

A Design of Low-Projection SCMA Codebooks for Ultra-Low Decoding Complexity in Downlink IoT Networks

Qu Luo, *Graduate Student Member, IEEE*, Zilong Liu, *Senior Member, IEEE*, Gaojie Chen, *Senior Member, IEEE*, Pei Xiao, *Senior Member, IEEE*, Yi Ma, *Senior Member, IEEE* and Amine Maaref, *Senior Member, IEEE*.

Abstract—This paper conceives a novel sparse code multiple access (SCMA) codebook design which is motivated by the strong need for providing ultra-low decoding complexity and good error performance in downlink Internet-of-things (IoT) networks, in which a massive number of low-end and low-cost IoT communication devices are served. By focusing on the typical Rician fading channels, we analyze the pair-wise error probability of superimposed SCMA codewords and then deduce the design metrics for multi-dimensional constellation construction and sparse codebook optimization. For significant reduction of the decoding complexity, we advocate the key idea of projecting the multi-dimensional constellation elements to a few overlapped complex numbers in each dimension, called low projection (LP). An emerging modulation scheme, called golden angle modulation (GAM), is considered for multi-stage LP optimization, where the resultant multi-dimensional constellation is called LP-GAM. Our analysis and simulation results show the superiority of the proposed LP codebooks (LPCBs) including one-shot decoding convergence and excellent error rate performance. In particular, the proposed LPCBs lead to decoding complexity reduction by at least 97% compared to that of the conventional codebooks, whilst owning large minimum Euclidean distance. Some examples of the proposed LPCBs are available at <https://github.com/ethanlq/SCMA-codebook>.

Index Terms—Sparse code multiple access (SCMA), golden angle modulation (GAM), codebook design, Internet-of-things (IoT), low complexity detection, Rician channels.

I. INTRODUCTION

THE Internet-of-things (IoT) represents a revolutionary paradigm shift from the legacy human-centric networks (e.g., in 3G and 4G) to massive machine-type communications, where the latter is a major use case in the 5G-and-beyond mobile networks [1]. Under this big picture, however, it is challenging to support the concurrent communications of massive IoT devices. Due to the limited time-frequency resources, traditional orthogonal multiple access (OMA) may be infeasible. A disruptive technique for addressing such a challenge is called non-orthogonal multiple access (NOMA) which permits several times of IoT devices larger than that in OMA systems communicating simultaneously [2].

Qu Luo, Gaojie Chen, Pei Xiao and Yi Ma are with 5G & 6G Innovation Centre, Institute for Communication Systems (ICS), University of Surrey, UK, email: {q.u.luo, gaojie.chen, p.xiao, m.yi}@surrey.ac.uk.

Zilong Liu is with the School of Computer Science and Electronic Engineering, University of Essex, UK, email: zilong.liu@essex.ac.uk.

Amine Maaref is with the Canada Research Center, Huawei Technologies Company Ltd., Ottawa, Canada, email: amine.maaref@huawei.com.

Existing NOMA techniques can be mainly categorized into two classes: power-domain NOMA [3] and code-domain NOMA (CD-NOMA) [4]. The former works by superimposing multiple users with distinctive power levels over the identical time-frequency resources, whereas the latter carries out multiplexing by employing different codebooks/sequences with certain intrinsic structural properties. This paper focuses on the sparse code multiple access (SCMA) [5], [6], which is a representative CD-NOMA scheme with sparse codebooks. In SCMA, every user's instantaneous input message bits are directly mapped to a multi-dimensional sparse codeword drawn from a pre-designed codebook [7]. SCMA is attractive as judiciously designed codebooks lead to constellation gain (due to signal space diversity), whilst their sparsity can be leveraged to enable near-optimum message passing (MPA) decoding [6].

Many studies have been carried out to reap the benefits of SCMA for the enabling of massive connectivity in IoT networks [8]–[13]. Recently, SCMA has also been applied in satellite IoT networks with novel random access design [14], [15] and new receiver schemes [12], [13]. To realize these SCMA based IoT systems, nevertheless, a fundamental question is: *how to design new sparse codebooks to enable highly efficient decoding in terms of complexity, energy/storage consumption, and convergence?* We will address this as detailed in the sequel.

