields/rings and q-PAM. The encoding operation is represented by

$$\mathbf{c}_m = \mathbf{G} \otimes \mathbf{w}_m, \ m = 1, 2, \cdots M, \tag{2}$$

where $\mathbf{G}$ is a $n \times k$ generator matrix, n is the length of the encoded sequence. We define $\oplus$ and $\otimes$ be the addition and multiplication operations separately in fields/rings, where $\mathbf{S} \oplus \mathbf{V} \triangleq \mathrm{mod}(\mathbf{S} + \mathbf{V}, q)$, $\mathbf{S} \otimes \mathbf{V} \triangleq \mathrm{mod}(\mathbf{SV}, q)$. The information rate is $R_m = \frac{k_m}{n} \log_2 q$ bits/symbol for user m.

For q-PAM constellation point with uniform spacing, the one-to-one mapping is given by

$$\mathbf{x}_m = \frac{1}{\gamma}\left(\mathbf{c}_m - \frac{q-1}{2}\right) \in \left\{\frac{1-q}{2\gamma}, \cdots, \frac{q-1}{2\gamma}\right\}^n, \tag{3}$$

where γ is a normalization factor to ensure unit average symbol energy. The mapping function is denoted by $\delta(\bullet)$ where $\mathbf{x}_m = \delta(\mathbf{c}_m)$. Note that this paper also considers non-uniformly spaced constellations.

The coding-modulation above was referred to as a q-ary *modulation code* in [34]. For simplicity, we just use the name "q-ary linear codes" instead of "modulation codes" in this paper. Practical q-ary irregular repeat accumulate (IRA) codes over finite field with low complexity encoding and decoding, and near-capacity performance have been reported [34], and are employed in this paper. The encoder consists of repeater, interleaver and accumulator with linear complexity, and the decoder utilizes iterative belief propagation (BP) algorithm. Such coding-modulation is in contrast to previous works that are based on off-the-shelf binary channel codes with many-to-one mapping to QAM symbol, such as those in the work on practical CFMA in [26]. he modulation codes have better performance than that of binary codes with high order modulation, with comparable computation cost [35], [36]. Moreover, modulation codes have a one-to-one mapping between codewords and constellation points, while binary codes with high order modulation do not have which leads to a complex non-linear mapping after the superposition of modulation symbols.

Remark 1 (Irregular zero padding): The zero padding strategy ensures that the q-ary linear codes w.r.t. the users are nested. We emphasize that the positions of zero padding need to be carefully designed, with an irregular zero padding (IZP) method. The IZP pattern significantly affects the code degree distribution and the error-rate performance, as will be detailed in Subsection E.

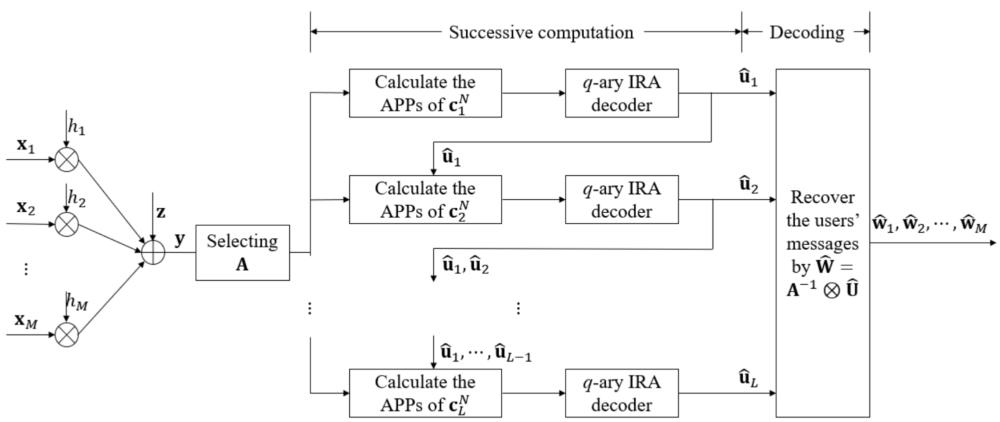


Fig. 2. Block diagram of LPNC-MA receiver.

C. Sequential Computation and Decoding (SCD)

Following the spirit of LPNC, the receiver is set to compute L linear combinations of users' messages over the finite field/ring, called *NC messages*. For MAC, $L = M$ and $\mathbf{A}$ has full rank M. For other setups such as distributed and relay networks, L may be smaller than M.

The l-th NC message sequence to be computed is written by

$$\mathbf{u}_l^T = \mathbf{a}_l \mathbf{W}, \ l = 1, 2, \cdots, L, \tag{4}$$

where $\mathbf{a}_l$ is the NC coefficient vector, and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_M]^T$ collects the message sequences of all users. All $L = M$ NC messages are collectively written as

$$\mathbf{U} = \mathbf{A} \otimes \mathbf{W}, \tag{5}$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_L \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{L1} & \cdots & a_{LM} \end{bmatrix}$ denotes the network coding coefficients matrix with entries belonging to $\{0, 1, \ldots, q-1\}$. Note that $\mathbf{A}$ is required to have rank M in the finite fields/rings such that $\mathbf{A}$ is invertible and $\mathbf{W}$ can be recovered for given $\mathbf{U}$.

The codewords of NC messages are given by

$$\mathbf{C}^N = \mathbf{G} \otimes \mathbf{U}, \tag{6}$$

With the property of linear superposition in feilds/rings, the linear combination of users' codewords is equal to the codewords of the linear combination of users' messages. The effective message length of the l-th NC message sequence is $k_l^N = \max_{a_{lm} \neq 0} k_m$, and the "computation rate" is $R_l^N = \frac{k_l^N}{n} \log_2 q$ bits/symbol [37].

The receiver is illustrated in Fig. 2, which consists of the following steps: 1) upon receiving $\mathbf{y}$ in (1), the receiver conducts symbol-by-symbol APP detection to obtain a soft estimate of the linear combination of the coded bits $\mathbf{c}_1^N, \cdots, \mathbf{c}_M^N$, with coefficients matrix $\mathbf{A}$. 2) The APPs of $\mathbf{c}_1^N, \cdots, \mathbf{c}_M^N$ are forwarded to the IRA decoder, which successively computes the linear combination of the message bits $\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_M$. 3) The decision on $\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_M$ are multiplied with $\mathbf{A}^{-1}$, yielding the decision on the users' messages $\hat{\mathbf{w}}_1, \cdots, \hat{\mathbf{w}}_M$. The detailed procedures of the SCD algorithm for the q-ary IRA coded LPNC-MA is given below:

Step 1) Calculate the symbol-wise a posterior probabilities (APPs) of the first NC message sequence. For convenience, let the t-th codeword of the l-th NC message sequence be denoted by $c_l^N[t]$, $t = 1, 2, \cdots, n$. This step is to estimate the probabilities of the codewords of NC messages from the received signals. Recall (1) and assume that $P_m = 1$. For a given NC coefficient matrix $\mathbf{A}$, the APPs are calculated by (7), similar to that in [19], where η is a normalization factor to ensure that $\sum_{i=0}^{q-1} p(c_1^N[t] = i|y) = 1$.

Step 2) Forward the APP sequence to the q-ary IRA decoder that carries out belief propagation (BP) algorithm to compute NC message sequence $\mathbf{u}_1$ (To avoid redundancy, the BP algorithm of the q-ary IRA code is not shown in this paper which can be found in [34]).). The computed result is denoted by $\hat{\mathbf{u}}_1$.

For the l-th ($l \geq 2$) NC message sequence, the previously computed $(l-1)$ NC messages are taken as

---

$$p\left(c_1^N[t] = i|y[t]\right) = \frac{1}{\eta} \sum_{a_{11}c_1[t] \oplus \cdots \oplus a_{1M}c_M[t] = i} \exp\left\{ -\frac{\left(y[t] - \sum_{m=1}^{M} h_m \delta\left(c_m[t]\right)\right)^2}{2\sigma^2} \right\}, \; i = 0, \cdots, q-1 \tag{7}$$

side information to help with computing the l-th NC message sequence's symbol-wise APPs. To do this, we convert the computed NC message sequences $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \cdots, \hat{\mathbf{u}}_{l-1}$ to NC codewords by

$$\hat{\mathbf{c}}_j^N = \mathbf{G} \otimes \hat{\mathbf{u}}_j, \ \ j = 1, 2, \cdots, l-1. \tag{8}$$

Next, they are used as side information to calculate the symbol-wise APPs for $\hat{\mathbf{c}}_l^N$ as in (9). After that, the APP sequence is delivered to the q-ary IRA decoder which computes the l-th NC message sequence $\hat{\mathbf{u}}_l$.

Step 3) After all M NC messages $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \cdots, \hat{\mathbf{u}}_M$ are sequentially computed, they are forwarded to a network coding decoder which solves M NC messages, that is

$$\hat{\mathbf{W}} = \mathbf{A}^{-1} \otimes \hat{\mathbf{U}}, \tag{10}$$

where $\mathbf{A}^{-1} \otimes \mathbf{A} = \mathbf{I}$, $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_M]^T$, which recovers all M users' messages. This completes the SCD.

The notion of network coding also applies to one-hop setup such as the multi-access channel. The processing of the receiver in this multiple-access scheme is viewed as a network consisting of multiple nodes. In the processing node l, it takes the $(l-1)$-th network coded message as a side information to compute the l-th NC message. At the final node of this processing network, all M NC messages are collected. By multiplying with the inverse of the coefficient matrix A, all users messages are recovered. Note that the complexity of the matrix inversion for recovering M users' messages is negligible w.r.t. the complexities of detection and decoding. The matrix inverse of computing is required once per block. The overhead of this is negligible for a moderate-to-large block size. The complexity of the multiplication of each network coded message M-tuple with is linear to M. This incurs an overhead that is minor compared to that of decoding and multi-user detection, whose complexity order that is a polynomial of M, in any multiple-access system. As such, the extra complexity in doing an inverse in GF(q) is not significant.

