IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 18, 2018, accepted October 3, 2018, date of publication October 11, 2018, date of current version November 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2875498

Rateless Multiple Access: Asymptotic Throughput Analysis and Improvement With Spatial Coupling

ZHAOYANG ZHANG[1], (Member, IEEE), XIANBIN WANG[1], YU ZHANG[2], (Member, IEEE), AND YAN CHEN[3], (Member, IEEE)

[1]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310058, China
[2]College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
[3]Huawei Technologies Co., Ltd., Shanghai 200121, China

Corresponding author: Zhaoyang Zhang (ning_ming@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61725104 and Grant 61631003, the National High-Tech R&D Project under Grant 2014AA01A702, and in part by Huawei Technologies Co., Ltd., under Grant HF2017010003, Grant YB2015040053, and Grant YB2013120029.

ABSTRACT A novel random code-domain non-orthogonal multiple access (NOMA) framework, rateless multiple access (RMA), is studied in this paper. In RMA, instead of granting each user specific radio resources in a fixed and centralized manner, the access point simply assigns to it a *random access control function*, according to which the user independently chooses a *pseudo-random* number of symbols each time and transmits their weighted sum until receiving an acknowledgment that indicates the successful recovery of its information. In this way, the transmission process of each individual user as well as that of all the users as a whole, resembles a special *linear superposition rateless encoder* and, thus, information can be retrieved with the low-complexity belief propagation (BP) algorithm at the access point. In this paper, the asymptotic throughput of RMA with BP decoding is analyzed. Then, in order to bridge the performance gap between BP decoding and the optimal MAP (maximum *a posteriori*) decoding, we apply the most recently developed coding technique of *spatial coupling* to RMA and propose an enhanced version, SC-RMA, in which all the user codewords are properly spatially-coupled across different subsets of REs. We prove that SC-RMA asymptotically approaches the channel sum-rate capacity with high adaptability and low-signaling overhead, which makes it a viable candidate for massive access.

INDEX TERMS Internet of Things (IoT), massive access, non-orthogonal multiple access.

I. INTRODUCTION

A. MOTIVATION

Massive access is one of the key technologies for future Internet of Things (IoT) which aims to provide reliable and adaptive connections for large number of devices over shared and dynamic resources [2]–[5]. Recently, non-orthogonal multiple access (NOMA) [6] has attracted a lot of research interest for its potential in enhancing the system capacity and meanwhile increasing the number of users by non-orthogonally sharing the resources among users. A typical NOMA scheme either exploits mainly the power domain and applies successive interference cancelation (SIC) to the received signal with differentiated user power [6], or exploits mainly the code domain and iteratively decodes the received signal by making use of certain codeword sparsity, such as low density signature multiple access (LDS) [7] and sparse code multiple access (SCMA) [8]. Due to its high spectrum efficiency, NOMA has been selected as the basic framework in 5G standardization.

However, several critical issues have yet to be solved in the design of massive access schemes. First, as the user number increases, centralized coordination becomes more difficult and inefficient. In particular, well-designed user pairing and accurate power control as usually required in conventional NOMA incurs considerable signalling and is not easy to realize in practice. Distributed coordination instead becomes more attractive in this sense. Second, the network becomes more dynamic with the increasing user number as well as the varying channel conditions in a wireless context. Fixed code/signature sequence design as in LDS, cannot adapt well to the network dynamics. High adaptability is thus of particular importance in the massive access protocol design.

Recently, random access, a technique for early computer networking, has been resurging due to its great advantages in signaling overhead and network adaptability [9]. Among all the possible approaches, the random access protocols incorporated with random coding and adaptive SIC have been of particular interests due to the increased achievable

throughput [10]. However, the spectrum efficiency of these protocols are still relatively low, since only single-user detection and decoding has been employed in these protocols.

In this work, we aim to establish a novel framework of random non-orthogonal massive access, namely, Rateless Multiple Access (RMA), which incorporates random access with the code-domain NOMA. In particular, in RMA, instead of granting each user specific resource elements (RE, i.e., a certain subcarrier-time slot pair), the access point (AP) assigns to it an *random access control function* (RACf) according to which each user independently and pseudo-randomly chooses a certain number of symbols for every RE and then transmits their linear combination over that RE. The transmission stops once the user receives an acknowledgement (ACK) from the AP indicating the successful recovery of its information. As a result, the transmission process of each individual user, as well as those of all the users as a whole, resembles a special *linear superposition rateless encoder* and BP algorithm can be used to retrieve the user information at the AP. Intuitively, RMA could work without specific RE allocation, which greatly reduces the signalling overhead. Moreover, due to the intrinsic rate adaptation property of rateless codes, RMA adapt well to the traffic and channel variations.

To further improve the performance, based on the above framework, we further incorporate RMA with the most recently developed idea of *spatial coupling* [11], a technique to improve the *saturation threshold* of iterative BP decoding or iterative detection [12], [13], and propose a novel Spatially-Coupled RMA (SC-RMA) protocol. SC-RMA retains the whole attractive features of RMA, and more importantly, it is proved to be capacity-approaching in an asymptotic sense.

B. RELATED WORKS

The original work of [14] studied in theory the achievability of gaussian multiple access channel capacity with rateless coding, which shows a new way for the design of multiple access protocols. To the best of our knowledge, the first series of attempts that incorporate rateless codes into practical multiple access protocol design were reported in [15]–[17]. Therein, instead of precoding the user information with a fixed-rate LDPC code as in IDMA [18], Raptor code [19] was employed to adapt the user transmission rate to the multiple access channel in which collision occurs unexpectedly since there lacks centralized coordination. Similarly, in [20], the authors introduced a frameless ALOHA that operates like a rateless code with simple repetitive encoding and iterative erasure decoding, and provided heuristic criteria to maximize the media access control (MAC) layer throughput. However, physical-layer properties of the wireless multiple access channel, such as linear superposition of collided signals and channel noise have not been fully considered therein. In [21], the authors proposed a similar idea in an analog fountain code framework. Therein empirical code degree and weight distribution are employed, the asymptotic achievable sum rates have not been analyzed, and the non-optimality of BP decoding and efficient spatial coupling of user code graphs have not been considered.

On the other hand, spatial coupling has been applied to random access protocols such as coded slotted ALOHA (CSA), to attain a performance close to the theoretical limit [22]. However, the physical-layer properties have not been fully considered there. In [12], the authors applied spatial coupling to sparsely spread CDMA to improve the detection performance. In [23] and [24], spatially coupling combined with two-stage decoding is proved in theory to achieve the channel capacity. Compared to the above schemes, we are not just applying another type of spreading codes (i.e., rateless codes vs. fixed-rate codes) which makes RMA adapt well to the traffic and channel variations and asymptotically achieve the sum-rate capacity, but its *rateless* nature brings a major conceptual shift with respect to the previous centralized spreading, which greatly reduces the signaling overhead especially when the number of users is large or the set of active users changes dynamically.

C. SUMMARY OF CONTRIBUTIONS

The contributions of this paper can be summarized as below:

- We establish a novel framework for random non-orthogonal massive access, namely RMA, which provides high spectrum efficiency with a relatively low signaling overhead and high adaptability. We derive the maximum achievable throughput under different signal-to-noise ratios (SNRs) and overloading factors based on the extrinsic information transfer (EXIT) theory [25], and develop the optimal degree distribution which provides some insights for the practical design of RACfs.
- We incorporate the coding technique of spatial coupling to RMA and propose the SC-RMA scheme, which provides increased system spectrum efficiency and reduced decoding complexity while retaining all the other attractive properties of basic RMA. Furthermore, we derive the achievable throughput of SC-RMA based on EXIT theory and prove that it is able to approach the channel capacity asymptotically.
- For further enhancing the performance of RMA in practical applications, we optimize system parameters including the RACfs, the spatial coupling pattern, and the coding rates and degree profiles of LDPC precoders using a two-step design method, which provides a feasible way for real system design.

The rest of the paper is organized as follows. In Section II, the system model, access procedure and decoding algorithm are introduced. In Section III, the achievable rate region of RMA is analyzed using the EXIT theory. In Section IV, SC-RMA is introduced and is proved to be capacity-approaching. In Section V, the key parameters in RMA are optimized using a practical two-step design method. Extensive simulation results are presented in Section VI and conclusions are drawn in Section VII.

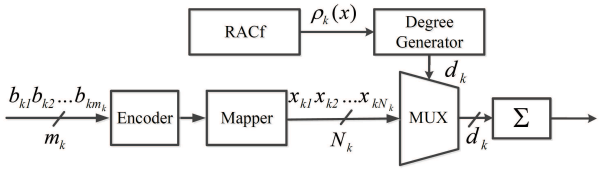

FIGURE 1. Block diagram of the user side in the RMA framework.

II. PROTOCOL AND ALGORITHM

Consider a cellular network with a number of potential users and an AP. All the users are assumed to be synchronized according to the down-link beacon. In case there still exist residue packet delays among users, such delays could be naturally compensated by the CP (cyclic prefix) if OFDM (Orthogonal Frequency Division Multiplexing) signalling is adopted. The users independently access the AP sporadically. At a given time, among all of them, K users, denoted by set $\mathscr{U}_A$ of $\{U_1, U_2, ..., U_K\}$, are *active*. As shown in Fig.1, during each access period, U_k $(k = 1, 2, ..., K)$ first encodes its m_k information bits with an LDPC encoder, and maps the resultant coded bits to a vector of symbols $\{x_{k,1}, x_{k,2}, ..., x_{k,N_k}\}$ with ± 1-valued indices, and then attempts to send them to the AP based on the access procedure as described below. The total length of information bits from all the active users is denoted by $m \triangleq \sum_{k=1}^{K} m_k$, and the total amount of resultant symbols is denoted by $N \triangleq \sum_{k=1}^{K} N_k$. At last, a quasi-static fading channel is considered, i.e., we assume that the channel gains from different users to the AP are independent and keep static during the transmission period of each packet.