A. Related works

In SCMA literature, to avoid the excessive design complexity, a widely adopted approach is to apply certain user-specific operations (e.g., interleaving, permutation, shuffling and phase rotations) to a common multidimensional constellation, called a mother constellation (MC), for multiple sparse codebooks [5], [16]. Following this approach, various types of SCMA codebooks have been developed in [17]–[23]. For excellent error performance, it is desirable to maximize the minimum product distance (MPD) or minimum Euclidean distance (MED) of an MC [24]. In general, a large MED leads to reliable detection in the Gaussian channel, whereas a large MPD is preferred for robust transmissions in the Rayleigh fading channel [25]. For example, the authors in [18] proposed a simple constellation rotation and interleaving codebook design scheme by using an MC with enlarged MED. Golden

angle modulation (GAM) constellation was adopted in [17] to construct SCMA codebooks with low peak-to-average power ratio (PAPR) properties. In [20], Star-QAM constellation was introduced to construct MC with large MED for downlink SCMA. In addition, a uniquely decomposable constellation group based codebook design approach was developed in [21] by maximizing the MED at each resource node. With the aid of generic algorithm, in [22], a novel class of power-imbalanced SCMA codebooks was developed in [22] by maximizing the MED of the superimposed codewords whilst maintaining a large MPD of the MC. Recently, new downlink quaternary sparse codebooks with large MED were obtained in [26] by a new iterative algorithm based on alternating maximization with exact penalty.

It is worth mentioning that the complexity of MPA decoding can be significantly reduced by designing an MC with low projection (LP) numbers [27], [28]. This is achieved by allowing certain overlapped constellation elements at the same dimension. The rationale is that, due to the multidimensional nature, any two MC vectors with certain overlapping dimension(s) may be well separated in other dimension(s). Importantly, constellation overlapping at certain dimension (or more) leads to less computations in the message passing based inference. By leveraging this advantage, the authors in [27] proposed a modified LP-MPA decoder which can utilise the constellation structure to reduce the number of calculations of belief messages at each iteration. Subsequently, the authors in [28] presented an LP codebook design based on the amplitude phase-shift keying constellation in uplink SCMA systems, where the the extrinsic information transfer of the constellation is used to aid the design. To the best of our knowledge, a comparative study on sparse codebooks with LP numbers in downlink channels is still missing. This is challenging because one needs to deal with the difficulties for joint optimization of the MC and user-specific codebook operators as well as the prohibitively high computational complexity in calculating the minimum Euclidean/product distance of the superimposed codewords, which significantly differs from the codebook design in uplink Rayleigh fading channels.

B. Motivations and contributions

This paper is driven by the key observation that a majority of IoT communication devices are usually low-end and lost-cost wireless sensors which are very constrained in terms of resources such as energy, CPU, and memory/storage capacity¹. In view of this and unlike many existing SCMA papers focusing on efficient receiver design, we aim for achieving extremely low decoding complexity and super-rapid convergence rate at the receiver from the codebook design perspective. In addition, in spite of numerous SCMA codebook designs for Gaussian channels or Rayleigh fading channels, they may not be optimal for other real-life channels. Typically, a practical wireless channel consists of a small-scale fading caused by non-line of sight (NLoS) propagation and static line of sight (LoS) path, which can be modeled as Rician fading

channels. The Rician fading channels are widely present in practical IoT networks, such as the terrestrial IoT networks [30], [31], satellite IoT networks [12], [13], [15], [32]–[35], and unmanned aerial vehicle (UAV) networks [36], [37].

The above observations motivate us to design enhanced codebooks that can substantially reduce the decoding complexity, while achieving good error performance for SCMA based downlink IoT networks under Rician fading channels. The main contributions of this work are summarized as follows:

- We analyze the system performance of SCMA in downlink Rician fading channels, and derive the minimum distance of the superimposed codewords by investigating and minimizing the conditional pairwise error probability (PEP). It is shown that our derived minimum distance generalizes the previous design metrics, i.e., the MED and MPD, which are for the AWGN and Rayleigh fading channels, respectively.
- We propose a novel class of LP codebooks (LPCBs) to enable ultra-low decoding complexity and excellent error performance. Specifically, an efficient algorithm is first proposed to construct a novel one-dimensional basic constellation with proper LP numbers based on GAM (LP-GAM)². Afterwards, we propose to jointly optimize the LP-GAM and users' constellation operators by considering the characteristics of LP-GAM.
- We pay a special attention on the application of the proposed LPCBs to a challenging design case where the codebook size and overloading factor are both large. The conventional state-of-the-art, however, may not be effective due to the prohibitively high decoding complexity in this case. Extensive simulations are conducted to evaluate the proposed LPCBs. Interestingly, it is shown that a converged decoding can be achieved in a single shot (called 'one-shot decoding' in this paper) for some of the proposed LPCBs, leading to dramatic reduction of the decoding complexity. In addition, to the best of our knowledge, some of the proposed LPCBs also achieve the largest MED values among all existing codebooks.