Furthermore, the encoding/decoding process in GF(q) does not incur significantly higher complexity. In particular, DFT/FFT based algorithm can be introduced to vastly decrease the check-node update in the iterative decoding over GF(q). When the DFT-based algorithm is not employed, in the CN update, we need

$$p\left(c_l^N[t] = i | y[t], c_1^N[t], \cdots, c_{l-1}^N[t]\right) = \frac{1}{\eta} \sum_{\substack{a_{j1}c_1[t] \oplus \cdots \oplus a_{jM}c_M[t] = \hat{c}_j^N[t], \\ j=1,\cdots,l-1, \ a_{l1}c_1[t] \oplus \cdots \oplus a_{lM}c_M[t] = i}} \exp\left\{ -\frac{\left(y[t] - \sum\limits_{m=1}^{M} h_m \delta\left(c_m[t]\right)\right)^2}{2\sigma^2} \right\} \tag{9}$$

to find the combination of all n-1 input edges satisfying the CN constraint, including q^{n-2} terms in total. Since there are q types of constraints, the total number of multiplications required is $(n-2)q^{n-1}$. When the DFT-based algorithm is adopted, a DFT operation is carried out over the probabilities of input edge, and an IDFT operation is carried out for those of output edge, which requires $2q^2$ multiplications for each edge. In addition, n-2 multiplications of DFT vectors are required, which lead to $(n-2)4q$ multiplications. The total required multiplications is decreased to $2nq^2+4(n-2)q$, which is much smaller than that without the DFT-based algorithm.

D. Achievable Rate Region of LPNC-MA

Before investigating the optimized design of the q-ary IRA coded LPNC-MA, here we present its achievable rate region that characterizes the performance limit.

Theorem 1: Consider prime q. Let W_m and X_m be the uncoded message and transmitted symbol of user m respectively, and Y be the received symbol. U_l denotes the l-th NC message where $l=1,2,\cdots,L$. Consider coded modulation method and Gaussian MAC, for a given coefficient matrix $\mathbf{A}$ and a certain modulation, the achievable rate region of M-user LPNC-MA is characterized by [38], [39]

$$\begin{aligned} R_m \leq & H(X_m)-\max\{\varphi(a_{1m})H(U_1|Y),\cdots, \\ & \varphi(a_{Lm})H(U_L|Y,U_1,\cdots,U_{L-1})\}, \end{aligned} \tag{11}$$

where $m=1,\cdots,M$,

$$\varphi(a)=\begin{cases} 0, & a=0 \\ 1, & a\neq 0 \end{cases},$$

and $H(\bullet)$ denotes the entropy function.

Proof. Recall that W_m is uniformly distributed from $\mathbb{Z}_q$, so $H(W_m)=\log(q)$. With coded modulation method where bijection relationship holds between $\mathbb{Z}_q$ and X_m, X_m also has uniform distribution. Therefore, $H(X_m)=H(W_m)=\log(q)$. According to (4), the probability of $U_l=i$ is denoted by

$$\begin{aligned} p(U_l=i) &= \sum_{a_{l1}j_1\oplus\cdots\oplus a_{lM}j_M=i}\prod_{m=1}^{M}p(W_m=j_m) \\ &= q^{M-1}\cdot\frac{1}{q^M}=\frac{1}{q}. \end{aligned}$$

Therefore, $H(U_l)=H(X_m)=\log(q)$. For given coefficient matrix $\mathbf{A}$, the achievable computation rate of the

l-th ($l > 1$) NC message is given by [39]

$$R_l^N \leq I(Y; U_l | U_1, \cdots, U_{l-1}),\ l = 2, \cdots, L,$$

where $I(\bullet)$ denotes the mutual information function and $R_1^N \leq I(Y; U_l)$. Let $\mathcal{R}^N$ be the collection of all L achievable computation rate, which is defined as

$$\mathcal{R}^N \triangleq \bigcup_{l=1,\cdots,L} \{R_l^N\}.$$

If $a_{lm} \neq 0$, the l-th NC message U_l includes the message W_m, which means that the rate of user m is no greater than the computation rate of the l-th NC message, written as

$$R_m \leq R_l^N.$$

For user m, we define $\mathcal{R}_m^N \triangleq \bigcup_{a_{lm} \neq 0} \{R_l^N\}$, which collects all the computation rates that constraint user m's rate R_m. Note that $\mathcal{R}_m^N \subseteq \mathcal{R}^N$. The rate of user m is bounded by

$$\begin{aligned}
R_m &\leq \min\{R_l^N | R_l^N \in \mathcal{R}_m^N\} \\
&\leq \min\{I(Y; U_l | U_1, \cdots, U_{l-1}) | a_{lm} \neq 0\} \\
&= \min\{H\left(U_l | U_1, \cdots, U_{l-1}\right) - H\left(U_l | Y, U_1, \cdots, U_{l-1}\right) | a_{lm} \neq 0\} \\
&\overset{(a)}{=} H(U_l) - \max\{H\left(U_l | Y, U_1, \cdots, U_{l-1}\right) | a_{lm} \neq 0\} \\
&\overset{(b)}{=} H(X_m) - \max\{H\left(U_l | Y, U_1, \cdots, U_{l-1}\right) | a_{lm} \neq 0\} \\
&= H(X_m) - \max\{\varphi(a_{1m}) H(U_1 | Y), \cdots, \varphi(a_{Lm}) H(U_L | Y, U_1, \cdots, U_{L-1})\},
\end{aligned}$$

where step (a) follows from the fact that $U_1, \cdots, U_L$ are independently uniform distributed, step (b) follows from the fact that X_m and U_l are independently uniform distributed. ■

Remark 2: From (11), user m's rate R_m is bounded by the minimum achievable computation rate whose corresponding NC message involves user m's message. When M users' rates satisfy (11), M NC messages can be correctly computed. Therefore, for given $\mathbf{A}$, all users' messages are able to be correctly recovered.

In [24], it is shown that by varying the NC coefficient matrix, the shaping lattice and the scaling factor, other rate-pairs can be achieved. In this paper, we consider q-PAM, where the treatment in [24] can be realized by introducing non-uniformly spaced q-PAM constellations [40], as will be detailed next.

Here we give a toy example to briefly explain the non-uniformly spaced constellation (NUC). Consider $q = 5$. Let the ratio of the origin-middle point distance to the origin-edge point distance be $\rho (0 < \rho < 1)$.

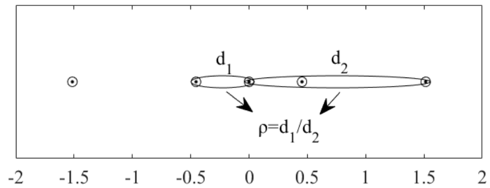


Fig. 3. An example for non-uniform constellations of q-PAM when $q = 5$.

Then, the constellation is $\left\{-\frac{1}{\gamma}, -\frac{\rho}{\gamma}, 0, -\frac{\rho}{\gamma}, -\frac{1}{\gamma}\right\}$, where γ is to normalize the average symbol energy. Let $\delta(i)$ denote i-th element of the constellation, thus the modulated symbols are

$$\mathbf{x} = \delta\left(\mathbf{c}\right). \tag{12}$$

Note that for a lager $q > 5$, the NUC needs multi-letters to describe.

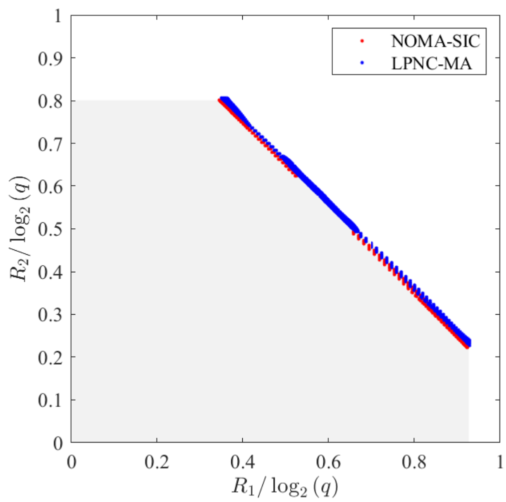


Fig. 4. An example of the achievable rate region of a $M = 2$ Gaussian MAC.

Consider $M = 2$, with $[\sqrt{P_1}, \sqrt{P_2}] = [1.3777, 1]$ and the noise variance $\sigma^2 = 0.06454$. Consider $\mathbf{A} = [1\ 1; 1\ 0]$. Then (11) is written as

$$\begin{aligned} R_1 &\leq H(X_1) - \max\left\{H(U_1|Y), H(U_2|Y, U_1)\right\}, \\ R_2 &\leq H(X_2) - H(U_1|Y). \end{aligned} \tag{13}$$

The rate pair $(\frac{1}{2}, \frac{2}{3})$ on the DF is achieved. By swapping the role of users 1 and 2 and let $\mathbf{A} = [1\ 1; 0\ 1]$, the rate pair $(\frac{2}{3}, \frac{1}{2})$ is achieved.

With non-uniformly spaced constellations for all ρ, the achievable rate-tuples are shown in Fig. 4. It is shown that LPNC-MA achieves all rate-tuples on the DF. In contrast, SIC with all choices of non-uniformly spaced constellation can only achieve those rate-tuples that are close to the corners of the DF.