A. ACCESS PROCEDURE OF RMA

Before starting the packet transmission, each active user accomplishes a certain *initialization/registration* process (please be referred to [9] for details), which enables the AP to obtain the current $\mathscr{U}_A$, the packet lengths and the corresponding channel gains, etc. Based on these parameters, the AP starts the procedure of data transmission by assigning each active user an RACf as follows:

$$\rho_k(x) = \sum_{d=0}^{d_{\max}} p_{k,d} x^d, \quad k = 1, 2, \ldots, K \tag{1}$$

where $p_{k,d}$ denotes the probability of User k to select d out of N_k symbols and $\sum_{d=0}^{d_{\max}} p_{k,d} = 1$. The RACfs can be viewed as some kind of *distribution profiles* and are used for the active users to share a block of REs in a distributive manner as described below.

As the transmission process starts, for every RE $t \in \{1, 2, \ldots, T\}$ where T denotes the total amount of REs, U_k first pseudo-randomly chooses a degree $d_{k,t}$ from 0 to $d_{\max}$ with probability $p_{k,d_{k,t}}$ according to its RACf. If $d_{k,t} = 0$, it transmits nothing over the RE. Otherwise, it uniformly selects $d_{k,t}$ symbols from $\{x_{k,1}, x_{k,2}, \ldots, x_{k,N_k}\}$, linearly combines them and then sends out the result over RE t. Denote the set of indices of the selected symbols as $\mathcal{V}(t,k) \subseteq \{1, 2, \ldots, N_k\}$. Thus the signal transmitted by U_k over RE t is

$$x'_{k,t} = \begin{cases} \sum_{n_k \in \mathcal{V}(t,k)} g_k x_{k,n_k} & d_{k,t} \neq 0 \\ 0 & d_{k,t} = 0, \end{cases} \tag{2}$$

in which g_k denotes the scaling factor to meet the power constraint. The signals transmitted over RE t from all the users are linearly added together in the air, and thus the signal received by the AP over RE t can be written as

$$y_t = \sum_{k=1}^{K} h_k x'_{k,t} + z_t = \sum_{k=1}^{K} \sum_{n_k \in \mathcal{V}(t,k)} h_k g_k x_{k,n_k} + z_t, \tag{3}$$

where z_t denotes the Gaussian noise at RE t with mean zero and variance σ_w^2. Since h_k and g_k have equivalent impacts on the received signal, we need not distinguish them, and hereafter we define $c_k = h_k g_k$ for brevity:

$$y_t = \sum_{k=1}^{K} \sum_{n_k \in \mathcal{V}(t,k)} c_k x_{k,n_k} + z_t. \tag{4}$$

At the AP, it consistently collects signals from all the available REs and keeps attempting to decode the user messages. Once a user message is successfully decoded, the AP feeds back to that user an ACK signal via the downlink to end its current transmission. Note that in this paper, we consider a half-duplex system in which the uplink and downlink transmissions are separated in time or frequency.

Remark 1: Obviously, the above transmission process resembles a special *linear superposition rateless encoder* (in contrast to the XOR-based conventional rateless encoders). On the one hand, the transmission of each user can be viewed as a *centralized* linear superposition rateless encoder, since each of its channel inputs is the *local* linear combination of a random number of randomly selected symbols. On the other hand, the transmissions of all active users as a whole can be viewed as a *distributed* linear superposition rateless encoder, since now the linear superposition is accomplished over the air. Due to such intrinsic features, the proposed framework is named RMA.

B. DECODING ALGORITHM OF RMA

Similar to Raptor codes [19], the transmission process of RMA can be elegantly represented by a single sparse factor graph, which brings the following two attractive features: (i) It enables the AP to retrieve the messages using a low-complexity BP algorithm; (ii) As elaborated in the following sections, it facilitates the design of the RMA schemes. Here we briefly describe the factor graph at first.

As shown in the graph representation of RMA in Fig.2, there are three kinds of nodes, including check nodes (CNs), variable nodes (VNs) and resource element nodes (RENs). In the precoding part, each subgraph within a rectangle represents the LDPC code applied at a user. At the bottom, each REN represents an RE. The edge between a VN and an

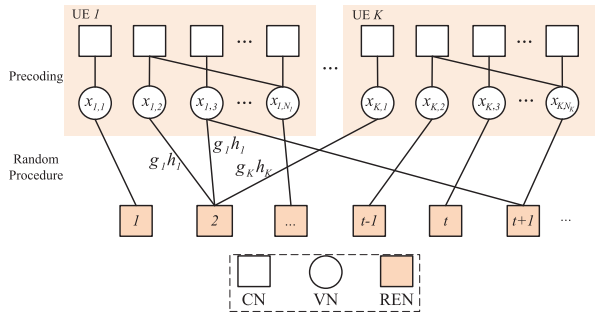

FIGURE 2. Graph representation of RMA.

REN means that the corresponding symbol is transmitted over that RE. The other nodes and edges are similarly defined as in Raptor codes [19]. As we have described that, at each time and for each active user, the number of symbols and their indices are *pseudo-randomly* generated with some predefined seed according to the RACf and the uniform distribution, respectively. Therefore, the AP has full knowledge about each user's graph structure once it is activated via the registration process, and hence is able to reconstruct the whole graph. In the case of no registration process available or in the so-called *grant-free* scenario, the graph structure can be recovered in conjunction with information decoding by exploiting the intrinsic and unique access pattern of each user, as elaborated in [28]. Here we only focus on the former case.

The AP then performs iterative BP algorithm based on which soft messages are exchanged on the graph iteratively to retrieve the original information bits. In this process, we need not distinguish which user a specific symbol is from, thus in the following we can drop the subscript k from x_{k,n_k} in (2) and denote them as $\{x_n, n = 1, 2, \ldots, N\}$ unless otherwise explicitly stated. The VN that represents x_n is referred to as VN n. Let $L^j_{n\leftarrow t}$ (resp. $L^j_{n\rightarrow t}$) denote the log-likelihood ratio (LLR) passed from REN t (resp. VN n) to VN n (resp. REN t) in iteration j. In particular, at the initial stage of the decoding algorithm, there is

$$L^0_{n\rightarrow t} = 0. \tag{5}$$

In the following, the iterative joint detection and decoding algorithm is described briefly.

First, denote the set of indices of VNs that are connected to REN t as $\mathcal{V}(t) \subseteq \{1, 2, \ldots, N\}$. For VN $n \in \mathcal{V}(t)$, the soft message from REN t to VN n can be calculated using the MAP (maximum a *posteriori* probability) criterion as follows:

$$\begin{aligned} L^{j+1}_{n\leftarrow t} &= \ln \frac{p(x_n = 1|y_t, \{L^j_{n'\rightarrow t}\}_{n'\in\mathcal{V}(t)\backslash n})}{p(x_n = -1|y_t, \{L^j_{n'\rightarrow t}\}_{n'\in\mathcal{V}(t)\backslash n})} \\ &= \ln \frac{p(y_t|x_n = 1, \{L^j_{n'\rightarrow t}\}_{n'\in\mathcal{V}(t)\backslash n})}{p(y_t|x_n = -1, \{L^j_{n'\rightarrow t}\}_{n'\in\mathcal{V}(t)\backslash n})} \end{aligned} \tag{6}$$

in which

$$\begin{aligned} p(y_t|x_n) &= \pm 1, \{L^j_{n'\rightarrow t}\}_{n'\in\mathcal{V}(t)\backslash n}) \\ &= \sum_{\{x_{n'}\}_{n'\in\mathcal{V}(t)\backslash n}} \prod_{n'\in\mathcal{V}(t)\backslash n} p(x_{n'}) p(y_t|x_n \\ &\qquad = \pm 1, \{x_{n'}\}_{n'\in\mathcal{V}(t)\backslash n}) \end{aligned} \tag{7}$$

and

$$p(x_{n'}) = \begin{cases} \dfrac{\exp(L^j_{n'\rightarrow t})}{\exp(L^j_{n'\rightarrow t}) + 1} & x_{n'} = 1 \\ \dfrac{1}{\exp(L^j_{n'\rightarrow t}) + 1} & x_{n'} = -1. \end{cases} \tag{8}$$

Note that the calculation of $p(y_t|x_n = \pm 1, \{x_{n'}\}_{n'\in\mathcal{V}(t)\backslash n})$ is slightly different from that in the conventional XOR-based Raptor codes since linear superposition is involved here.

Second, the soft message that VN n gets from all the connected RENs can be expressed as

$$L^{j+1}_n = \sum_{t\in\mathcal{Q}(n)} L^{j+1}_{n\leftarrow t}, \tag{9}$$

in which $\mathcal{Q}(n) \subseteq \{1, 2, \ldots, T\}$ denotes the set of indices of RENs that are connected to VN n. Based on this, the LDPC coding part can be updated iteratively as in [19] and feeds back $L^{j+1}_{n\rightarrow t}$ to RENs to further improve the detection performance as shown in (6).

III. ACHIEVABLE THROUGHPUT OF RMA UNDER BP DECODING

The main object of this section is to derive the maximal achievable throughput of RMA under the low complexity BP decoding algorithm. Due to the random structure of the factor graph and the iterative manipulations applied herein, it is difficult to pursue an explicit result by convergence analysis. Here we resort to the EXIT theory [26] which provides an effective approach for analysis of the performance of iterative decoding algorithms.