C. Organization

The rest of the paper is organized as follows. In Section II, the system model of downlink SCMA based IoT network along with the multiuser detection technique are presented. Section III analyzes the PEP of SCMA based IoT system under Rician fading channels and derives the criteria of sparse codebooks. In Section IV, the proposed sparse codebook design and optimization are elaborated. The numerical results are given in Section V. Finally, conclusions are made in Section VI.

D. Notations

$\mathbb{C}^{k \times n}$ and $\mathbb{B}^{k \times n}$ denote the $(k \times n)$ -dimensional complex and binary matrix spaces, respectively. Z_M stands for the integer set $\{1, 2, \dots, M\}$. $\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$ and $\|\cdot\|$ denote the floor

¹One may refer to [29] for an excellent survey on low-end IoT devices as well as their major challenges posed to the operating system design.

²GAM is an emerging modulation that leads to higher mutual information and lower PAPR performance over pulse-amplitude modulation and square quadrature amplitude modulation (QAM) [38].

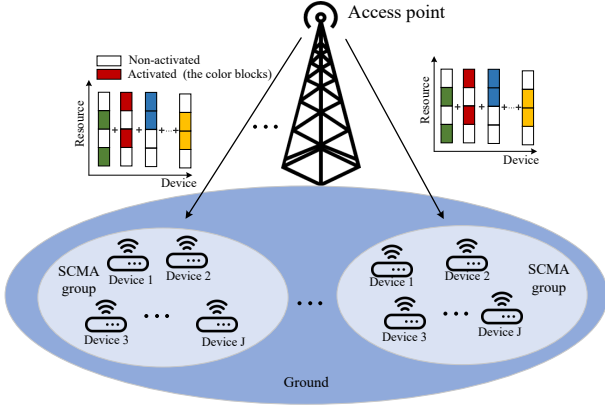


Fig. 1: An illustration of SCMA deployment in an IoT network.

function, absolute value, and the ℓ_2 – norm, respectively. $\mathcal{CN}(0,1)$ denotes complex Gaussian distribution with zero-mean and unit-variance. $\text{diag}(\cdot)$ denotes the diagonalization of a matrix.

II. SYSTEM MODEL

A. Signal Model

We consider an SCMA system in a downlink IoT network, where the access point serves multiple SCMA groups³, as shown in Fig. 1. Each SCMA group consists of J IoT devices occupying K orthogonal resource nodes, and the overloading factor is defined as $\lambda = \frac{J}{K} > 100\%$. At the transmitter side, the SCMA encoder maps $\log_2(M)$ binary bits to a length- K codeword \mathbf{x}_j drawn from pre-defined codebook $\mathcal{X}_j \in \mathbb{C}^{K \times M}$, where M denotes the modulation order. The mapping process is defined as

$$f_j: \mathbb{B}^{\log_2 M \times 1} \rightarrow \mathcal{X}_j \in \mathbb{C}^{K \times M}, \text{ i.e., } \mathbf{x}_j = f_j(\mathbf{b}_j), \quad (1)$$

where $\mathcal{X}_j = \{\mathbf{x}_{j,1}, \mathbf{x}_{j,2}, \dots, \mathbf{x}_{j,M}\}$ is the codebook set for the j th user with cardinality of M and $\mathbf{b}_j = [b_{j,1}, b_{j,2}, \dots, b_{j,\log_2 M}]^T \in \mathbb{B}^{\log_2 M \times 1}$ stands for the j th user's incoming binary message vector. The K -dimensional complex codewords in the SCMA codebook are sparse vectors with N non-zero elements and $N < K$. Let \mathbf{c}_j be a length- N vector drawn from $\mathcal{C}_j \subset \mathbb{C}^{N \times M}$, where \mathcal{C}_j is obtained by removing all the zero elements in \mathcal{X}_j . We further define the mapping from $\mathbb{B}^{\log_2 M}$ to \mathcal{C}_j as [39]

$$g_j: \mathbb{B}^{\log_2 M \times 1} \mapsto \mathcal{C}_j, \text{ i.e., } \mathbf{c}_j = g_j(\mathbf{b}_j). \quad (2)$$