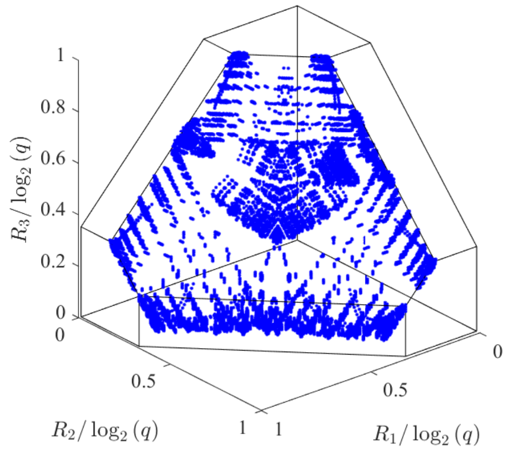


Fig. 5. Achievable rate region for $M = 3$ deterministic MAC. Here we only display the result where the step size of ρ is 0.05 for the limit of our evaluation time. Utilizing more accurate step size, the whole dominant face can be achieved.

Fig. 5 shows the achievable rate region of LPNC-MA for $M = 3$ Gaussian MAC where $[\sqrt{P_1}, \sqrt{P_2}, \sqrt{P_3}] = [1.4537, 1.3479, 1]$ and the noise variance is $\sigma^2 = 0.03791$, where the step size of ρ is 0.05. By utilizing smaller step size, the whole dominant face can be achieved.

E. Joint Nested Code Constructions and Degree Distribution Optimization

In order to achieve all the points on the DF, users need to transmit different rates, which leads to different generating matrices among users. Nested linear (or lattice) codes hold can be designed to ensure that the linear combination with integer coefficients of multiple codewords belong to the expanded codebook (or the same fine lattice) [41]. Then, the NC messages can be computed via single-user decoding.

In this part, we present a new design method of generating jointly nested and optimized IRA codes, aiming at approaching the rate-region limit presented above. Consider $M = 2$. Let (R_1, R_2) be the rates of the two users and assume $R_1 < R_2$. Now we need to design two IRA codes $\mathcal{C}_1, \mathcal{C}_2$ that satisfy: 1) $\mathcal{C}_1 \subseteq \mathcal{C}_2$; 2) $\mathcal{C}_1$ and $\mathcal{C}_2$ have "good" code degree profiles for the convergence of BP algorithm in decoding.

The Tanner graph of our nested IRA codes construction is shown in Fig. 6, consisting of three steps:

Step 1) Code degree profile optimization for $\mathcal{C}_1$: We first optimize $\mathcal{C}_1$ which is the same as the optimization for the single NC message case [19]. The degree profiles of the component node of q-ary IRA codes is obtained by using the EXIT chart curve fitting via linear programming [2] [19], [42], [43]. The nodes of $\mathcal{C}_1$ are depicted with solid black circles and squares, and the repeat nodes (RNs) are connected to the check nodes (CNs) through interleaver 1 with edges denoted by solid lines. With the optimized degree profile w.r.t. $\mathcal{C}_1$, the edges of these nodes are determined.

[2]To avoid redundancy, the details on EXIT chart curving fitting design is not shown in this paper, which can be found in [19].

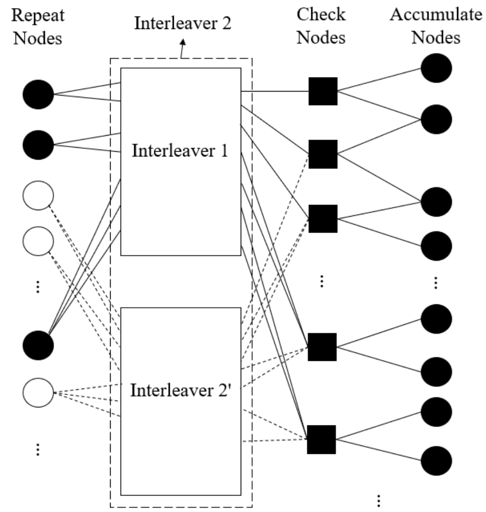


Fig. 6. Tanner graph of the nested IRA codes for the two-user case.

Step 2) $\mathcal{C}_1$'s degree profile as a constraint to optimize $\mathcal{C}_2$'s degree profile: We next consider the optimization of $\mathcal{C}_2$. Recall that $R_2 > R_1$, thus user 2 has a longer message sequence, i.e., $k_2 > k_1$. There are $k_2 - k_1$ zero-padded positions for the user 1's message sequence, which are depicted with the hollow circles. Since the k_1 solid black circles have been determined in Step 1, we still need to optimize the degree profile of the $k_2 - k_1$ hollow circles. Note that the edges w.r.t. the nodes of $\mathcal{C}_1$ impose an extra constraint to the optimization of degree distribution of $\mathcal{C}_2$. Consider RNs of degree 2. It requires that the number of degree-2 RNs in $\mathcal{C}_2$ should be no less than that in $\mathcal{C}_1$, translated into a lower limit on the proportion of degree 2 when optimizing $\mathcal{C}_2$'s code degree distribution. For example, consider $k_1 = 5000$ and $k_2 = 10000$. If $\mathcal{C}_1$ has 100 RNs of degree 2, $\mathcal{C}_2$ should have at least 100 RNs of degree 2, which imposes a "1%" lower limit on degree 2 when optimizing $\mathcal{C}_2$. In general, all RNs with different degrees need to satisfy such constraints, which are included in the EXIT chart curve fitting, obtaining the $\mathcal{C}_2$'s optimized code degree distribution.

Step 3) Interleaver design and nested construction: Now we obtain the optimized degree profiles of $\mathcal{C}_1$ and $\mathcal{C}_2$. For each degree, the number of RNs in $\mathcal{C}_2$ is more than that in $\mathcal{C}_1$, where the extra RNs are those hollow circles in Fig. 6 and are the positions adding zero for user 1. There are $k_2 - k_1$ such zero padded positions in total. However, if interleaver 2 is random, the connections of all lines (no matter solid or dashed lines) are random, thus the degree distribution of CNs in $\mathcal{C}_1$ will change randomly when adding zero in the positions of those hollow circles for user 1. In other words, optimized $\mathcal{C}_1$ and $\mathcal{C}_2$ are no longer nested when interleaver 2 is random. Therefore, interleaver 2 needs to be designed to ensure that the optimized $\mathcal{C}_1$ and $\mathcal{C}_2$ are nested. Our method for this is that the connecting relation (solid lines) of RNs and CNs in $\mathcal{C}_1$ is random through interleaver 1, while it is fixed for $\mathcal{C}_2$. The rest of the lines (the dashed lines) in $\mathcal{C}_2$ are randomly connected through interleaver 2'. This design of interleaver 2 ensures the optimized code degree distribution of $\mathcal{C}_1$ even though user 1 with zero padding utilizes $\mathcal{C}_2$ as the channel code. This completes the joint nested construction and optimization of $\mathcal{C}_1$ and $\mathcal{C}_2$.

Remark 3: The interleavers 1 and $2'$ are randomly selected while interleaver 2 is not random in overall. This may lead to a performance degradation for $\mathcal{C}_2$ relative to the single-user case. In order to reduce the degradation, we carry out the followings: 1) restrict the proportion of check nodes with degree equal to 1 from 2% to 40% [3]; 2) avoid adding three or more lines on the check nodes. Our EXIT chart and simulation results show the performance loss is fairly small.

Here we briefly explain how the above code construction and optimization enables $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Let the generator matrices of $\mathcal{C}_1$ and $\mathcal{C}_2$ be denoted by $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively. The codewords of user 1 and user 2 are $\mathbf{c}_1 = \mathbf{G}_1 \otimes \mathbf{w}_1'$ and $\mathbf{c}_2 = \mathbf{G}_2 \otimes \mathbf{w}_2$. Also, the construction method ensures that $\mathbf{c}_1 = \mathbf{G}_2 \otimes \mathbf{w}_1$, where $\mathbf{w}_1$ denotes the zero padded version of $\mathbf{w}_1'$ (add zeros in the positions of the hollow circles shown in Fig. 6). Define $\mathcal{W}_1$ as the sequence set of $\mathbf{w}_1$, which contains all possible sequences of $\mathbf{w}_1$. In the same way, define $\mathcal{W}_2$ as the sequence set of $\mathbf{w}_2$. It is apparent that $\mathcal{W}_1$ is a subset of $\mathcal{W}_2$ ($\mathcal{W}_1 \subseteq \mathcal{W}_2$) as some deterministic positions of $\mathbf{w}_1$ are always zeros. Consequently, after multiplying the same matrix $\mathbf{G}_2$, $\mathcal{C}_1 \subseteq \mathcal{C}_2$ is obtained.

For $M > 2$, the constructions and optimizations are performed layer-by-layer. Let $(R_1, R_2, \cdots, R_M)$ be the transmitting rate tuple of M users and assume $R_1 < R_2 < \cdots < R_M$ without loss of generality. We aim to design M IRA codes $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_M$ such that $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \cdots \subseteq \mathcal{C}_M$ and they have code degree profiles that favor the convergence of BP algorithm in decoding. For the m-th $(m = 2, \cdots, M)$ layer, take the optimized degree distribution of $\mathcal{C}_{m-1}$ and fixed connections of $\mathcal{C}_{m-1}$'s RNs and CNs as constraints to optimize code degree distribution of $\mathcal{C}_m$ using EXIT chart curve fitting, until all M IRA codes are designed.