A. DEFINITIONS AND NOTATIONS

For useful insight as well as tractability, we assume that all the active users have the same Quality of Service (QoS) requirement and the same LDPC code rate $R = \frac{m}{N}$. Moreover, by proper power control all the user signals arrive at the AP with nearly the same power, i.e., $c_k (k \in \{1, 2, \ldots, K\})$ defined in (4) are equal and thus can be denoted as c with subscript k dropped hereafter. In such a case, *overloading factor* β, which is defined as the ratio of the amount of symbols from all the users to the amount of REs, can be calculated by:

$$\beta \triangleq \frac{N}{T} = \frac{m}{RT}, \tag{10}$$

and, throughput Θ can be calculated by

$$\Theta \triangleq \frac{m}{T} = R\beta. \tag{11}$$

As to the transmission process, the total number of symbols transmitted on RE $t \in \{1, 2, \ldots, T\}$ is denoted as d_t and therefore, $d_t = \sum_{k=1}^{K} d_{k,t}$. Without loss of generality, we assume that the noise variance $\sigma_w^2 = 1$. In such a case, the average SNR over all the REs, denoted by γ, can be calculated by

$$\gamma \triangleq \frac{\sum_{k=1}^{K} \mathrm{E}(d_{k,t})(c_k)^2}{\sigma_w^2} = c^2 \mathrm{E}(d_t). \tag{12}$$

Moreover, we denote d_n as the degree of VN n, namely, $d_n = |\mathcal{Q}(n)|$. Note that $\mathcal{Q}(n)$, defined in (9), denotes the set of indices of RENs that are connected to VN n. As we have assumed that the symbols are uniformly selected during the access procedure, we have

$$d_n = \frac{\mathrm{E}(d_t)T}{N}, \quad \sum_{t \in \mathcal{Q}(n)} c^2 = \frac{c^2 \mathrm{E}(d_t)T}{N} = \frac{\gamma}{\beta}. \tag{13}$$

Consider the Gaussian channel $y = x + z$ in which x is the binary input and z is the Gaussian noise with distribution $\mathcal{N}(0, \sigma^2)$. In this case, the soft message can be modeled as log-likelihood ratio L_x which is Gaussian distributed as $\mathcal{N}(\frac{2}{\sigma^2}x, \frac{4}{\sigma^2})$. Furthermore, we define $\sigma_L = \frac{2}{\sigma}$, and therefore, the distribution of L_x can be expressed as:

$$L_x \sim \mathcal{N}(\frac{\sigma_L^2}{2}x, \sigma_L^2). \tag{14}$$

Based on this, the mutual information (MI) between L_x and x can be calculated by:

$$J(\sigma_L) \triangleq I(L_x; x) = 1 - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_L^2}} \exp\{-\frac{(\xi - \frac{\sigma_L^2}{2})^2}{2\sigma_L^2}\} \cdot \log_2(1 + e^{-\xi}) d\xi. \tag{15}$$

Note that $J(\sigma_L)$ is a monotonically increasing function of σ_L defined on $[0, +\infty)$, which means that the less the noise is, the more information about x we have. Specifically, we have $J(0) = 0$ and $J(+\infty) = 1$.

B. THEORETICAL ANALYSIS OF RMA

For ease of analysis, we consider the two-stage detection-and-then-decoding algorithm for RMA. In the first stage, the subgraph in the bottom of Fig.2 is iteratively updated according to (6) until the MI of VNs converge. In the second stage, the subgraphs of pre-coding part are iteratively updated to retrieve the original information bits. Now we start to derive the theoretical achievable throughput in this case.

We first analyze the first stage of the algorithm. The main object is to derive the accurate relationship between the evolution process MI of VNs and the specific degree distribution (for more information about degree distribution, please refer to [19] and the references therein). We denote the edge-degree distribution of RENs as

$$\Omega(x) = \sum_{i=1}^{l} \omega_i x^{k_i d_a - 1}, \tag{16}$$

where l denotes the number of **different** REN degrees in the factor graph (therefore, $k_{i_1} \neq k_{i_2}$ for $i_1 \neq i_2$), and ω_i denotes the probability that a randomly chosen edge is connected to an REN of degree $k_i d_a$ (here d_a is a given natural number), thus $\sum_{i=1}^{l} \omega_i = 1$. If $d_a \to \infty$, it is referred to as *the asymptotic case*. If there is an $i^* \in \{1, 2, \ldots, l\}$ satisfying $\omega_{i^*} = 1$, the edge degree distribution is then uniform. If both of them hold, it is referred to as *the uniform asymptotic case*.

Before stepping into more details, we first discuss the case $\omega_i = 1$, i.e., each RE is only accessed by one VN. In this case, we can easily verify that $L_n^j \sim \mathcal{N}(\frac{2\gamma}{\beta}x_n, \frac{4\gamma}{\beta})$, which means that the channel one user faces can be approximated as a Gaussian channel. In the following, it is revealed that such result also holds for the RMA in which each RE is accessed by a random number of VNs.

Lemma 1: After the multiuser detection, the channel that one user faces can be approximated as a Gaussian channel with certain SNR degradation, i.e.,

$$L_n^j \sim \mathcal{N}(\frac{2\eta^j\gamma}{\beta}x_n, \frac{4\eta^j\gamma}{\beta}). \tag{17}$$

The degradation factor η can be calculated by:

$$\eta^j = \sum_{i=1}^{l} \frac{\omega_i}{1 + \frac{\beta k_i d_a c^2}{\gamma} f(\eta^{j-1})}, \tag{18}$$

in which,

$$f(x) \triangleq \frac{\gamma}{\beta\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(1 - \tanh(x\frac{\gamma}{\beta} + \sqrt{x\frac{\gamma}{\beta}}\xi)\right)^2 \times \exp\{-\frac{\xi^2}{2}\} d\xi. \tag{19}$$

Proof: See Appendix A. ■

Corollary 1: In the uniform asymptotic case, (18) can be further simplified as:

$$\eta^j = \frac{\omega_1}{1 + \frac{\beta k_1 d_a c^2}{\gamma} f(\eta^{j-1})} \overset{(a)}{=} \frac{1}{1 + \beta f(\eta^{j-1})} \tag{20}$$

where (a) follows from (13).

In the following, we prove that degradation factor η^j is maximized only in the uniform asymptotic case and thus the uniform asymptotic case is optimal in terms of throughput.

Lemma 2: For the degree distributions of RENs, the maximization of η^j is achieved only when there is an $i^* \in \{1, 2, \ldots, l\}$ satisfying $\omega_{i^*} = 1$.

Proof: See Appendix B. ■

Thereby, to derive the achievable throughput in the asymptotic case, we only consider the uniform asymptotic case in the following. It can be easily seen from (18) that, there is $\eta^1 > \eta^0$ when $\eta^0 = 0$, and $\eta^{j+1} < \eta^j$ when $\eta^j = 1$. Invoking the monotonicity of $f(\eta)$, we can conclude that $\eta = \frac{1}{1+\beta f(\eta)}$ has at least one solution within (0, 1]. We denote the minimum solution as $\eta_{\beta,\min}$. As a result, invoking the

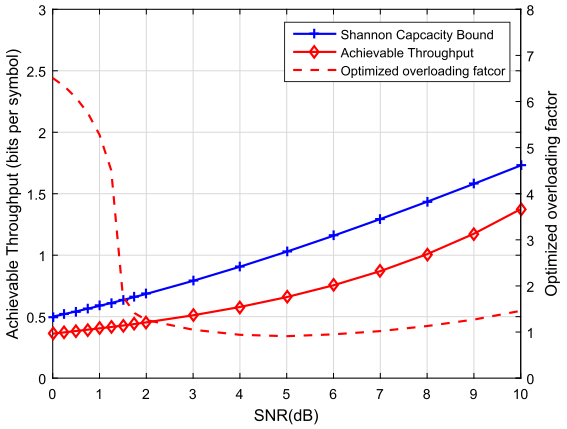

FIGURE 3. Achievable throughput with optimized overloading factor.

MI defined in (15), the converged MI of VNs may be expressed as

$$J(\sqrt{\frac{4\eta_{\beta,\min}\gamma}{\beta}}). \tag{21}$$

Then, we take the LDPC coding part into consideration, based on which, we derive the theoretical achievable throughput of RMA.

Theorem 1: In the RMA system with overloading factor β and SNR γ, the asymptotic achievable throughput, denoted by Θ_{ac}, can be calculated by

$$\Theta_{\text{ac}} = \beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_{\beta,\min}), \tag{22}$$

in which $C_{\text{BIAWGNC}}(x)$ denotes the capacity of the binary input Gaussian channel with SNR x.

Proof: Based on previous discussions, in the RMA system with overloading factor β and SNR γ, the channel that one user faces can be approximated as a Gaussian channel with SNR $\frac{\gamma}{\beta}\eta_{\beta,\min}$. In such a case, to guarantee the decodability of LDPC codes, the maximum coding rate may be expressed as $C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_{\beta,\min})$. Hence the conclusion follows by invoking (11). ■

Note that since the linear combination may happen among the coded symbols from both different active users and an individual user itself, the analysis result in this section is tenable regardless of the specific user number.

Remark 2: For the given γ, different β results in different $\eta_{\beta,\min}$ and thus different achievable throughput Θ_{ac}. For further enhancing the performance, we optimize β to maximize the asymptotic achievable throughput. Define β_{opt} as

$$\beta_{\text{opt}} = \arg\max_{\beta}\{\Theta_{\text{ac}} : \Theta_{\text{ac}} = \beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_{\beta,\min})\}. \tag{23}$$

Since it is hard to obtain the explicit expressions of β_{opt} and Θ_{ac}, numerical results are illustrated in Fig.3, according to which, a notable gap between the achievable throughput and the sum capacity can be observed even with optimal overloading. This is mainly due to that the relatively low complexity BP algorithm achieves only local optimum which leads to possible performance loss when the degrees of the underlying graph increase [11]. In the following section, the problem is addressed by employing the most recently developed *spatial coupling* theory [11], [13], which results in a capacity-approaching RMA scheme with reserved attractive features of low signaling overhead and high adaptability as the basic RMA framework.