Thus, the SCMA mapping defined in (1) now can be rewritten as

$$f_j \equiv \mathbf{V}_j g_j, \text{ i.e., } \mathbf{x}_j = \mathbf{V}_j g_j(\mathbf{b}_j), \quad (3)$$

where $\mathbf{V}_j \in \mathbb{B}^{K \times N}$ is a mapping matrix that maps the N -dimensional vector to a K -dimensional sparse SCMA codeword. The sparse structure of the J SCMA codebooks can be represented by an indicator (sparse) matrix $\mathbf{F}_{K \times J} = [\mathbf{f}_1, \dots, \mathbf{f}_J] \in \mathbb{B}^{K \times J}$ where $\mathbf{f}_j = \text{diag}(\mathbf{V}_j \mathbf{V}_j^T)$. The variable node j is connected to resource node k if and only

if $f_{k,j} = 1$. Furthermore, let $\mathcal{I}_r(k) = \{j | f_{k,j} = 1\}$ and $\mathcal{I}_u(j) = \{k | f_{k,j} = 1\}$ are the set of user indices sharing resource node k and the set of resource indices occupied by user j , respectively. In this paper, the following two factor indicator matrices with $\lambda = 150\%$ and $\lambda = 200\%$ are employed [17], [22]:

$$\mathbf{F}_{4 \times 6} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}, \quad (4)$$

$$\mathbf{F}_{5 \times 10} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \quad (5)$$

In the downlink channel, J users' data are first superimposed and then transmitted over K orthogonal subcarriers. The received signal of user u can be expressed as

$$\mathbf{y}_u = \text{diag}(\mathbf{h}_u) \sum_{j=1}^J \mathbf{x}_j + \mathbf{n}_u, \quad (6)$$

where $\mathbf{h}_u = [h_{u,1}, h_{u,2}, \dots, h_{u,K}]^T \in \mathbb{C}^{K \times 1}$ denotes Rician fading channel vector between the transmitter and u th IoT user. Namely, $h_{u,k} \sim \mathcal{CN}\left(\sqrt{\frac{\kappa}{1+\kappa}}, \sqrt{\frac{1}{1+\kappa}}\right)$ with κ represents the ratio of average power in the LoS path over that in the scattered component. The probability density function of the normalized Rician random variable is given by [40]

$$f_{|h_{u,k}|}(x) = 2x(1+\kappa) \exp(-\kappa - x^2(1+\kappa)) \times I_0\left(2x\sqrt{\kappa(1+\kappa)}\right), \quad x \geq 0, \quad (7)$$

where $I_0(\cdot)$ is the first order modified Bessel function of the first kind. $\mathbf{n}_u = [n_{u,1}, n_{u,2}, \dots, n_{u,K}]^T$ is the complex additive white Gaussian noise vector each element of which has zero mean and variance N_0 , i.e., $n_{u,k} \sim \mathcal{CN}(0, N_0)$. For simplicity, we assume the K subcarriers follow the same Rician distribution of κ , and hereafter the subscript u is omitted throughout this paper.

B. SCMA Detection

Similar to many previous works, we assume that the channel state information is perfectly known by receiver [17]–[23]. Define $I_{r_k \rightarrow u_j}^{(t)}(\mathbf{x}_j)$ as the belief message associated with codewords \mathbf{x}_j which is transmitted from resource node r_k to variable node u_j at the t th iteration. Similar, denoted by $I_{u_j \rightarrow r_k}^{(t)}(\mathbf{x}_j)$ the belief message from variable node u_j to resource node r_k . We assume equal probability for the input message \mathbf{x}_j , i.e., $I_{r_k \rightarrow u_j}^{(0)}(\mathbf{x}_j) = \frac{1}{M}, \forall j = 1, \dots, J, \forall k \in \mathcal{I}_u(j)$. The iterative messages exchanged between resource nodes and variable nodes are computed as

$$\begin{aligned} & I_{r_k \rightarrow u_j}^{(t)}(\mathbf{x}_j) \\ &= \sum_{\substack{\mathbf{x}_j = \mathbf{x} \\ i \in \mathcal{I}_r(k) \setminus \{j\} \\ \mathbf{x}_i \in \mathcal{X}_i}} \frac{1}{\pi N_0} \exp \left\{ -\frac{|y_k - h_k \sum_{i \in \mathcal{I}_r(k)} x_{k,i}|^2}{N_0} \right\} \\ & \quad \times \prod_{i \in \mathcal{I}_r(k) \setminus \{j\}} I_{u_i \rightarrow r_k}^{(t-1)}(\mathbf{x}_i), \end{aligned} \quad (8)$$

³The users in the same group generally have similar locations, and the detailed user grouping scheme is beyond the scope of this paper.

and

$$I_{u_j \rightarrow r_k}^{(t)}(\mathbf{x}_j) = \alpha_j \times \prod_{\ell \in \mathcal{I}_u(j) \setminus \{k\}} I_{r_\ell \rightarrow u_j}^{(t-1)}(\mathbf{x}_j), \quad (9)$$

where α_j is a normalization factor.