Table I presents a number of code construction and optimization results. Here we briefly explain the construction and optimization process for $M = 3$, as that for $M = 2$ is a subcase. Assume $q = 5$. Given that $[\sqrt{P_1}, \sqrt{P_2}, \sqrt{P_3}] = [1.5425, 1.46255, 1]$ and $\sigma^2 = 0.11605$, the target code rates are $R_1 = 0.2238, R_2 = 0.4762, R_3 = 0.5000$. The process is as following:

1) Optimize the code degree distribution of q-ary IRA codes with $R_1 = 0.2238$, using the EXIT chart curve fitting method, named as $\mathcal{C}_1$;
2) Then $\mathcal{C}_1$'s degree profile serves as a constraint to optimize $\mathcal{C}_2$'s degree profile;
3) Design $\mathcal{C}_2$'s interleaver to guarantee the nested relation between $\mathcal{C}_1$ and $\mathcal{C}_2$;
4) $\mathcal{C}_2$'s degree profile serves as a constraint to optimize $\mathcal{C}_3$'s degree profile;

[3]We use nonsystematic IRA codes in this work, where check nodes need degree 1 to enable a successful start to the iterative decoding [43].

TABLE I
DEGREE DISTRIBUTION OF NESTED IRA CODES

	Coefficients	Code Rate	Check Node Degree	Information Node Degree
$M = 2$, $q = 5$	$[\sqrt{P_1}, \sqrt{P_2}] = [1.3777, 1]$, $\sigma^2 = 0.06454$	$R_1 = 0.5000$	$0.0571x + 0.9429x^2$	$0.2721x^2 + 0.1474x^3 + 0.1623x^5 + 0.0621x^6 + 0.0933x^7 + 0.1424x^{11} + 0.1205x^{12}$
		$R_2 = 0.6667$	$0.0413x + 0.2511x^2 + 0.4663x^3 + 0.2413x^4$	$0.2427x^2 + 0.1938x^3 + 0.1174x^5 + 0.0449x^6 + 0.0675x^7 + 0.0266x^8 + 0.1030x^{11} + 0.1073x^{12} + 0.0968x^{14}$
$M = 2$, $q = 5$	$[\sqrt{P_1}, \sqrt{P_2}] = [1.3777, 1]$, $\sigma^2 = 0.06454$	$R_1 = 0.5798$	$0.0216x + 0.5856x^2 + 0.3928x^5$	$0.1714x^2 + 0.1694x^3 + 0.2633x^5 + 0.0499x^6 + 0.1805x^{12} + 0.1655x^{15}$
		$R_2 = 0.5798$	$0.0216x + 0.5856x^2 + 0.3928x^5$	$0.1714x^2 + 0.1694x^3 + 0.2633x^5 + 0.0499x^6 + 0.1805x^{12} + 0.1655x^{15}$
$M = 3$, $q = 5$	$[\sqrt{P_1}, \sqrt{P_2}, \sqrt{P_3}] = [1.5425, 1.46255, 1]$, $\sigma^2 = 0.11605$	$R_1 = 0.2238$	$0.3612x + 0.3864x^2 + 0.2082x^3 + 0.0442x^5$	$0.0824x^2 + 0.0948x^3 + 0.1332x^5 + 0.1066x^8 + 0.0818x^{10} + 0.4647x^{24} + 0.0365x^{29}$
		$R_2 = 0.4762$	$0.0400x + 0.2982x^2 + 0.5776x^3 + 0.0843x^5$	$0.1244x^2 + 0.1955x^3 + 0.0899x^5 + 0.1585x^8 + 0.0488x^9 + 0.0583x^{10} + 0.2939x^{24} + 0.0305x^{29}$
		$R_3 = 0.5000$	$0.0378x + 0.2800x^2 + 0.3893x^3 + 0.2095x^4 + 0.0834x^5$	$0.1277x^2 + 0.1897x^3 + 0.0878x^5 + 0.1540x^8 + 0.0481x^9 + 0.0573x^{10} + 0.2847x^{24} + 0.0304x^{29} + 0.0201x^{50}$

5) Design $\mathcal{C}_3$'s interleaver to guarantee the nested relation between $\mathcal{C}_2$ and $\mathcal{C}_3$;

This finishes the process. Their error-rate performance will be shown in Section IV.

III. LPNC-MA for Fading MAC

In this section, we study efficient communication over M-user fading MACs. Block fading is considered where the coefficients remain constant in each coding block and vary over blocks. The users have the same statistical channel characteristics and signal-to-noise ratio. The symmetric coding rate is denoted by R. The receiver is assumed to have perfect estimation of the CSI, while the users do not know the CSI. Equal power and equal target rates among the users are considered.

A. *Encoding and Decoding*

Each user adopts the same channel coding and modulation with the identical rate R. The encoding operations are the same as those depicted for Gaussian MAC, but no efforts are required for code nesting[4].

The receiver carries out the SCD depicted in the previous section. The key difference to Gaussian MAC is that, for each fading channel realization, the NC coefficient matrix $\mathbf{A}$ needs to be carefully selected such that the error-rate performance is optimized. This is a challenging task when taking into account practical q-ary

[4]The optimization of node degree distribution of q-ary IRA codes is based on q-PAM modulation for single-user AWGN channel.

IRA codes, q-PAM/QAM modulation with a certain range of coding rates. Note that the selection of $\mathbf{A}$ needs to be done once for each block of n symbols.

B. Achievable Symmetric Rate

According to Theorem 1 and [38], the symmetric rate R is given by

$$\begin{aligned} &R < I(U_1; Y), \\ &R < I(U_2; Y|U_1), \\ &\quad \vdots \\ &R < I(U_M; Y|U_1, \cdots, U_{M-1}), \end{aligned} \tag{14}$$

where $I(\bullet)$ denotes mutual information (MI) function. To simplify the representation, the MIs are represented by $I_1, I_2, \cdots, I_M$, then R can be expressed as

$$R < \min\left\{I_1, I_2, \cdots, I_M\right\}. \tag{15}$$

Different network coded coefficients matrices yield different symmetric rates. A key issue is to find $\mathbf{A}$ that maximizes (15), that is

$$\mathbf{A}_{opt} = \max_{\text{Rank}(\mathbf{A})=M} \min\left\{I_1, I_2, \cdots, I_M\right\}, \tag{16}$$

where the entries of $\mathbf{A}$ belong to $\{0, \cdots, q-1\}$.

According to the chain rule of MI, the sum of the MIs is equal to a constant value

$$I_1 + I_2 + \cdots + I_M = I(U_1, U_2, \cdots, U_M; Y) = I(W_1, W_2, \cdots, W_M; Y) \triangleq I_{sum}. \tag{17}$$

Here, $\mathbf{A}$ is of full rank, such that U and W can be obtained from each other with a given $\mathbf{A}$.

C. A Pragmatic Solution to The Optimized NC Coefficient Matrix

The matrix $\mathbf{A}$ has M^2 elements. A brute-force search of all $\mathbf{A}$ involves q^{M^2} possible candidates. The computational complexity increases exponentially with M^2. Here, we propose a pragmatic algorithm for selecting $\mathbf{A}$. This is a algorithm similar to greedy algorithm, which finds the coefficients layer by layer. Recall that the NC coefficient of the l-th NC message is $\mathbf{a}_l^T = [a_{l1}, \cdots, a_{lM}]$. For the first layer, select $\mathbf{a}_1^T$ by

$$\mathbf{a}_1^T = \max_{\mathbf{a}_1^T \neq \mathbf{0}} \min\left\{I_1, \frac{I_{sum} - I_1}{M-1}\right\}. \tag{18}$$

For the l-th layer, select $\mathbf{a}_l^T$ by

$$\mathbf{a}_l^T = \max_{\text{Rank}([\mathbf{a}_1,\cdots,\mathbf{a}_l]^T)=l} \min \left\{ I_l, \frac{I_{sum} - \sum_{i=1}^{l} I_i}{M-l} \right\}. \tag{19}$$

The first term represents the MI w.r.t. the current NC message, and the second term represents the maximum value of the minimum MI of the subsequent $M-l$ NC messages.

Note that the coefficient matrix obtained by the above pragmatic algorithm is not necessarily the global optimal solution $\mathbf{A}_{opt}$. Yet it yields the near optimal solution as we will see from the numerical results momentarily. Next, we illustrate the complexity of the proposed algorithm.

Lemma 1: If the NC coefficient vectors of the first l NC messages can be written as $[\mathbf{a}_1,\cdots,\mathbf{a}_{l-1},\mathbf{a}_l] = [\mathbf{a}_1,\cdots,\mathbf{a}_{l-1},\sum_{i=1}^{l}\alpha_i\mathbf{a}_i]^T$ ($\alpha_i \in \{0,1,\cdots,q-1\}, i=1,2,\cdots,l, \alpha_l \neq 0$), the conditional entropies satisfy $H(Y|U_1,\cdots,U_{l-1},U_l) = H(Y|U_1,\cdots,U_{l-1},U_l')$ where $U_l' = (\sum_{i=1}^{l}\alpha_i\mathbf{a}_i)^T[W_1,W_2,\cdots,W_N]^T$.