IV. SPATIALLY COUPLED RMA

As shown in [11], the *locally* optimal BP algorithm generally has a threshold lower than the *globally* optimal but highly complicated MAP (maximum *a posteriori*) algorithm, where the threshold is a unique parameter reflecting the tolerable channel condition, such as erasure probability or noise level, etc. Fortunately, it is revealed in [11] that by properly spatially coupling individual codes, the BP threshold of the resultant new ensemble can approach to its maximum possible value, namely the MAP threshold of the underlying ensemble. Such a phenomenon of "*threshold saturation*" gives an entirely new way to approach capacity.

Motivated by the above theory, in this section, we propose a novel spatially coupled RMA scheme, which naturally produces a new superposition rateless code with a spatially coupled graph by properly mapping the coded symbols of different users to some specific subsets of REs. According to the analysis, SC-RMA approaches asymptotically the channel capacity with reserved attractive features as the basic RMA framework.

A. ACCESS PROCEDURE OF SC-RMA

To highlight its intrinsic structure, we present a *protograph* [11] representation of SC-RMA in Fig.4, in which the precoding parts correspond to the LDPC precoders, and the differently colored squares denote different *types* of REs. Each RE is randomly assigned a unique *type* value within $\{1-w, 2-w, \ldots, K+w\}$, where $w \ll K$ is a given natural number,[1] i.e., there are $K+2w$ types of REs in total. The arrow denotes that the user can access the corresponding type of REs. As a result, each type of REs can only be accessed by a specific subset of users. In particular, for the REs with type s, only the subset of users $\{U_k : |k-s| \leq w\}$ have access to them, while any other users are not allowed to transmit over them. For example, for the RE with type $1-w$, only U_1 has access to it. Similarly, only the set of users $\{U_1, \ldots, U_{2w+1}\}$ have access to the RE of type $w+1$. Note that, since the basic RMA allows all the active users to access the *only* type of REs, it can be regarded as a special case of SC-RMA.

The access process is much alike that of the basic RMA. Each user pseudo-randomly selects its precoded symbols according to its RACf, and then transmits the linear sum over the allowed types of REs and stops the transmission once it receives an ACK from the AP. The AP also resorts

[1] On the other way around, in the case $K << w$, SC-RMA would degrade to the basic RMA, since all the users have equal access to all the REs approximately in such a case.

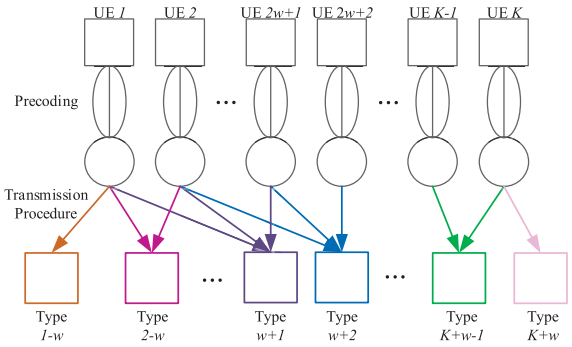

FIGURE 4. Protograph representation of SC-RMA: each type of REs can only be accessed by a specific subset of the users.

to the BP decoding on the spatially coupled factor graph to retrieve the information bits. Evidently, the change brought by the spatial coupling in the transmission process is subtle, which brings attractive features for SC-RMA that it inherits the low signaling overhead and high adaptability from the basic RMA.

B. THEORETICAL ANALYSIS OF SC-RMA

In the following, we derive the achievable throughput of SC-RMA. For the same reasons, we consider the two-stage decoding algorithm.

First, we consider the first stage of the decoding algorithm. In this stage, the subgraph in the bottom of Fig.4 is updated iteratively until the MI of VNs converge. Note that in Section III, we have proved that the uniform REN degree distribution is optimal in terms of throughput. In the SC-RMA scheme, the degrees of RENs of type s, $s \in \{1-w, 2-w, \ldots, w\} \cup \{K-w+1, K-w+2, \ldots, K+w\}$, are different from RENs of other RENs since a different number of users have access to these REs, as depicted in Fig.4. However, as elaborated in the following, this does not affect the performance visibly since the $w \ll K$. The converged MI of VNs in this case can be determined as below.

Lemma 3: In the SC-RMA scheme, the channel that one user faces can be approximated as a Gaussian channel with certain SNR degradation. The evolution process of the degradation factor can be expressed as follows:

$$\eta_k^{j+1} = \frac{1}{1+2w} \sum_{i_1=-w}^{w} \frac{1}{1+\frac{\beta(K+2w)}{(1+2w)K} \sum_{i_2=-w}^{w} f(\eta_{k+i_1+i_2}^{j})}, \tag{24}$$

where η_k^j ($k \in \{1, 2, \ldots, K\}$) denotes the degradation factor of U_k in iteration j, and $\eta_k^0 = 0$ for $k \in \{1, 2, \ldots, K\}$. For the extreme case where $\frac{w}{K} \to 0$, (24) is reduced to

$$\eta_k^{j+1} = \frac{1}{1+2w} \sum_{i_1=-w}^{w} \frac{1}{1+\frac{\beta}{(1+2w)} \sum_{i_2=-w}^{w} f(\eta_{k+i_1+i_2}^{j})}. \tag{25}$$

Proof: See Appendix C. ■

We can see that for different $k \in \{1, 2, \ldots, K\}$, η_k^j as well as the MI of VNs $J(\sqrt{\frac{4\eta_k^j \gamma}{\beta}})$, are not equal. According to the access procedure of SC-RMA as described before, since U_1 and U_K suffer the least interferences, η_1^j and η_K^j will be larger than η_k^j ($k \in \{2, 3, \ldots, K-1\}$). However, when analyzing the achievable throughput for SC-RMA, we only need to consider U_1 (or U_K) as described below.

Lemma 4: In the SC-RMA scheme where all the users share the same LDPC precoder, other users can also be reliably decoded iff U_1 and U_K are reliably decoded.

Proof: Once U_1 (or U_K) has been correctly decoded, other users can then be decoded in a successive interference cancellation manner. In particular, the AP can remove the subgraph of U_1 (or U_K) from the protograph, which leads to a situation that U_2 (respectively U_{K-1}) faces the same interference as U_1 (respectively U_K) does (see the protograph in Fig. 4), indicating that U_2 (respectively U_{K-1}) can also be successfully decoded the same way.

On the other hand, in the protograph, the subgraphs of U_1 and U_K are connected to dedicated REs which have no inputs from other users. Therefore, U_1 and U_K always suffer the least interference in average and thus can be decoded successfully w.h.p. if all other users can be reliably recovered in the BP decoding. ■

Based on these lemmas, we derive the achievable throughput for SC-RMA as follows.

Theorem 2: For SC-RMA, the asymptotic achievable throughput Θ_{ac} can be expressed as

$$\Theta_{\text{ac}} = \beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta} \eta_1^{\infty}). \tag{26}$$

In particular, when $K \to \infty$, $w \to \infty$ and $\frac{w}{K} \to 0$, we have

$$\lim_{\beta \to \infty} \Theta_{\text{ac}} = \lim_{\beta \to \infty} (\beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta} \eta_1^{\infty})) = \frac{1}{2} \log_2(1+\gamma). \tag{27}$$

Proof: See Appendix D. ■

Remark 3: Although each user only takes very simple binary modulation, RMA can still approach the Gaussian capacity asymptotically. This is due to that the channel input is the sum of several randomly selected symbols and the distribution of the resultant symbol tends to be Gaussian when the number of selected symbols is relatively large.

Remark 4: When $K \to \infty$, the numerical results for achievable throughputs with given w are presented in Fig.5, which shows that the achievable throughput increases with w and approaches asymptotically the channel sum capacity. In particular, when $w = 0$, the system can be regarded as a collection of K independent RMA subsystems, each of which occupies $\frac{1}{K}$ of the total REs and has the same throughput. Therefore, it is equivalent to the basic RMA system which has only a single type of REs (i.e., all the users have access to all the REs). Thus from Fig.5, we can see that the basic RMA system has the least achievable throughput under BP decoding, and spatial coupling ($w > 0$) can bring significant gain in throughput. Moreover, we note that the gap to the capacity increases as the SNR increases, which shows that the required w increases as SNR increases.

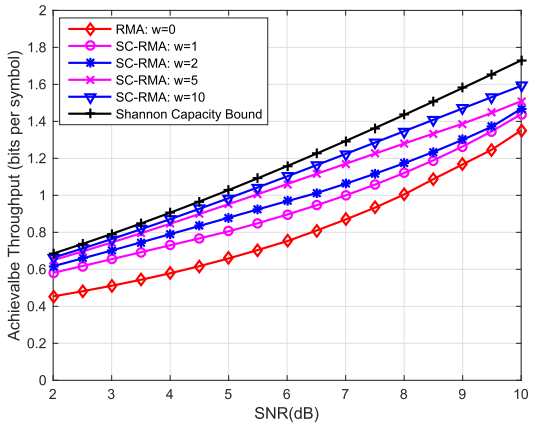

FIGURE 5. Achievable throughput increases with w and approaches to the channel sum capacity.

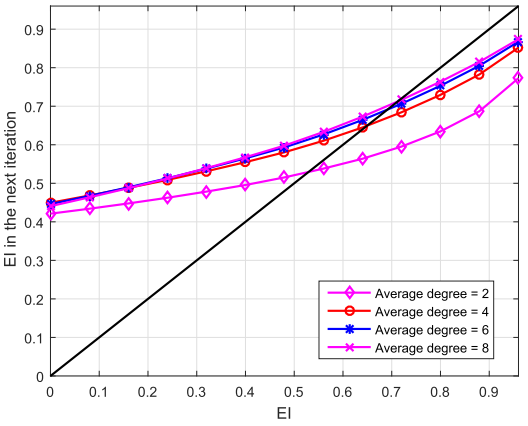

FIGURE 6. The evolution of extrinsic information of VNs in one iteration: the intersection point that determines the converged MI gradually approaches some stationary point as d_{RE} increases.