At each iteration, the main complexity of MPA is dominated by the message updating at the resource node, which can be approximated as $\mathcal{O}(KM^{d_f})$, where d_f is the number of users which collide over a resource node. To reduce the decoding complexity of MPA at each iteration, one possible approach is to reduce the effective M for the codebook, which is the key idea of the proposed LPCBs.

III. PROPOSED DESIGN CRITERIA OF SPARSE CODEBOOK IN DOWNLINK RICIAN CHANNELS

In this section, we first analyze the PEP performance of SCMA in the downlink Rician channel. Then, we present the corresponding codebook design criteria. We also show that the proposed design criteria can generalize the existing design criteria of MED and MPD in AWGN and Rayleigh fading channels, respectively.

A. PEP analysis in downlink Rician fading channels

In downlink SCMA systems, users' data are first superimposed over K orthogonal resources, which constitute a superimposed constellation, denoted as Φ_{MJ} . Let $\mathbf{w} = \sum_{j=1}^J \mathbf{x}_j$ be a superimposed codeword of Φ_{MJ} . Assume the erroneously decoded codeword is $\hat{\mathbf{w}}$ when \mathbf{w} is transmitted, where $\mathbf{w}, \hat{\mathbf{w}} \in \Phi_{MJ}$, and $\mathbf{w} \neq \hat{\mathbf{w}}$. Furthermore, let us define the element-wise distance $\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k) = \left| \sum_{j \in \mathcal{I}_r(k)} (x_{j,k} - \hat{x}_{j,k}) \right|^2$ and Euclidean distance $\delta_{\mathbf{w} \rightarrow \hat{\mathbf{w}}} = \sum_{k=1}^K \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)$. Then, the PEP conditioned on the channel fading vector for a maximum-likelihood receiver is given as [24]

$$\begin{aligned} \Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}} | \mathbf{h}\} &= Q\left(\sqrt{\frac{\|\text{diag}(\mathbf{h}) \sum_{j=1}^J (\mathbf{x}_j - \hat{\mathbf{x}}_j)\|^2}{2N_0}}\right) \\ &= Q\left(\sqrt{\frac{\sum_{k=1}^K h_k^2 \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{2N_0}}\right), \end{aligned} \quad (10)$$

where $Q(\cdot)$ is the Gaussian Q -function i.e., $Q(x) = (2\pi)^{-1/2} \int_x^{+\infty} e^{-t^2/2} dt$. A tight upper bound of Q -function can be expressed by $Q(x) \leq \sum_{i=1}^L a_i e^{-b_i x^2}$, where L, a_i, b_i are constant. Then, by applying the above upper bound and taking expectation with respect to \mathbf{h} on both sides of (10), we arrive

$$\begin{aligned} \Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}}\} &\leq \mathbb{E}_{\mathbf{h}} \left[\sum_{i=1}^L a_i \prod_{k=1}^K \exp\left(-\frac{b_i h_k^2 \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{2N_0}\right) \right] \\ &= \sum_{i=1}^L a_i \prod_{k=1}^K \mathbb{E}_{\mathbf{h}} \left[\exp\left(-\frac{b_i h_k^2 \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{2N_0}\right) \right] \\ &\stackrel{(i)}{=} \sum_{i=1}^L a_i \left\{ \prod_{k=1}^K \frac{1 + \kappa}{1 + \kappa + \frac{b_i \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{2N_0}} \exp\left(\frac{-\kappa \frac{b_i \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{2N_0}}{1 + \kappa + \frac{b_i \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{2N_0}}\right) \right\}, \end{aligned} \quad (11)$$

where step (i) is derived based on the fact that the distribution of h_k^2 relates to the non-central chi-square distribution with its

moment generating function given by [41]

$$M_{|h_k|^2}(s) = \frac{1 + \kappa}{1 + \kappa + s} \exp\left(-\frac{\kappa s}{1 + \kappa + s}\right). \quad (12)$$