Proof. The conditional entropy is written as

$$H(Y|U_1,\cdots,U_{l-1},U_l) = \frac{1}{q^l}\sum_{j_1,\cdots,j_l} H(Y|[U_1,\cdots,U_l]=[j_1,\cdots,j_l]). \tag{20}$$

Since U is the linear combination of user symbols W, we have

$$U_l' = (\sum_{i=1}^{l}\alpha_i\mathbf{a}_i)^T[W_1,W_2,\cdots,W_N]^T = \sum_{i=1}^{l}\alpha_i(\mathbf{a}_i{}^T[W_1,W_2,\cdots,W_N]^T) = \sum_{i=1}^{l}\alpha_i U_i. \tag{21}$$

Therefore, the constellation points corresponding to $[U_1,\cdots,U_l]=[j_1,\cdots,j_l]$ are the same as those corresponding to $[U_1,\cdots,U_l'] = [j_1,\cdots,\sum_{i=1}^{l}\alpha_i j_i]$. Condition $\alpha_l \neq 0$ ensures that when $[j_1,\cdots,j_l]$ takes different values, $[j_1,\cdots,\sum_{j=1}^{l}\alpha_i j_i]$ also has different values. So the final values of $H(Y|U_1,\cdots,U_{l-1},U_l)$ and $H(Y|U_1,\cdots,U_{l-1},U_l')$ are the same, although their summation order are different. ■

For the NC coefficient $\mathbf{a}_l^T$ of l-th NC message, different selections of $\sum_{i=1}^{l}\alpha_i\mathbf{a}_i^T$ ($\alpha_i \in \{0,1,\cdots,q-1\}, i = 1,2,\cdots,l,\alpha_l \neq 0$) have no difference in the value of I_l ($I_l = H(Y|U_1,\cdots,U_{l-1}) - H(Y|U_1,\cdots,U_l)$). Thus, we only need to select one of $q^{l-1}(q-1)$ choices in (19).

For the first NC message, there are q^M-1 possible combinations of non-zero coefficients. However, considering the cyclic invariance of modular q operation, there are only $\frac{q^M-1}{q-1}$ independent coefficient vectors. For the l-th NC message, considering the rank condition that the NC coefficient needs to meet, there are $(q^M - q^{l-1})$ possible combination coefficients (q^{l-1} represents the combination of the previous $l-1$ coefficients). Therefore, there are $\frac{q^M-q^{l-1}}{q^{l-1}(q-1)} = \frac{q^{M+1-l}-1}{q-1}$ independent coefficients to be calculated to obtain suitable $\mathbf{a}_l^T$.

The coefficients determined in previous layer do not change in subsequent layer, thus the complexity is

$$O(\sum_{l=1}^{M}\frac{q^{M+1-l}-1}{q-1})=O(\frac{q(q^M-1)-M(q-1)}{(q-1)^2})\approx O(q^M). \tag{22}$$

In contrast, the total choices in the brute-force search method are their product $\prod_{l=1}^{M}\frac{q^{M+1-l}-1}{q-1}$, whose complexity is $O(q^{\frac{M^2}{2}})$. For example, for a practical $q=5$ and $M=4$, the complexity can be saved by $(1-\frac{q^4}{q^8}=1-\frac{1}{q^4})$, which is over 99%.

D. Non-uniformly spaced Constellation

Similar to the Gaussian case, here we illustrate nonuniformly spaced constellation for the fading case. For the Gaussian MAC, one is free to choose the constellation for each rate-pair target. For fading MAC, one needs to find a certain constellation that performs well on average for all fading realizations.

Consider $M=2$. Assume that the equidistant q-PAM mapping is adopted, the channel gain $\mathbf{h}=[h_1,h_2]=[1.1,1]$, finite field size $q=5$ and the noise variance $\sigma^2=0.03162$ (SNR=15dB). If $\mathbf{a}_1^T=[1,1]$ is selected, then $I_1=0.9195\log_2(5)$ and $I_2=0.3360\log_2(5)$, this will result in that $\min\{I_1,I_2\}$ is dominated by I_2, so that the symmetric rate R is quite small. Therefore, the idea here is to balance I_1 and I_2 by selecting non-uniform constellations, which maximizes $\min\{I_1,I_2\}$.

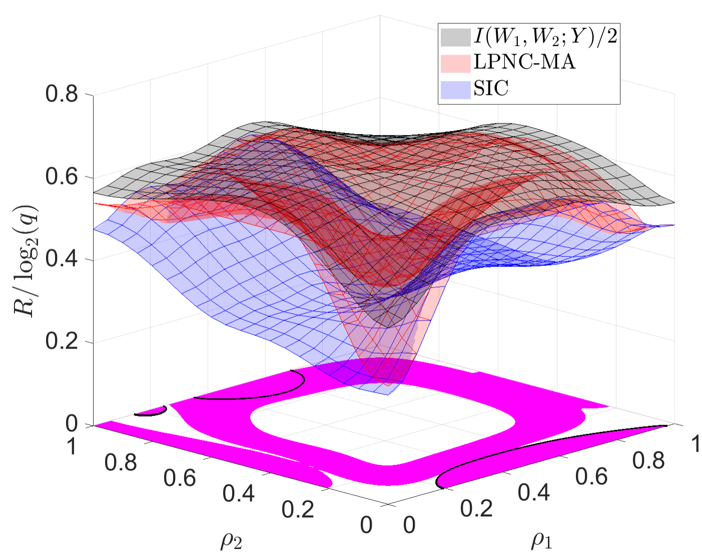


Fig. 7. Achievable symmetric rate R varies with the constellation parameters ρ of the two users.

Fig. 7 shows the achievable symmetric rates with different constellation parameters ρ of the two users. The gray surface represents the maximum symmetry rate of the MAC that can be achieved by joint ML decoding.

It can be seen that the maximum rate of LPNC-MA changes greatly, while the sum rate $I(W_1, W_2; Y)$ changes little. Thus, it is possible for $\min\{I_1, I_2\}$ to approach $I(W_1, W_2; Y)/2$ with a specific constellation parameter.

The parameter regions of LPNC-MA and SIC whose achievable rates exceed 90% of the ML decoding are drawn on the xOy axis plane separately. The purple region represents LPNC-MA and the region enclosed by black curve represents SIC. Note that the constellation parameter is determined in advance and used for all fading realizations. The channel coefficient varies randomly, and is assumed to follow a Rayleigh distribution. The realization of $\mathbf{h} = [1.1, 1]$ and $\mathbf{h} = [1, 1.1]$ are equally likely. The figure corresponding to $\mathbf{h} = [1, 1.1]$ is to exchange the x-axis and y-axis of Fig. 7. The projection areas of SIC will not coincide under the two channel gains. However, the projection areas of LPNC-MA scheme will coincide because its shape is similar to a ring. Here, we include Fig. 7 for a specific channel realization. This is just to provide an intuitive explanation on why LPNC-MA outperforms NOMA-SIC (for the fading setup), which helps with the understanding of the gain from LPNC-MA.

For the cases with $M > 2$, the constellation mapping is optimized by focusing on equal channel gain of each user to maximize $\max\limits_{\text{Rank}(\mathbf{A})=M} \min\{I_1, I_2, \cdots, I_M\}$. Note that the parameter ρ is continuous. Here we can discretize several different constellation mapping, which have different spacing characteristics as far as possible. Then the NC coefficient selection is combined with different constellation mapping for different users for calculating $\max\limits_{\text{Rank}(\mathbf{A})=M} \min\{I_1, I_2, \cdots, I_M\}$ to find the optimized mapping scheme.

E. Outage Probability and FER

Here we briefly illustrate the outage probability (OP) of LPNC-MA, which sets the lower bound on the FER of our designed q-ary IRA codes based LPNC-MA for fading MAC. For a target symmetric rate R_{set}, the OP of LPNC-MA is

$$p_{outage}(R) = \Pr(\min\{I_1, I_2, \cdots, I_M\}_{\mathbf{A}_{opt}} < R_{set}), \tag{23}$$

which is averaged over the p.d.f. of the fading channel coefficients. The OP of SIC is equivalent to LPNC-MA with $\mathbf{A}$ being a the permutation matrix[5].

The OP lower bound is given via the evaluation of the MAC capacity region, that is

$$Pr\left[mR < I\left(\{W_i\}_{i\in\Omega_m}; Y | \{W_j\}_{j\in\{M\}/\Omega_n}\right)\right], m = 1, 2, \cdots, M, \tag{24}$$

[5]The permutation matrix is determined by the order of SIC.

where $\{M\}$ denotes a collection of indexes from 1 to M and Ω_m denotes all combinations of m elements selected from $\{M\}$. With the optimized $\mathbf{A}$ and constellation parameter, we are ready to implement the q-ary IRA codes based LPNC-MA for fading MAC, and evaluate its practical FER performance.

IV. Numerical Results

A. Discussions for the assumptions of LPNC-MA

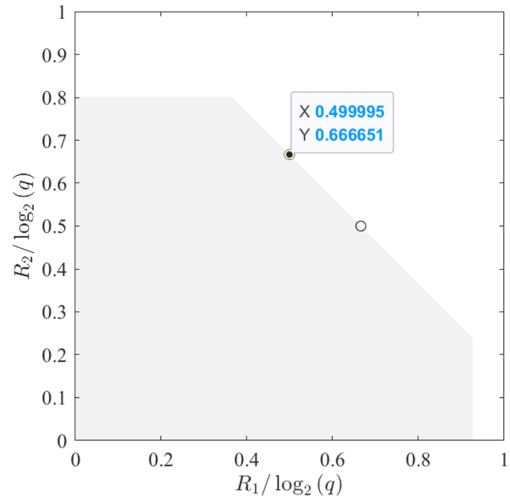
1) Assumptions and coordination for deterministic Gaussian MAC: Following the convention of NOMA, we assume that the receiver knows the number of users M in the coordination process before the transmission. We also assume that accurate CSI is obtained via the training process with sufficient preamble length. Based on the CSI, the receiver determines a coding strategy, from a discrete set of nested IRA codes candidates, PAM constellations and NC matrix $\mathbf{A}$, which meets the rate requirements of users. For the deterministic Gaussian MAC, we consider close-loop coordination, i.e., the choice of nested IRA codes and PAM constellations is reliably delivered to the users, in a fashion similar to adaptive modulation and coding (AMC).