Moreover, when the number of users K is relatively small, according to (24), we have

$$\eta_1^1 = \frac{1}{2w+1}\sum_{i_1=1}^{2w+1}\frac{1}{1+\frac{i_1}{2w+1}\frac{K+2w}{K}\gamma}. \tag{28}$$

When $2w+1 < K(\ll \infty)$, η_1^1 and consequently the achievable throughput, increase with w at first and then drop down. For further enhancing the performance in such a case, we need to optimize w carefully, which will be elaborated in Section V.

V. PRACTICAL DESIGN OF THE SYSTEM PARAMETERS

The main system parameters including the RACfs, the spatial coupling pattern w, and the coding rates and degree profiles of LDPC precoders, should be carefully designed in real applications. However, due to the prohibitive complexity, it is difficult to achieve a joint optimization of these system parameters. Instead, for tractability and flexibility, here we provide a practical two-step approach to determine the system parameters.

First, we consider the design of RACf. Since it is a nonlinear multi-variate optimization problem with unknown dimensions, the optimal analytical RACfs are hard to obtain. With the aid of EXIT analysis, here we present an empirical but feasible solution. In particular, from the theoretical analysis in previous section, we know that the sum capacity could be approached for the asymptotic case in which the average degree of RENs d_{RE} tends to infinity. However, a large average REN degree usually results in high decoding complexity, and moreover, it also brings much difficulty in the determination of an RACf with a large number of unknowns. Therefore, for ease of treatment and near-optimal performance, we choose to seek a relatively simple but practical RACf with reduced dimensions and sufficiently large average REN degree, i.e., $\rho(x) = (1-p) + px^d$, which has only two non-zero coefficients. With this specific RACf, the user only needs to choose whether to transmit on certain RE, which further greatly simplifies the access procedure. Note that for

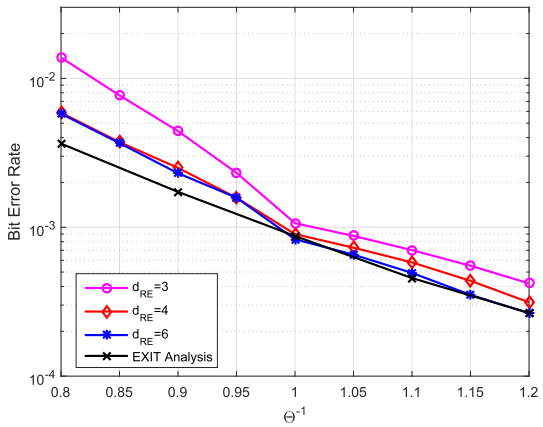

FIGURE 7. BER versus Θ^{-1}: the decoding performance improves as d_{RE} increases.

such an RACf, the average degree of RENs $d_{\mathrm{RE}} = pKd$, increases with K. This in fact gives us an opportunity to adapt p and d to the dynamic change in K.

Now using the EXIT theory, the evolution of the MI of VNs for different d_{RE} can be predicted and the numerical results are provided in Fig.6. For each specific EXIT curve, the VNs will get more MI after each decoding iteration when the EXIT curve is above the black diagonal line. Otherwise, less MI will they get. As a result, the MI of VNs finally converges to the intersection point of the EXIT curve and the diagonal line, which determines the overall decoding performance. As d_{RE} increases, the intersection point will gradually approach the stationary point that corresponds to the asymptotic throughput. Thus in practice, as a rule of thumb, in order to achieve a good tradeoff between the throughput performance and the decoding complexity, we can choose d_{RE} as the one that makes the intersection point close enough to the stationary point, i.e., 6 or 8 as shown in the figure. This observation is also verified by simulation results in Fig.7 in Section VI.

Second, we jointly optimize the spatial coupling pattern w and the LDPC pre-coding parameters to maximize the

throughput. Let $\{\lambda_i\}$ (resp. $\{\rho_i\}$) denote the edge-degree distributions of VNs (resp. CNs) of LDPC codes.[2] Let $I^j_{V\to C_k}$ (resp. $I^j_{C\to V_k}$) denote the extrinsic information (EI) passed from VNs (resp. CNs) to CNs (resp. VNs) of U_k. In the initial case, we have

$$I^0_{V\to C_k}=0,\quad I^0_{C\to V_k}=0. \tag{29}$$

Invoking the Gaussian approximation [25], the variance of the soft messages that a VN of U_k gets from the connected CNs, can be expressed as $\sum\lambda_i i(J^{-1}(I^j_{C\to V_k}))^2$. Therefore, taking these information into account and invoking (24), we have

$$\eta^j_k=\frac{1}{1+2w}\sum_{i_1=-w}^{w}\frac{1}{1+\frac{\beta(K+2w)}{(1+2w)K}\sum_{i_2=-w}^{w}A_{k,i_1,i_2}} \tag{30}$$

in which

$$A_{k,i_1,i_2}\triangleq f\Big(\frac{\sum\lambda_i i(J^{-1}(I^j_{C\to V_{k+i_1+i_2}}))^2+4\frac{\eta^{j-1}_{k+i_1+i_2}\gamma}{\beta}}{4\frac{\gamma}{\beta}}\Big). \tag{31}$$

On the other hand, the evolution process of the EI during the LDPC decoding can be expressed as follows (please be referred to [25] for more details):

$$\begin{aligned}I^{j+1}_{V\to C_k}&=\sum\lambda_i J\Big(\sqrt{(i-1)(J^{-1}(I^j_{C\to V_k}))^2+4\frac{\eta^{j+1}_k\gamma}{\beta}}\Big),\\ I^{j+1}_{C\to V_k}&=1-\sum\rho_i J\Big(\sqrt{(i-1)(J^{-1}(1-I^{j+1}_{V\to C_k}))^2}\Big).\end{aligned} \tag{32}$$

As a result, with the initial value $I^0_{V\to C_k}=0$, $I^j_{V\to C_k}$ may be iteratively calculated based on (30) - (32). Furthermore, we define $T_{\min}$ as

$$T_{\min}\triangleq\min_T\{T: I^{j+1}_{V\to C_1}-I^j_{V\to C_1}>0, I^j_{V\to C_1}\in(0,1)\}. \tag{33}$$

If and only if, $T>T_{\min}$, $I^j_{V\to C_1}$ increases as the algorithm iterates until it approaches 1, which means that U_1 can be reliably decoded [25]. Invoking Lemma 4, $T_{\min}$ is thus the threshold of T above which the AP can recover all the original information reliably. Therefore, to maximize the throughput, we may optimize system parameters to minimize $T_{\min}$ as follows:

$$\min_{\{\lambda_i,R,w\}}\ T_{\min}\quad \text{s.t.}\ \begin{cases}0<R<1\\ \sum_i\lambda_i=1\\ w\in\mathbb{N},\quad w\ll K.\end{cases} \tag{34}$$

Note that, with given $\{w,R\}$, the optimization problem may be simplified as a standard LDPC optimization problem, which can be solved based on the differential evolution (DE) method [25]. Then the exhaustive two dimensional search approach can be employed to find the optimized w and R. The phenomenon that the achievable throughput increases with w at first and then drops down, as discussed in Remark 4, can be used to simplify the search process. The algorithm is summarized in Algorithm 1.

Algorithm 1 Algorithm to Solve the Optimization Problem

Input:
System parameters: m, γ and K;
Output:
Optimized parameters: $\{w,\lambda_i,R\}$;
1: Initialize $w=0$ and $T_{\min}=\infty$;
2: For any given R, optimize λ_i using the DE method to minimize $T_{R,w}$. Denote the minimal $T_{R,w}$ as T^* ;
3: **if** $(T^*<T_{\min})$ **then**
4: Keep $\{w^*,R^*,\lambda_i^*\}$ that corresponds to the minimal T^* as the candidate;
5: $w=w+1$, $T_{\min}=T^*$ and go to step 2;
6: **end if**
7: **return** $\{w^*,R^*,\lambda_i^*\}$.

VI. SIMULATION RESULTS

We first present simulation results to confirm Lemma 1. We set that $K=6$ and $\gamma=10$dB. In the access procedure, each active user transmits to the AP a packet of length 100 bits without precoding. The simulation results for Bit Error Rate (BER) performance are presented in Fig.7. Note the inflection points in Fig.7 at $\Theta^{-1}=1$ (especially when $d_{\text{RE}}=3$) are due to the uneven VN node degree distribution among users which leads to certain differences in their equivalent symbol power when d_{RE} is relatively small. As d_{RE} increases, the inflection points vanish gradually. According to the simulation results, the BER performance improves as d_{RE} increases, which matches well with the previous theoretical analysis. Since the performance at $d_{\text{RE}}=6$ is very close to the asymptotic case, we then use it for the following simulations.

Then, we show simulation results for the scenario in which the user number is relatively small. Again, we set $K=6$, i.e., there are 6 active users, and each of which transmits to the AP a packet of 240 bits. The LDS scheme is taken as a benchmark [7] and the well designed signature matrix reported in [30] is chosen for the simulation. For RMA, we choose RACf $\rho(x)=0.5+0.5x^2$, thereby we have $d_{\text{RE}}=6$. On the other hand, the optimized w is 0 since K is relatively small. In the precoding part, the same regular LDPC code of rate 0.6 is adopted for these two schemes. According to the simulation results in Fig.8, RMA achieves better performances using the same number of decoding iterations, without needing to design the rather complicated signature matrix as in LDS.

Finally, we evaluate RMA for the scenario in which there are a relatively large number of users. In this case, we set that the received SNR $\gamma=5$dB. Furthermore, there are

[2]In the following, we take the LDPC codes with uniform CN degree distributions as an example. In such a case, $\{\rho_i\}$ is determined by $\{\lambda_i\}$ and rate R.