By choosing $L = 1$, $a_1 = b_1 = \frac{1}{2}$ we can obtain the Chernoff bound⁴. Accordingly, the PEP can be written in the form

$$\Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}}\} \leq \frac{1}{2} \exp\left(-\frac{d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2}{4N_0}\right), \quad (13)$$

with

$$d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2 = \sum_{k=1}^K \left\{ \underbrace{\frac{\kappa \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{1 + \kappa + \frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0}}}_{d_{1, \mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k)} + 4N_0 \ln \left(1 + \frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0(1 + \kappa)} \right) \right\}. \quad (14)$$

B. Design criteria of multi-dimensional codebooks

Define $n_e(\mathbf{w}, \hat{\mathbf{w}})$ as the erroneous bits when $\hat{\mathbf{w}}$ is decoded at receiver, the union bound of the average bit error rate (ABER) for SCMA systems is given by

$$P_b \leq \frac{1}{M^J \cdot J \log_2(M)} \sum_{\mathbf{w}} \sum_{\hat{\mathbf{w}} \neq \mathbf{w}} n_e(\mathbf{w}, \hat{\mathbf{w}}) \Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}}\}. \quad (15)$$

At large SNRs, as the ABER is mainly dominated by the largest value of PEP in (13), we assert that improving the minimum distance of $d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2$ among all pairs will lead to lower ABER. Hence, we formulate our codebook design of the SCMA system with structure $\mathcal{S}(\mathcal{V}, \mathcal{G}; J, M, N, K)$, $\mathcal{V} := [\mathbf{V}_j]_{j=1}^J$ and $\mathcal{G} := [g_j]_{j=1}^J$ in downlink Rician fading channels as

$$\begin{aligned} \mathcal{P}_1 : \quad \mathcal{V}^*, \mathcal{G}^* &= \arg \max_{\mathcal{V}, \mathcal{G}} \Delta_{\min}(\mathcal{X}), \\ \text{Subject to} \quad &\sum_{j=1}^J \mathcal{X}_j = MJ, \end{aligned} \quad (16)$$

where $\mathcal{X}_j, j = 1, 2, \dots, J$ are the generated codebook of j th user based on (3), $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_J\}$ and $\Delta_{\min}(\mathcal{X})$ denotes the minimum distance of the codebook \mathcal{X} in Rician channels, which is obtained by calculating $M^J(M^J - 1)/2$ mutual distances between the superimposed codewords in Φ_{MJ} , i.e.,

$$\Delta_{\min}(\mathcal{X}) \triangleq \min \{d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2 \mid \forall \mathbf{w}, \hat{\mathbf{w}} \in \Phi_{MJ}, \mathbf{w} \neq \hat{\mathbf{w}}\}. \quad (17)$$

Define the multiple error events (MEEs) as the detection errors occurred with multiple users, and the single error event (SEE) as the detection errors occurred with a single user. Then, we introduce the following lemma.

Lemma 1: At sufficiently high SNR values, for $(1 + \kappa) \gg \frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0}$, $d_{1, \mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k) \rightarrow 0$ and the SEE dominates the ABER, whereas for $(1 + \kappa) \ll \frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0}$, $d_{2, \mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k) \rightarrow 0$ and the MEEs dominate the ABER. Moreover, the proposed design criteria can also generalize the previous design metric of MED and MPD in the AWGN and Rayleigh fading channels, respectively.

Proof: Refer to Appendix A.

⁴Alternatively, one can also use $L = 2$, $a_1 = \frac{1}{12}$, $b_1 = \frac{1}{2}$, $a_2 = \frac{1}{4}$, $b_2 = \frac{2}{3}$ for the approximation. In general, the Chernoff bound may be a little loose compared to the approximation of $L \geq 2$. However, this does not affect the codebook optimization [20], [22], [23].

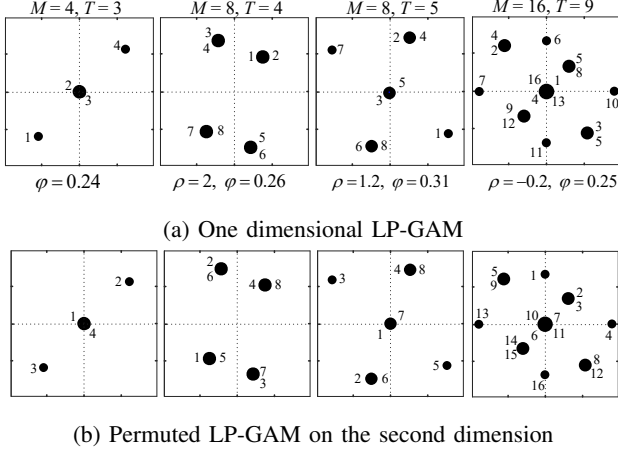


Fig. 2: Examples of $\mathcal{A}_{M,T}$ and $\pi_2(\mathcal{A}_{M,T})$.
IV. PROPOSED LOW-PROJECTION CODEBOOK DESIGN