2) Assumptions and coordination for fading MAC: For non-deterministic fading MAC setup, we again assume that the receiver knows M and has accurate CSI. Following the convention, we consider open-loop system for the fading MAC setup, where the feedback is unavailable thus AMC cannot be implemented. In this setup, the users have no indication about the channel conditions and permitted rates, so each user just simply transmits at a designated target rate for all fading blocks. In general, different users may have different target rates, but for clarity and simplicity we just consider that all users have identical target rates as in conventional treatment. Then, the core issue becomes identifying the optimal NC coefficient matrix $\mathbf{A}$, which has been discussed in Section III. C.

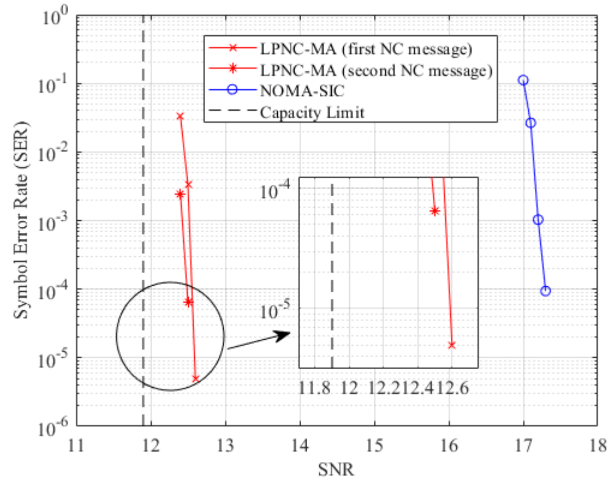
3) Discussion on the overheads: For Gaussian MAC, the extra overheads lie in a) finding NC matrix $\mathbf{A}$, b) selecting the nested IRA codes and c) non-uniformly spacing of the PAM constellations. Note that the nested IRA codes are pre-designed off-line, which provides a table containing a discrete set of choices of codes and PAM constellations. In the online MAC transmission process, the extra overhead for delivering the indices of the choice of coding and modulation is (almost) identical to that of conventional AMC.

For fading MAC where AMC is not conducted and thus coordination is not required, the extra overhead of the online MAC transmission process is due to identifying the optimal NC matrix $\mathbf{A}$, which is performed once per fading block. Using our developed greedy-algorithm based method, the cost for identifying the best $\mathbf{A}$ can be made reasonably small, i.e., negligible compared to the detection and decoding complexity. Such extra overhead is almost unnoticeable when the code block length becomes relatively large in practice.

B. Performance of LPNC-MA for Gaussian MAC



(a) Rate-pairs $(\frac{1}{2}, \frac{2}{3})$ and $(\frac{2}{3}, \frac{1}{2})$ on the dominant face.



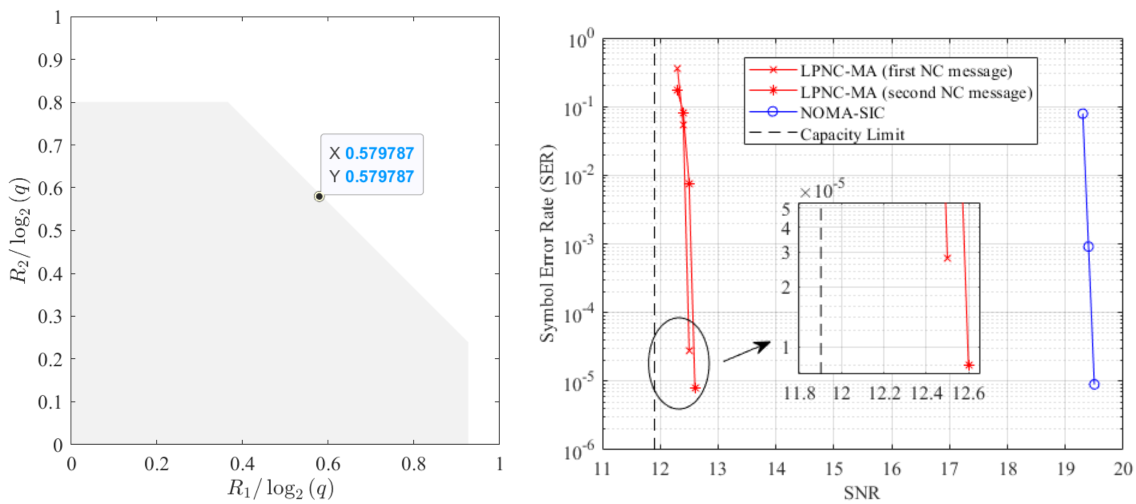
(b) Error-rate performance.

Fig. 8. Achievable rate-pair and error-rate performance for $M = 2$ with uniformly spaced q-PAM constellation, $q = 5$. The received powers are $[\sqrt{P_1}, \sqrt{P_2}] = [1.3777, 1]$. The SNR w.r.t. the capacity limit is at 11.9017dB.

1) Two-user Cases: First recall the $M = 2$ example previously shown in Fig. 4, with received powers of the two users given by $[\sqrt{P_1}, \sqrt{P_2}] = [1.3777, 1]$ and noise variance $\sigma^2 = 0.06454$. LPNC-MA was shown to achieve all the rate-pairs on the DF therein. With uniformly spaced q-PAM, $q = 5$, LPNC-MA achieves the rate-pairs $(\frac{1}{2}, \frac{2}{3})$ and $(\frac{2}{3}, \frac{1}{2})$ on the DF [6]. The corresponding NC coefficients matrix for rate-pair $(\frac{1}{2}, \frac{2}{3})$ is $\mathbf{A} = [1\ 1; 1\ 0]$, i.e., first computes the modulo-q sum and then uses it as side information to decode user 1's message. The nested IRA codes with q-PAM are used to characterize its practical performance. The optimized code degree distributions can be found in Table I. The block length k is 100,000, and the fidelity is that more than 100 frame errors are collected before the termination of the simulation for each simulation point. Fig. 8(b) shows the error-rate performance of LPNC-MA. In particular, the symbol error-rate (SER) of both the first and second NC messages are shown separately. Note that if the gap between these two is significant, the overall error-rate suffers as it is dominated by the worst of the two. Using our developed nested code construction and optimization, the gap between these two are minimized which yields the best overall performance. Under the same conditions, we also show the performance of NOMA-SIC in Fig. 8(b). Note that for the two-user case, NOMA-SIC decodes one user's message first and then cancels it from the received signal to decode the other user's message. Moreover, the performance of NOMA-SIC can be further improved by selecting the best ordering. In the simulation result in Fig. 8(b), we show NOMA-SIC with the best ordering, where the equivalent matrix $\mathbf{A} = [1\ 1; 0\ 1]$. We see that our designed LPNC-MA performs within 0.7dB the MAC capacity at SER of 10^{-4}, which overwhelmingly outperforms SIC by about 5 dB.

[6]The coding rates $(\frac{1}{2}, \frac{2}{3})$ are of practical interests and this is why this example is selected for illustration purpose.

Here, we give an intuitive explanation on the huge performance gap between LPNC-MA and NOMA-SIC. In Fig. 8(a), NOMA-SIC can achieve the corner rate-pair $(0.37, 0.8)$. However, the user 1's target rate is $\frac{1}{2}$, which is 0.13 bits more than NOMA-SIC's achievable rate on the set SNR. To achieve the target rate, a larger SNR is required for NOMA-SIC. This raises an enormous gap between the curve for decoding user 1's message and the capacity limit. On the other hand, the user 2's achievable rate of NOMA-SIC is higher than the target rate on the set SNR, thus NOMA-SIC can achieve the target rate at a lower SNR for the second decoding step. This causes a phenomenon in NOMA-SIC that the performance of the first decoding step is pretty bad, but that of the second is especially good. In this case, the overall performance of NOMA-SIC is pretty bad. In contrast, LPNC-MA has balanced performances in all decoding steps, due to its ability to achieve the whole dominant face. Next, we give another example to show that LPNC-MA can achieve all points on the dominant face with non-uniform constellations.

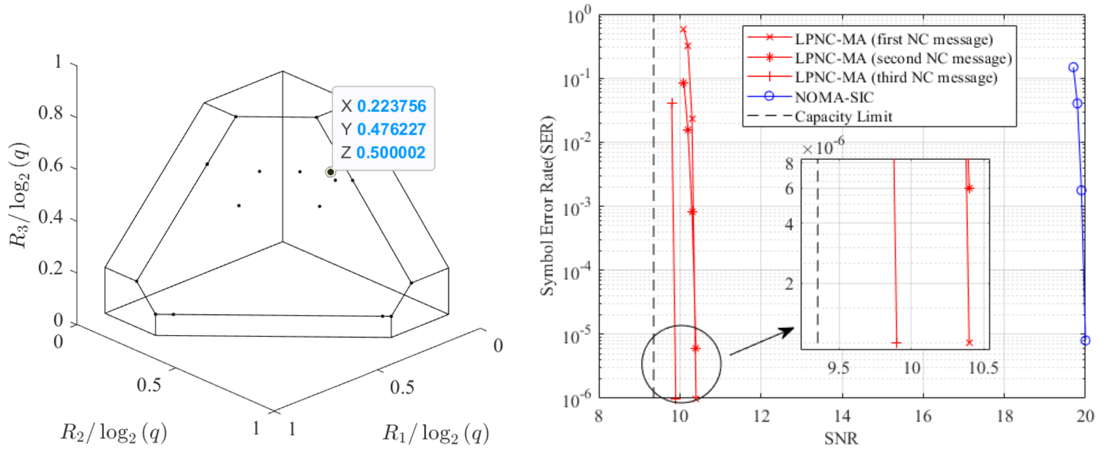


(a) Rate-pair $(\frac{7}{12}, \frac{7}{12})$ on the dominant face.