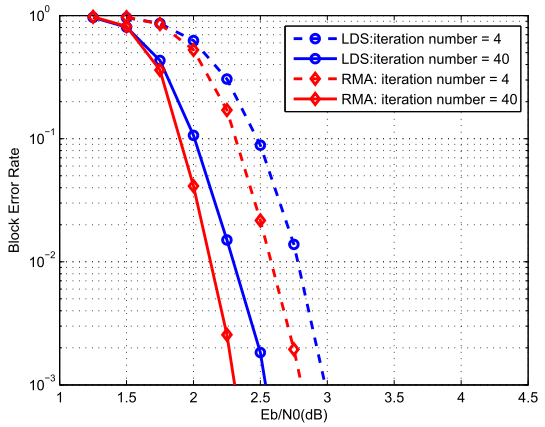

FIGURE 8. Performance comparison between RMA and LDS.

TABLE 1. Optimized LDPC codes and the achievable throughput for the system with $K = 100$ and $\gamma = 5dB$.

	C2 for w=2	C0 for w=0	C1 for w=1
R	0.4	0.7	0.5
Θ_{ac}	0.82645	0.70922	0.80645

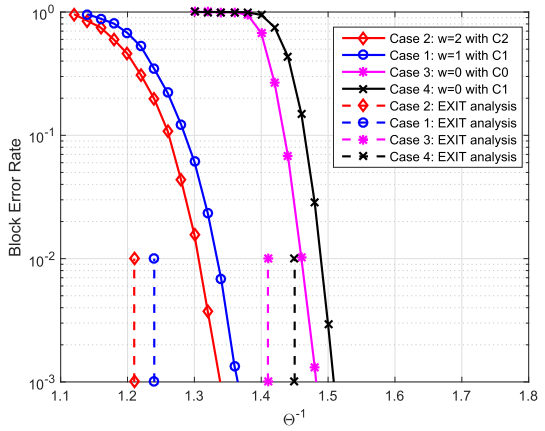

FIGURE 9. BLER versus Θ^{-1} for RMA with optimized LDPC codes for $w \in \{0, 1, 2\}$.

$K = 100$ users, each of which transmits to the AP a packet of 100 bits. By solving the optimization problem, we obtain the optimized spatial coupling pattern $w = 2$ and the corresponding optimized LDPC codes, namely, C2. For ease of comparison, we optimize LDPC codes for the case $w \in \{0, 1\}$ and denote the corresponding LDPC codes as C0 and C1 respectively. Their rates, theoretical throughput and distribution profiles can be found in Table 1 and at the top of next page, respectively. As discussed in Section IV, the SC-RMA with $w = 0$ corresponds to the basic RMA system. The analytical and simulation results are presented in Fig. 9. As Θ^{-1} increases, i.e., as we increase the number of channel uses or equivalently, REs, to beyond the EXIT thresholds as shown in Table 1 in a rateless manner, the block error rate (BLER) falls quickly. It also shows the BLER decreases as w increases, which corroborates that the performance of RMA can be further improved by proper spatial coupling. Moreover, it makes sense to compare the practical decoding performance with channel capacity, which is 1.0287 bits per channel use for a received SNR $\gamma = 5$dB. The gap between the system throughput and channel capacity get closer as the spatial coupling get tighter, which provides simple but sufficient clues that proper spatial coupling improves the system performance.

VII. CONCLUSION

In this work, we have established the novel rateless multiple access framework, which has attractive features of low signaling overhead, low decoding complexity and high adaptability to network traffic and channel changes. Based on the EXIT theory, we analyzed the asymptotic achievable throughput under BP decoding. By exploiting the recent theory of spatial coupling, we proposed the spatially-coupled RMA which further enhances the system performance. A practical two-step design approach for the system parameters is provided. According to the theoretical analysis and simulation results, RMA achieves high spectrum efficiency with high adaptability and low signaling overhead, which makes it a viable candidate for future random massive access.

APPENDIX A
PROOF OF LEMMA 1

The proof mainly follows the approach used in [27]. First, (3) can be equivalently formulated as

$$y_t = cx_n + \sum_{n' \in \mathcal{V}(t) \backslash n} cx_{n'} + z_t. \tag{35}$$

Define

$$W_t = \sum_{n' \in \mathcal{V}(t) \backslash n} cx_{n'}, \tag{36}$$

and

$$\Psi_\iota(y_t) = \sum_{\{x_{n'}\}_{n' \in \mathcal{V}(t) \backslash n}} \prod_{n' \in \mathcal{V}(t) \backslash n} p(x_{n'}) \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - W_t)^2}{2}} (y_t - W_t)^\iota, \tag{37}$$

for $\iota \in \{0, 1\}$. For the asymptotic case in which $d_t \to \infty$ and thereby $c \to 0$, we have

$$\begin{aligned} p(y_t | x_n) &= a, \{L^j_{n' \to t}\}_{n' \in \mathcal{V}(t) \backslash n}) \\ &= \sum_{\{x_{n'}\}_{n' \in \mathcal{V}(t) \backslash n}} \prod_{n' \in \mathcal{V}(t) \backslash n} p(x_{n'}) \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - W_t - ac)^2}{2}} \\ &\overset{(a)}{=} \sum_{\{x_{n'}\}_{n' \in \mathcal{V}(t) \backslash n}} \prod_{n' \in \mathcal{V}(t) \backslash n} p(x_{n'}) \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - W_t)^2}{2}} \\ &\quad \times \Big(1 + (y_t - W_t)ac + o(c)\Big) \\ &= \Psi_0(y_t) + \Psi_1(y_t)ac + o(c), \end{aligned} \tag{38}$$

$$\begin{aligned}
\text{C2 for } w &= 2:\lambda(x)=0.3552x^{19}+0.0010x^{17}+0.14x^{9}+0.3473x^{2}+0.1565x^{1}\\
\text{C0 for } w &= 0:\lambda(x)=0.3863x^{19}+0.0337x^{17}+0.1444x^{8}+0.3206x^{2}+0.1150x^{1}\\
\text{C1 for } w &= 1:\lambda(x)=0.4748x^{14}+0.0024x^{7}+0.0021x^{5}+0.0082x^{4}+0.3623x^{2}+0.1502x^{1}
\end{aligned}$$

where (a) follows from the Taylor series expansion. Hence,

$$\begin{aligned}
L_{n\leftarrow t}^{j+1} &= \log\frac{p(y_t|x_n=1,\{L_{n'\to t}^{j}\}_{n'\in\mathcal{V}(t)\backslash n})}{p(y_t|x_n=-1,\{L_{n'\to t}^{j}\}_{n'\in\mathcal{V}(t)\backslash n})}\\
&= \log(1+\frac{2\Psi_1(y_t)c}{\Psi_0(y_t)-\Psi_1(y_t)c})\\
&= \frac{2\Psi_1(y_t)c}{\Psi_0(y_t)-\Psi_1(y_t)c}\\
&\approx 2\frac{\Psi_1(y_t)}{\Psi_0(y_t)}c. \qquad (39)
\end{aligned}$$

Then the expectation and the variance of $L_{n\leftarrow t}^{j}$ can be calculated by

$$\begin{aligned}
&\mathrm{E}(L_{n\leftarrow t}^{j+1})\\
&= \int_{-\infty}^{+\infty}2\frac{\Psi_1(y_t)}{\Psi_0(y_t)}cp(y_t|x_n,\{L_{n'\to t}^{j}\}_{n'\in\mathcal{V}(t)\backslash n})dy_t\\
&\overset{(a)}{=} \int_{-\infty}^{+\infty}\Psi_1(y_t)dy_t+2\int_{-\infty}^{+\infty}\frac{\Psi_1^2(y_t)}{\Psi_0(y_t)}dy_tc^2x_n+o(c^2)\\
&\overset{(b)}{=} 2\int_{-\infty}^{+\infty}\frac{\Psi_1^2(y_t)}{\Psi_0(y_t)}dy_tc^2x_n+o(c^2) \qquad (40)
\end{aligned}$$

and

$$\mathrm{var}(L_{n\leftarrow t}^{j+1})=4\int_{-\infty}^{+\infty}\frac{\Psi_1^2(y_t)}{\Psi_0(y_t)}dy_tc^2+o(c^2), \qquad (41)$$

where (a) follows (38) and (b) follows from that $\Psi_1(y_t)$ is odd function.

Thereby, the expectation and the variance of L_n^{j+1}, defined in (9), can be calculated by

$$\mathrm{E}(L_n^{j+1})=\sum_{t\in\mathcal{Q}(n)}\mathrm{E}(L_{n\leftarrow t}^{j+1})=2\eta^{j+1}x_n\sum_{t\in\mathcal{Q}(n)}c^2+o(c^2) \qquad (42)$$

and

$$\mathrm{var}(L_n^{j+1})=\sum_{t\in\mathcal{Q}(n)}\mathrm{var}(L_{n\leftarrow t}^{j+1})=4\eta^{j+1}\sum_{t\in\mathcal{Q}(n)}c^2+o(c^2), \qquad (43)$$

in which $\eta^{j+1}\triangleq\int_{-\infty}^{+\infty}\mathrm{E}\Big(\frac{\Psi_1^2(y_t)}{\Psi_0(y_t)}\Big)dy_t$ where the expectation is taken over different d_t. Note that $\int_{-\infty}^{+\infty}\mathrm{E}\Big(\frac{\Psi_1^2(y_t)}{\Psi_0(y_t)}\Big)dy_t$ varies in j since $\Psi_\iota(y_t)$ depends on W_t at iteration j as indicated in (36) and (37).