This section introduces our proposed LPCBs for low complexity detection by maximizing the proposed minimum distance metric, i.e., Δ_{\min} . The main design steps include: generate a one-dimensional LP-GAM vector, permute LP-GAM to obtain an N -dimensional MC, optimize the constellation operators by addressing (16), and optimize the bit-to-symbol labeling. To improve $\Delta_{\min}(\mathcal{X})$, we propose to jointly optimize the MC and the constellation operators.

A. Design of one-dimensional LP-GAM

Denote $\mathcal{A}_{M,T}$ as a length- M one-dimensional LP-GAM vector with T distinct elements, i.e., $M - T$ constellation points are overlapped. To obtain $\mathcal{A}_{M,T}$, we first construct a constellation \mathcal{A}_T with T points, then repeat $M - T$ points in \mathcal{A}_T by following certain rules. The selection of the \mathcal{A}_T is of vital importance for downlink SCMA codebook design. In this paper, we employ GAM to design the \mathcal{A}_T due to the attractive features of enhanced mutual information, distance and PAPR performance [38]. The n th constellation point of GAM with N_p points can be generated according to $x_n = r_n e^{i2\pi\omega n}$, where $r_n = c_{\text{norm}}\sqrt{n}$, $c_{\text{norm}} = \sqrt{\frac{2P}{N_p+1}}$, P is the power constraint and $\omega = \frac{1-\sqrt{5}}{2}$ is the golden angle in radians. The design procedure of $\mathcal{A}_{M,T}$ is specified as follows:

1) *Generate \mathcal{A}_T* : Generate the GAM points according to $x_n = c_{\text{norm}}\sqrt{n} + \rho e^{i2\pi(\varphi+\omega)n}$, where φ is an arbitrary angle and ρ is the amplitude factor. Note that in the proposed LP-GAM, we have two additional degrees of freedom for enhanced codebook design, i.e., ρ and φ . Such a constellation is referred to as (ρ, φ) -GAM. The construction steps of \mathcal{A}_T are as follows: 1) For even value of T , generate $N_p = \frac{T}{2}$ (ρ, φ) -GAM points, denoted by $\mathcal{A}_T^{(m)}$, $m = 1, 2, \dots, \frac{T}{2}$. The remaining $\frac{T}{2}$ points are obtained by symmetry: $\mathcal{A}_T^{(m+\frac{T}{2})} = -\mathcal{A}_T^{(m)}$ for $m = 1, 2, \dots, \frac{T}{2}$; 2) For odd value of T , generate $N_p = \frac{T-1}{2}$ (ρ, φ) -GAM points. Similarly, the other $\frac{T-1}{2}$ points are obtained by symmetry. Finally, the zero point in the complex domain is added to \mathcal{A}_T .

2) *Obtain $\mathcal{A}_{M,T}$ based on \mathcal{A}_T* : After generating the \mathcal{A}_T based on GAM, the next problem is how to allocate \mathcal{A}_T with overlapped signal points to a length- M vector. To proceed, let us define $\mathcal{P}_{M,T} = \{p_1, p_2, \dots, p_L\}$ and $\mathcal{T}_{M,T} =$

$\{t_1, t_2, \dots, t_L\}$ as the overlapped signal set and corresponding projected numbers, respectively, where $p_l \in \mathcal{A}_T$, $1 \leq l \leq L$ and $\sum_{l=1}^L t_l = M - T$. Namely, the constellation point p_l is overlap t_l times in $\mathcal{A}_{M,T}$. The following rules are adopted to determine \mathcal{P} and \mathcal{T} :

- The overlapping number for a particular constellation point, i.e., t_l , should be as small as possible, such that the constructed MC has a good distance property.
- The constellation points with lower energy in \mathcal{A}_T are preferentially added to $\mathcal{P}_{M,T}$, such that the constellation \mathcal{A}_M has a smaller average energy.
- \mathcal{P} and \mathcal{T} should exhibit certain symmetry, such that the obtained \mathcal{A}_M is also symmetrical and with a zero mean.