(b) Error-rate performance.

Fig. 9. Achieved rate-pair of $(\frac{7}{12}, \frac{7}{12})$ and error-rate performance. The channel parameters are the same as in Fig. 8.

We next consider another (symmetric) rate-pair $(\frac{7}{12}, \frac{7}{12})$ on the dominant face, as shown in Fig. 9(a). Here $\rho = 0.323$. The corresponding NC coefficients matrix is $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. Note that the purpose of this example in Fig. 9(a) with symmetric rate-pair is to show a different rate pair relative to that in Fig. 8(a) under the same channel parameter. In this particular example we consider symmetric rate. In such case, the code nesting can be avoided. It is noteworthy that non-uniformly spaced PAM constellation is generally required to achieve the symmetric rate-pair on the dominant face of MAC region. For fair comparison, NOMA-SIC utilizes the same modulation codes as the LPNC-MA scheme, with an optimal decoding order. Fig. 9(b) shows the error-rate performance of LPNC-MA and NOMA-SIC under this scenario. We see that LPNC-MA is only 0.65dB away from the capacity limit at SER of 10^{-4}, and outperforms NOMA-SIC by about 7dB. Clearly, SIC is not a good candidate in terms of symmetric rate, which will tremendously impact the FER performance in fading MAC as we will see in the next subsection.



(a) Rate-tuple $(0.2238, 0.4762, 0.5)$ on the dominant face.

(b) Error rate performace.

Fig. 10. Achieved rate-tuple of $(0.2238, 0.4762, 0.5)$ and error-rate performance for the three-user MAC. The received powers are $[\sqrt{P_1}, \sqrt{P_2}, \sqrt{P_3}] = [1.5425, 1.46255, 1]$, $q = 5$ and noise variance $\sigma^2 = 0.11605$. The capacity limit is at 9.3535dB.

2) Three-user Cases: Next we proceed to LPNC-MA for three-user MAC. Here we consider the powers $[\sqrt{P_1}, \sqrt{P_2}, \sqrt{P_3}] = [1.5425, 1.46255, 1]$ and noise variance $\sigma^2 = 0.11605$. With uniformly spaced q-PAM, $q = 5$, LPNC-MA achieves the rate-tuples on the DF shown in Fig. 10(a). Among those, we illustrate the rate-tuple $(0.2238, 0.4762, 0.5)$, with the corresponding NC coefficients matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{25}$$

For fair comparison, NOMA-SIC utilizes the same code as in LPNC-MA, with an optimal decoding order. Fig. 10(b) shows that the error-rate performance. Efforts are made to minimize the gap between the three curves w.r.t. the three NC messages using our optimized nested code construction. We see that LPNC-MA is 1.1dB away from the capacity limit at SER of 10^{-4}, and outperforms SIC by more than 9 dB.

In this section, we considered prime q. In most current standards, 2^m-PAMs are usually employed which calls for non-prime q. Recently we developed a doubly-irregular RA codes over integer rings that apply to any non-prime q, which can be used in LPNC-MA. For complex-valued model, LPNC-MA can be straightforwardly employed by sending two independent streams along the I and Q phases of q^2-QAM. A M-user complex-valued model is equivalent to a $2M$-user real-valued model, so the operations and design remain unchanged.

C. LPNC-MA for Fading MAC

This part shows the frame error rate (FER) performance of LPNC-MA for fading MAC. The CSI is assumed to be perfectly estimated at the receiver side, but not available to the transmitters. The M users are assumed to have the same target rate in the simulations, while it can be easily extended to non-identical target rates. We consider Rayleigh fading and block fading, where the channel coefficients remain constant within each coding block of length n and varies over blocks. For comparison, NOMA-SIC uses the same codes and modulations as used in corresponding LPNC-MA, with an optimally successive decoding order. The fidelity is that more than 500 frame errors are collected before termination of the simulation for each SNR point.

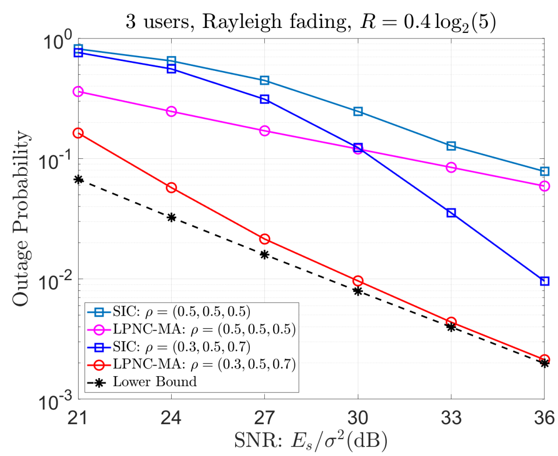


Fig. 11. Outage probability of LPNC-MA for three-user fading MAC.

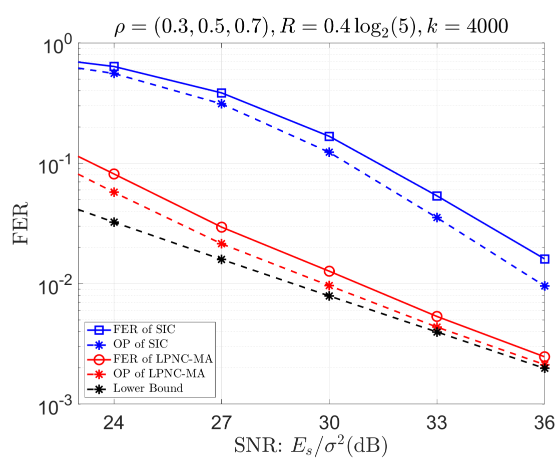


Fig. 12. FER of LPNC-MA for three-user fading MAC.

Consider symmetric target rate of $R = 0.4\log_2 5$ bits/symbol/user. We first evaluate the outage probability (OP) of LPNC-MA, that is, the probability that the target rate is out of the achievable rate-region (or greater than the symmetric rate) for the given channel realization. Then we proceed to simulate the FER of our developed LPNC-MA scheme, with block length $k = 4000$. Both uniformly spaced and non-uniformly spaced (with parameter $\rho = (0.3, 0.5, 0.7)$) constellations are considered. The OP of $M = 3$ users LPNC-MA is shown in Fig. 11. We see that at 10^{-2} (or 3×10^{-3}) , the OP of LPNC-MA with non-uniformly spaced constellation is only 0.8dB (or 0.3dB) from the lower bound of fading MAC. The FER performance is shown in Fig. 12. Comparing the curves of FER and outage probability of LPNC scheme, there is a gap of about 0.7dB. With the increase of code length, the gap may be further narrowed. Tremendously improved performance over NOMA-SIC is observed for both OP and FER. It can be seen that the outage probability can be used as a fairly accurate performance metric for the implementation and optimization of LPNC-MA.

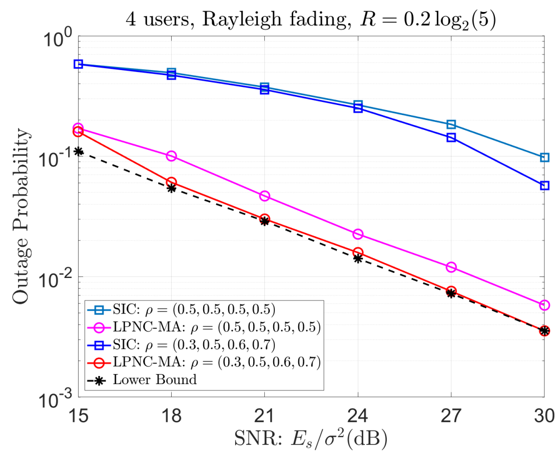


Fig. 13. Outage probability of LPNC-MA for four-user fading MAC.

The OP of LPNC with $M = 4$ is shown in Fig. 13. The target rate is $R = 0.2\log_2 5$ bits/symbol/user. The nonuniformly spaced constellations with parameters $\rho = (0.3, 0.5, 0.6, 0.7)$ is considered. It is seen that at OP of 0.0032, the OP of LPNC-MA scheme reaches the MAC lower bound. A greater improvement over NOMA-SIC is also observed.

D. Comparison with IDMA

1) Extending to A Large Number of Users: Our developed LPNC-MA and design method can be applied to MACs with any number of users. In practice, one simple way to extend to a large number of users would be to append a spreading code onto the LPNC-MA originally designed for three or four users. For example, each user may append a length-N Hadamard sequence [14]. Then the total number of users of LPNC-MA becomes MN, where each user's effective rate is reduced to $1/N$. However, this way of extending LPNC-MA to MN users, although simpler, may not compete with directly designing of LPNC-MA for MN in terms of rate-region and symmetric rate.

2) Complexity Comparison: The proposed scheme features lower complexity and less processing delay compared to existing code-domain NOMA system that utilizes big-loop receiver iterations. To see this, note that the computational complexity is primarily due to the multi-user detection and channel-code decoding. Let I_{sl} denotes the number of iterations of the sum-product algorithm of channel-code decoding, and I_{bl} denotes the number of big-loop receiver iterations. The total number of channel-code decoding is given by $M * I_{bl} * I_{sl}$, and the total number of detection is I_{bl}. Note that LPNC-MA needs $M * I_{sl}$ decoding and M detection, whereas the typical value of I_{bl} is usually 4 to 10 in existing code-domain NOMA such as iterative APP detection and decoding, IDMA and SCMA.