According to the central limit theorem, when number of nodes in $\mathcal{Q}(n)$ tends to infinity, the distribution of $L_{n\to t}^{j+1}$ tends to be Gaussian. In this case, based on (13), we have $\sum_{t\in\mathcal{Q}(n)}c^2=\frac{\gamma}{\beta}$ and thereby,

$$L_n^{j+1}\sim\mathcal{N}(\frac{2\eta^{j+1}\gamma}{\beta}x_n,\frac{4\eta^{j+1}\gamma}{\beta}). \qquad (44)$$

Moreover, by invoking (14), when symbol x_n is passed through a Gaussian channel with SNR $\frac{\eta^{j+1}\gamma}{\beta}$, the distribution of the LLR will be identical to (44). Therefore, for RMA, the channel that one user faces can be approximated as a Gaussian channel with an SNR degradation factor of $\lim_{j\to\infty}\eta^j$.

At last, we calculate η^{j+1}. By invoking the central limit theorem again, we have $W_t\sim\mathcal{N}(\mu_{W_t},\sigma_{W_t}^2)$, and

$$\Psi_\iota(y_t)=\mathrm{E}\Big(\frac{1}{\sqrt{2\pi}}e^{\frac{-(y_t-W_t)^2}{2}}(y_t-W_t)^\iota\Big). \qquad (45)$$

Consequently,

$$\frac{\Psi_1^2(y_t)}{\Psi_0(y_t)}=\frac{1}{\sqrt{2\pi}}\frac{(y_t-\mu_{W_t})^2}{(1+\sigma_{W_t}^2)^{\frac{5}{2}}}\exp\Big[-\frac{(y_t-\mu_{W_t})^2}{2(1+\sigma_{W_t}^2)}\Big], \qquad (46)$$

and η^{j+1} can be calculated as

$$\eta^{j+1}=\mathrm{E}(\frac{1}{1+\sigma_{W_t}^2})=\sum_{i=1}^{l}\frac{\omega_i}{1+\sigma_{W_t|d_t=k_id_a}^2}. \qquad (47)$$

Furthermore, since $\sigma_{W_t|d_t=k_id_a}^2$ can be viewed as the estimation variance of the binary inputs of the related equivalent Gaussian channels in the form of $y=\sqrt{\eta^j\frac{\gamma}{\beta}}x+z$ where $z\sim\mathcal{N}(0,1)$, i.e., $\sigma_{W_t}^2=(k_id_a-1)c^2\mathrm{E}_z\Big((1-\tanh(\eta^j\frac{\gamma}{\beta}+\sqrt{\eta^j\frac{\gamma}{\beta}}z))^2\Big)$, thus we can get (18) after some manipulation.

APPENDIX B
PROOF OF LEMMA 2

Based on the edge degree distribution of RENs, defined in (16), $\mathrm{E}(d_t)$ can be expressed as

$$\mathrm{E}(d_t)=\sum_{i_1=1}^{l}\frac{\frac{\omega_{i_1}}{k_{i_1}}}{\sum_{i_2=1}^{l}\frac{\omega_{i_2}}{k_{i_2}}}k_{i_1}d_a=\frac{d_a}{\sum_{i_2=1}^{l}\frac{\omega_{i_2}}{k_{i_2}}}, \qquad (48)$$

based on which, we have

$$d_a=\mathrm{E}(d_t)\sum_{i_2=1}^{l}\frac{\omega_{i_2}}{k_{i_2}}\overset{(a)}{=}\frac{\gamma}{c^2}\sum_{i_2=1}^{l}\frac{\omega_{i_2}}{k_{i_2}} \qquad (49)$$

where (a) follows from (12).

Thereby, the degradation factor η^{j+1} can be calculated by:

$$\begin{aligned}\eta^{j+1} &= \sum_{i_1=1}^{l} \frac{\omega_{i_1}}{1+\frac{\beta k_{i_1} d_a c^2}{\gamma} f(\eta^j)} \\ &= \sum_{i_1=1}^{l} \frac{\omega_{i_1}}{1+k_{i_1}\beta f(\eta^j)\sum_{i_2=1}^{l}\frac{\omega_{i_2}}{k_{i_2}}}. \end{aligned} \tag{50}$$

The optimal $\{\omega_i\}$ refers to the one that maximizes η^{j+1} for the given η^j.

For clarity, we define $\tau \triangleq \beta f(\eta^j)$. The problem to optimize $\{\omega_i\}$ may be expressed as

$$\begin{aligned}&\max_{\{\omega_i\}} \sum_{i_1=1}^{l} \frac{\omega_{i_1}}{1+\tau k_{i_1}\sum_{i_2=1}^{l}\frac{\omega_{i_2}}{k_{i_2}}} \\ &\text{s.t.} \begin{cases} \sum_{i_1=1}^{l} \omega_{i_1} = 1, \\ \omega_{i_1} >= 0. \end{cases} \end{aligned} \tag{51}$$

Then we prove that the optimal solution is the uniform distribution by *induction*.

First, we consider the case that $l = 2$. Evidently, when $(\omega_1, \omega_2) = (1, 0)$ or $(0, 1)$, we have

$$\sum_{i_1=1}^{2} \frac{\omega_{i_1}}{1+\tau k_{i_1}\sum_{i_2=1}^{2}\frac{\omega_{i_2}}{k_{i_2}}} = \frac{1}{1+\tau} \tag{52}$$

for any k_i and τ. Define

$$\mathcal{M}(x) = \frac{x}{1+\tau k_1(\frac{x}{k_1}+\frac{1-x}{k_2})} + \frac{1-x}{1+\tau k_2(\frac{x}{k_1}+\frac{1-x}{k_2})} - \frac{1}{1+\tau}. \tag{53}$$

Therefore, $\mathcal{M}(x) = 0$ has at least two solutions: $x = 0$ and $x = 1$. Furthermore, we have

$$\mathcal{M}'(x=0) = \frac{-\tau(\frac{k_1}{k_2}-1)^2}{\frac{k_1}{k_2}(1+\frac{k_1}{k_2}\tau)(1+\tau)^2} < 0. \tag{54}$$

Because of the properties of quadratic equations, $\mathcal{M}(x) < \frac{1}{1+\tau}$ for $0 < x < 1$, which proves the lemma for $l = 2$.

Assume that the lemma is correct when $l = l^*$.

Then we prove that it is correct when $l = l^*+1$ by *contradiction*. If it is not correct, the optimal solution of (51) must be achieved in the case that $\omega_i \neq 0$ for any $i \in \{1, 2, \ldots, l^*+1\}$. Otherwise, it reduces to the case $l = l^*$ which has been proved correct. To find the optimal $\{\omega_i\}$ for $l = l^* + 1$, the following optimization problem can be formulated:

$$\begin{aligned}&\min_{\{\omega_i\}} \sum_{i_1=1}^{l^*+1} \frac{-\omega_{i_1}}{1+\tau k_{i_1}\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} \\ &\text{s.t.} \begin{cases} \sum_{i_1=1}^{l^*+1} \omega_{i_1} = 1 \\ -\omega_{i_1} < 0. \end{cases} \end{aligned} \tag{55}$$

The optimal solution satisfies the following KKT conditions [31]: for $i \in \{1, 2, \ldots, l^*+1\}$,

$$\begin{aligned}&\frac{-1}{1+\tau k_i\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} + \frac{1}{k_i}\sum_{i_1=1}^{l^*+1}\frac{i_1\tau\omega_{i_1}}{1+\tau k_{i_1}\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} \\ &\qquad -\mu_i + \varphi = 0, \end{aligned} \tag{56}$$

$$\sum_{i_1=1}^{l^*+1}\omega_{i_1} = 1, \quad \mu_i a_i = 0, \quad -\omega_i + a_i^2 = 0, \tag{57}$$

where μ_i, φ, a_i are auxiliary variables introduced by the KKT conditions [31].

Since $\omega_i \neq 0$ for $i \in \{1, 2, \ldots, l^*+1\}$, we have $a_i^2 \neq 0$. Thus, $\mu_i = 0$ and

$$\frac{-k_i}{1+\tau k_i\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} + \sum_{i_1=1}^{l^*+1}\frac{i_1\tau\omega_{i_1}}{1+\tau k_{i_1}\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} + k_i\varphi = 0, \tag{58}$$

for $i \in \{1, 2, \ldots, l^*+1\}$. Therefore, for any $i, i' \in \{1, 2, \ldots, l^*+1\}$, we have

$$\frac{-k_i}{1+\tau k_i\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} + k_i\varphi = \frac{-k_{i'}}{1+\tau k_{i'}\sum_{i_2=1}^{l^*+1}\frac{\omega_{i_2}}{k_{i_2}}} + k_{i'}\varphi. \tag{59}$$

However, there are no such $\{\omega_i\}$, which results in a *contradiction*. Thus we complete the proof.

APPENDIX C
PROOF OF LEMMA 3

Proof: Let $L^j_{(k,n)\leftarrow(t,s)}$ denote the LLR passed from REN t of type s to VN n of U_k in iteration j. Based on (39), it can be approximated as

$$L^j_{(k,n)\leftarrow(t,s)} = 2\frac{\Psi_1(y_t)}{\Psi_0(y_t)}c. \tag{60}$$

Then based on (40)-(41), its expectation and the variance can be calculated by

$$\mathrm{E}(L^j_{(k,n)\leftarrow(t,s)}) = \frac{2}{\phi_s^j}c^2 x_{k,n}, \quad \mathrm{var}(L^j_{(k,n)\leftarrow(t,s)}) = \frac{4}{\phi_s^j}c^2 \tag{61}$$

in which,

$$\begin{aligned}\phi_s^j &\triangleq 1+\frac{1}{2w+1}\frac{\frac{N}{K}}{\frac{T}{K+2w}}\sum_{i_1=-w}^{w} f(\eta^j_{s+i_1}) \\ &= 1+\frac{\beta(K+2w)}{(1+2w)K}\sum_{i_1=-w}^{w} f(\eta^j_{s+i_1}). \end{aligned} \tag{62}$$

Let $\mathcal{Q}(k, n)$ denote the set of indices of RENs that are connected to VN n of U_k. Let $L^{j+1}_{(k,n)}$ denote the overall LLR that VN n of U_k get from the connected RENs. Based on (61), its expectation and the variance can be calculated by

$$\begin{aligned}\mathrm{E}(L^{j+1}_{(k,n)}) &= 2x_{k,n}\frac{\gamma}{(1+2w)\beta}\sum_{i_1=-w}^{w}\frac{1}{\phi^j_{k+i_1}}, \\ \mathrm{var}(L^{j+1}_{(k,n)}) &= 4\frac{\gamma}{(1+2w)\beta}\sum_{i_1=-w}^{w}\frac{1}{\phi^j_{k+i_1}}. \end{aligned} \tag{63}$$

The distribution of $L_{(k,n)}^{j+1}$ tends to be the Gaussian distribution as the number of nodes in $\mathcal{Q}(k,n)$ tends to infinity. Thereby, with SC-RMA, the channel that one user faces can also be approximated as a Gaussian channel. In particular, the degradation factor of U_k in iteration $n+1$ can be calculated by:

$$\eta_k^{j+1} = \frac{1}{1+2w}\sum_{i_1=-w}^{w}\frac{1}{\phi_{k+i_1}^{j}}. \tag{64}$$

Based on (62),(64), we have

$$\eta_k^{j+1} = \frac{1}{1+2w}\sum_{i_1=-w}^{w}\frac{1}{1+\frac{\beta(K+2w)}{(1+2w)K}\sum_{i_2=-w}^{w} f(\eta_{k+i_1+j_i}^{j})}. \tag{65}$$

■

APPENDIX D
PROOF OF THEOREM 2

Proof: At first, we consider the subgraph in the bottom part of Fig.4. Based on the previous analysis, the soft message that VN n of U_1 gets from the connected RENs, denoted by $L_{(1,n)}^{\infty}$, can be approximated as

$$L_{(1,n)}^{\infty} = \frac{2\eta_1^{\infty}\gamma}{\beta}x_{1,n} + 2\sqrt{\frac{\eta_1^{\infty}\gamma}{\beta}}z, \tag{66}$$

in which $z \sim \mathcal{N}(0,1)$. Thereby, to guarantee the decodability, the maximum pre-coding rate that U_1 can take is $C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_1^{\infty})$.

Then based on Lemma 4, the overall achievable system throughput Θ_{ac} can be expressed as

$$\Theta_{\text{ac}} = \beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_1^{\infty}). \tag{67}$$

In the initial case, we do not have any MI of the VNs, i.e., $\eta_k^0 = 0 (k \in \{1, 2.., K\})$. Based on (62), we have

$$\phi_s^0 = \begin{cases} 1+\dfrac{(s+w)\beta}{2w+1}f(0) & (s<w) \\ 1+\beta f(0) & (w\le s\le K-w) \\ 1+\dfrac{(K+w+1-s)\beta}{2w+1}f(0) & (s>K-w), \end{cases} \tag{68}$$

in which $f(0)$ can be calculated by $f(0) = \frac{\gamma}{\beta}$ according to (19). Then we have

$$\begin{aligned} \eta_1^1 &= \lim_{w\to\infty}\frac{1}{2w+1}\sum_{i_1=1}^{2w+1}\frac{1}{1+\frac{i_1}{2w+1}\gamma} \\ &= \int_0^1 \frac{1}{1+\xi\gamma}d\xi \\ &= \frac{\ln(1+\gamma)}{\gamma}. \end{aligned} \tag{69}$$

Thereby, we have

$$\begin{aligned} \Theta_{\text{ac}} &= \beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_1^{\infty}) \ge \beta C_{\text{BIAWGNC}}(\frac{\gamma}{\beta}\eta_1^{1}) \\ &\overset{(a)}{=} \beta C_{\text{BIAWGNC}}(\frac{\ln(1+\gamma)}{\beta}) \\ &\overset{(b)}{\ge} \beta(\frac{1}{2\ln 2}\frac{\ln(1+\gamma)}{\beta} + A_1(\frac{\ln(1+\gamma)}{\beta})^2), \end{aligned} \tag{70}$$

where (a) follows from (69) and (b) follows from [32]. Here A_1 is certain constant. When $\beta \to \infty$, we have

$$\Theta_{\text{ac}} \ge \beta\frac{1}{2\ln 2}\frac{\ln(1+\gamma)}{\beta} = \frac{1}{2}\log_2(1+\gamma). \tag{71}$$

Furthermore, based on the information theory [33], we have

$$\Theta_{\text{ac}} \le \frac{1}{2}\log_2(1+\gamma). \tag{72}$$

Thus we complete the proof. ■

ACKNOWLEDGMENT

This paper was presented in part at the Globecom 2015 [1].

REFERENCES

[1] X. Wang, Z. Zhang, Y. Zhang, L. Zhang, and Y. Chen, "Multi-carrier rateless multiple access: A novel protocol for dynamic massive access," in *Proc. IEEE Globecom*, Dec. 2015, pp. 1–6.

[2] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] X. Chen and D. Guo, "Gaussian many-access channels: Definition and symmetric capacity," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2013, pp. 1–5.

[4] A. Zanella *et al.*, "M2M massive wireless access: Challenges, research issues, and ways forward," in *Proc. IEEE Globecom Workshops*, Dec. 2013, pp. 151–156.

[5] G. Durisi, T. Koch, and P. Popovski. (2015). "Towards massive, ultra-reliable, and low-latency wireless communication with short packets." [Online]. Available: https://arxiv.org/abs/1504.06526

[6] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. Int. Symp. Wireless Commun. Syst.*, Aug. 2014, pp. 781–785.

[7] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.

[8] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE PIMRC*, Sep. 2013, pp. 332–336.

[9] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[10] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815–6832, Dec. 2015.

[11] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[12] K. Takeuchi, T. Tanaka, and T. Kawabata, "Performance improvement of iterative multiuser detection for large sparsely spread CDMA systems by spatial coupling," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1768–1794, Apr. 2015.

[13] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.

[14] U. Niesen, U. Erez, D. Shah, and G. W. Wornell, "CTH05-2: Rateless codes for the Gaussian multiple access channel," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov./Dec. 2006.

[15] K. Wu, Z. Zhang, and S. Chen, "Rateless multiple access over erasure channel," in *Proc. IEEE Veh. Technol. Conf.*, May 2010, pp. 1–6.

[16] K. Wu, Z. Zhang, and S. Chen, "Rateless multiple access over noisy channel," in *Proc. Int. Wireless Commun. Mobile Comput. Conf.*, 2010, pp. 271–275.

[17] Z. Zhang, S. Chen, and K. Wu, "Non-coded rateless multiple access," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Oct. 2012, pp. 1–6.

[18] 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.

[19] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.

[20] C. Stefanović and P. Popovski, "ALOHA random access that operates as a rateless code," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4653–4662, Nov. 2013.

[21] M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Multiple access analog fountain codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 2167–2171.

[22] G. Liva, E. Paolini, M. Lentmaier, and M. Chiani, "Spatially-coupled random access on graphs," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 478–482.

[23] D. Truhachev, "Universal multiple access via spatially coupling data transmission," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1884–1888.

[24] D. Truhachev, "Achieving AWGN multiple access channel capacity with spatial graph coupling," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 585–588, May 2012.

[25] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[26] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[27] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 3, pp. 421–431, Apr. 2008.

[28] Z. Zhang, X. Wang, Y. Zhang, and Y. Chen, "Grant-free rateless multiple access: A novel massive access scheme for Internet of Things," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2019–2022, Oct. 2016.

[29] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[30] Y. Wu, S. Zhang, and Y. Chen, "Iterative multiuser receiver in sparse code multiple access systems," in *Proc. IEEE ICC*, Jun. 2015, pp. 2918–2923.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004, p. 244.

[32] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.

ZHAOYANG ZHANG (M'00) received the Ph.D. degree in communication and information systems from Zhejiang University, Hangzhou, China, in 1998. Since 1998, he has been with the College of Information Science and Electronic Engineering, Zhejiang University, where he became a Full Professor in 2005 and the Qiushi Distinguished Professor in 2017. He has a wide variety of research interests including information theory and coding theory, signal processing for communications and in networks, and fundamental aspects of computation-and-communication convergence. He was a recipient of the National Science Fund for Distinguished Young Scholars and a co-recipient of five international conference best paper awards. He is currently serving as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, *IET Communications* and some other international journals. He served as the General Chair, the TPC Co-Chair, or the Symposium Co-Chair for many international conferences and workshops such as ChinaCOM 2008, ICUFN 2011/2012/2013, WCSP 2013/2018, the Globecom 2014 Wireless Communications Symposium, and the VTC 2017 Workshop HMWC.

XIANBIN WANG received the B.S.Eng. and Ph.D. degrees in information engineering from the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China, in 2012 and 2017, respectively. He is currently with Huawei Technologies Co., Ltd. His research interests include massive access and channel coding in 5G.

YU ZHANG (M'14) received the B.S. degree in communication engineering and the Ph.D. degree in communication and information systems from Zhejiang University, Hangzhou, China, in 2008 and 2013, respectively. From 2013 to 2015, he was a Post-Doctoral Researcher with the Department of Information Science and Electronic Engineering, Zhejiang University. From 2013 to 2014, he visited the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, as a Post-Doctoral Researcher. Since 2015, he has been with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China, where he is currently an Assistant Professor. His current research interests include rateless coding, network coding, cooperative communication, and massive MIMO.

YAN CHEN received the B.Sc. and Ph.D. degrees from Zhejiang University in 2004 and 2009, respectively. She has been a Visiting Researcher with HKUST from 2008 to 2009. She joined Huawei Technologies, Shanghai, in 2009, where she was the Project Leader of the Green Radio Research from 2010 to 2013. She is currently the Technical Leader of multiple access research and standardization. She has received the award for the IEEE Advances in Communications in 2017.