In particular, we provide a detailed complexity analysis in q-ary codes based LPNC and binary codes based IDMA, whose performances. IDMA has linear detection complexity $O(M)$ with interference Gaussian approximation, which requires more iterations to converge (e.g., 10 or even 20 iterations). LPNC-MA has $O(q^M)$ detection complexity, which is minor compared to decoding complexity when M is moderate in the LPNC framework considered in this work. The decoding complexity mainly comes from multiplications in check nodes. For each iteration, the complexity is $O(nq\log_2 q)$ [35], where n is the length of codewords. For fair comparison, binary IRA codes need $n\log_2 q$ codewords. Consider that LPNC-MA requires M serial computation with one q-ary IRA decoder and IDMA requires I_{bl} big-loop iterations with M parallel binary IRA decoders. Assume two types of decoders have the same I_{sl}, thus the complexity ratio is given by

$$r = \frac{Mnq\log_2 q}{2n\log_2 qMI_{bl}} = \frac{q}{2I_{bl}}. \tag{26}$$

For example, for $q = 5$ and $I_{bl} = 10$, the computation cost of LPNC-MA is only 1/4 of IDMA's.

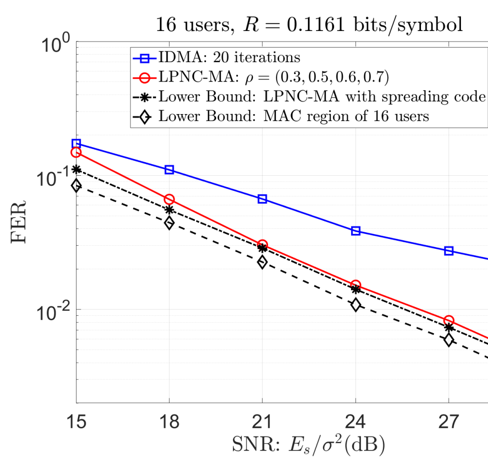


Fig. 14. Frame error rate of LPNC-MA and IDMA for 16-user fading MAC.

3) Performance Comparison: In Fig. 14, we present numerical results of a 16-user system where a 4-user LPNC-MA scheme is extended to 16-user using length-4 orthogonal Hadamard sequences. 16 users are divided into 4 groups, where in each group we utilize LPNC-MA together with a length-4 Hadamard sequence for spreading. The result shows that FER performance of LPNC-MA approaches its lower bound, as well as within 1.5dB gap to the Gaussian lower bound. Then, we compare the FER performance of LPNC-MA to that of IDMA. Here for the IDMA system, the repetition code has length-6 and each user employs a rate 0.6966 binary IRA code optimized for single-user AWGN channel. For LPNC-MA, the rate 0.2 5-ary IRA code (used in the previous example) and length-4 Hadamard spreading sequences are used. The sum rate is 1.8576 bits/symbol for both systems (as for a fair comparison). Note that this can be regarded as operating in a medium-to-high spectral efficiency. It is demonstrated that at FER lower than 10^{-2}, the LPNC-MA significantly outperforms IDMA. In this situation, the IDMA system operates with a load of 185%, which is very challenging for the iterative elementary signal estimator decoder to converge. Also note that big-loop iteration is avoided in the LPNC-MA system.

V. Conclusion and Future Works

A linear physical-layer network coding multiple access (LPNC-MA) scheme for efficient communication over Gaussian and fading MACs was studied. q-ary irregular repeat accumulate (IRA) codes over finite integer rings was utilized as the underlying coding-modulation. A practical sequential computation and decoding (SCD) algorithm was developed. A joint nested code construction and optimization method was developed. For a number of typical rate-tuples, the performance is within 0.7dB and 1.1 dB the capacity limits, for two and three users respectively. These near capacity-region performances were achieved with single-user decoding, without receiver-iterations and time-sharing. For fading MAC where the CSI is only available at the receiver, a pragmatic method for identifying the mutual information maximizing network coding coefficient matrix was presented. Numerical results demonstrated that the frame error rate of LPNC-MA is within a fraction of dB from the outage probability of the fading MAC, and remarkable improvement or complexity reduction over NOMA-SIC and other code-domain NOMA schemes is achieved. Open problems along this research direction includeFor multiple-antenna base station, how to exploit the notion of LPNC-MA to devise efficient detector and precoder for the uplink and downlink? How LPNC-MA will impact grant-free random-access in massive connectivity scenario and cell-free systems? We will look into these problems in the near future.

Acknowledgement

We would like to appreciate Dr. Jingge Zhu of Melbourne university for explaining the details of CFMA.

REFERENCES

[1] C. E. Shannon, “Two-way communication channels,” in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.*, vol. 1, pp. 611–644, The Regents of the University of California, 1961.

[2] R. Ahlswede, “Multi-way communication channels,” in *2nd Int. Symp. Inform. Theory, Tsahkadsor, SSR Armenia, 1971*, pp. 23–52, Publishing House of the Hungarian Academy of Science, 1973.

[3] T. M. Cover and J. A. Thomas, “Elements of information theory,” *John Wiley & Sons, Inc.*, 1991.

[4] W. Yu and J. M. Cioffi, “FDMA capacity of Gaussian multiple-access channels with ISI,” *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 102–111, 2002.

[5] D. D. Falconer, F. Adachi, and B. Gudmundson, “Time division multiple access methods for wireless personal communications,” *IEEE Commun. Mag.*, vol. 33, no. 1, pp. 50–57, 1995.

[6] S. Hara and R. Prasad, “Overview of multicarrier CDMA,” *IEEE Commun. Mag.*, vol. 35, no. 12, pp. 126–133, 1997.

[7] A. J. Viterbi, *CDMA: principles of spread spectrum communication*. Addison Wesley Longman Publishing Co., Inc., 1995.

[8] R. Prasad, *OFDM for wireless communications systems*. Artech House, 2004.

[9] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, “Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,” *IEEE Commun. Surv. Tuts.*, vol. 19, no. 2, pp. 721–742, 2016.

[10] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, “Application of non-orthogonal multiple access in lte and 5G networks,” *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, 2017.

[11] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, “Exploiting full/half-duplex user relaying in noma systems,” *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 560–575, 2018.

[12] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems,” *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1077–1091, 2017.

[13] 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, 1999.

[14] 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, 2006.

[15] H. Nikopour and H. Baligh, “Sparse code multiple access,” in *Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, pp. 332–336, 2013.

[16] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, “SCMA codebook design,” in *Proc. IEEE Veh. Tech. Conf*, pp. 1–5, 2014.

[17] T. Yang, J. Yuan, and Z. Shi, “Rate optimization for IDMA systems with iterative joint multi-user decoding,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1148–1153, Mar. 2009.

[18] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, “Pattern division multiple access–A novel nonorthogonal multiple access for fifth-generation radio networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, 2016.

[19] L. Yang, T. Yang, J. Yuan, and J. 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, 2015.

[20] L. Yang, T. Yang, Y. Xie, J. Yuan, and J. An, “Linear physical-layer network coding and information combining for the K-user fading multiple-access relay network,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5637–5650, 2016.

[21] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, “Compute-forward for DMCs: simultaneous decoding of multiple combinations,” *IEEE Trans. Inform. Theory*, vol. 66, no. 10, pp. 6242–6255, 2020.

[22] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, “Integer-forcing linear receivers,” *IEEE Trans. Inform. Theory*, vol. 60, no. 12, pp. 7661–7685, 2014.

[23] A. Sakzad, J. Harshan, and E. Viterbo, “Integer-forcing MIMO linear receivers based on lattice reduction,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4905–4915, 2013.

[24] J. Zhu and M. Gastpar, “Gaussian multiple access via compute-and-forward,” *IEEE Trans. Inform. Theory*, vol. 63, no. 5, pp. 2678–2695, 2017.

[25] T. Yang, L. Yang, Y. J. Guo, and J. Yuan, “A non-orthogonal multiple-access scheme using reliable physical-layer network coding and cascade-computation decoding,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1633–1645, 2017.

[26] E. Sula, J. Zhu, A. Pastore, S. H. Lim, and M. Gastpar, “Compute–forward multiple access (CFMA): Practical implementations,” *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1133–1147, 2019.

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

[28] N. S. Ferdinand, M. Nokleby, and B. Aazhang, “Low-density lattice codes for full-duplex relay channels,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2309–2321, 2014.

[29] G. Caire, G. Taricco, and E. Biglieri, “Bit-interleaved coded modulation,” *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[30] L. Ping, J. Tong, X. Yuan, and Q. Guo, “Superposition coded modulation and iterative linear MMSE detection,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 995–1004, 2009.

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

[32] Y. Cheng, L. Liu, and L. Ping, “Orthogonal AMP for massive access in channels with spatial and temporal correlations,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 726–740, 2021.

[33] J. J. Guo, T. Yang and 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, May 2015.

[34] M.-C. Chiu, “Bandwidth-efficient modulation codes based on nonbinary irregular repeat-accumulate codes,” *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 152–167, 2010.

[35] A. Bennatan and D. Burshtein, “Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 549–583, 2006.

[36] D. Sridhara and T. E. Fuja, “LDPC codes over rings for PSK modulation,” *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3209–3220, 2005.

[37] B. Nazer and M. Gastpar, “Compute-and-forward: Harnessing interference through structured codes,” *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.

[38] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, “A joint typicality approach to compute–forward,” *IEEE Trans. Inform. Theory*, vol. 64, no. 12, pp. 7657–7685, 2018.

[39] B. Nazer and M. Gastpar, “Computation over multiple-access channels,” *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3498–3516, 2007.

[40] M.-C. Chiu, “Bandwidth-efficient modulation codes based on nonbinary irregular repeat–accumulate codes,” *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 152–167, 2009.

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

[42] S. Ten Brink, “Convergence behavior of iteratively decoded parallel concatenated codes,” *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, 2001.

[43] S. ten Brink and G. Kramer, “Design of repeat-accumulate codes for iterative detection and decoding,” *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2764–2772, 2003.