Next, we give an example of the constructed LP-GAM in Fig. 2(a). One can see that there are $M - T$ constellation points overlapped in LP-GAM, where the overlapped constellation set, i.e., $\mathcal{P}_{M,T}$ and $\mathcal{T}_{M,T}$ are predefined according to the above rules. Note that the LP-GAM of $M = 4$, $T = 3$ is a special case, in which the constellation is only determined by φ .

B. Construction of the N -dimension MC

After obtaining the one-dimensional constellation $\mathcal{A}_{M,T}$, the remaining $N - 1$ dimensions can be obtained by the repetition of $\mathcal{A}_{M,T}$. To obtain a higher gain, permutation is further applied to $\mathcal{A}_{M,T}$. Let $\pi_n : [1, 2, \dots, M]^T \rightarrow [\pi_{n,1}, \pi_{n,2}, \dots, \pi_{n,M}]^T$ denote the permutation mapping of the n th dimension, where $\pi_{n,m}$, $1 \leq \pi_{n,m} \leq M$ denotes the permutation index, and for $m \neq l$, $\pi_{n,m} \neq \pi_{n,l}$. Namely, the $\pi_{n,m}$ th constellation point in $\mathcal{A}_{M,T}$ is permuted to the m th position of $\pi_n(\mathcal{A}_{M,T})$ at n th dimension. Then the N -dimensional MC $\mathcal{C}_{MC} = [\bar{c}_1^T, \bar{c}_2^T, \dots, \bar{c}_M^T]^T \in \mathbb{C}^{N \times K}$ can be obtained as

$$\mathcal{C}_{MC} = [\pi_1(\mathcal{A}_{M,T}), \pi_2(\mathcal{A}_{M,T}), \dots, \pi_N(\mathcal{A}_{M,T})]^T. \quad (18)$$

To proceed, we further define the N -dimensional Cartesian product over the one-dimensional \mathcal{A}_T as

$$\begin{aligned} \mathcal{A}_T^N &= \underbrace{\mathcal{A}_T \times \mathcal{A}_T \times \dots \times \mathcal{A}_T}_N \\ &\triangleq \{(a_1, a_2, \dots, a_N) | \forall a_i \in \mathcal{A}_T, \forall i \in \{1, 2, \dots, N\}\}, \end{aligned} \quad (19)$$

where “ \times ” denotes the Cartesian product. By transforming every element in \mathcal{A}_T^N into a column vector, we obtain an N -dimensional matrix $\mathbf{A} \in \mathbb{C}^{N \times T^N}$. For simplicity, we denote the above Cartesian mapping as $f_{\mathbf{A}} : \mathcal{A}_T^N \rightarrow \mathbf{A}$. Then, we introduce the following lemma:

Lemma 2: To construct an N -dimensional \mathcal{C}_{MC} , the LP number should satisfy $\lceil \sqrt[N]{M} \rceil \leq T \leq M$. For the case of $T = \sqrt[N]{M} \in \mathbb{Z}$, the permutation π_n , $n = 1, 2, \dots, N$ are given by the mapping of $f_{\mathbf{A}}$.

Proof: Refer to Appendix B.

For the other values of T and M , we first fix the constellation at the first dimension with a randomly chosen permutation. Accordingly, for the second dimension, there will be a total of $T = M!$ pairing options. Consequently, there are $T = (M!)^{N-1}$ different choices of the permutations. The MC constrained PEP can be employed to aid the

design of permutation metric. To this end, we further define $d_{\min,t}^{\mathcal{C}_{MC}}, t = 1, 2, \dots, T$ as the minimum effective distance from the $M(M-1)/2$ mutual codeword distances arising from \mathcal{C}_{MC} . Based on the results in (13) and (14), we have

$$d_{\min,t}^{\mathcal{C}_{MC}} = \arg \min_{1 \leq i < l \leq M} \sum_{n=1}^N \frac{\kappa |c_{n,i}^{(t)} - c_{n,l}^{(t)}|^2}{1 + \kappa + \frac{|c_{n,i}^{(t)} - c_{n,l}^{(t)}|^2}{4N_0}} + 4N_0 \ln \left(1 + \frac{|c_{n,i}^{(t)} - c_{n,l}^{(t)}|^2}{4N_0(1 + \kappa)} \right), \quad (20)$$

where $c_{n,i}^{(t)}$ denotes the i th symbol at the n th entry of \mathcal{C}_{MC} . Hence, the permutation should be chosen to maximise (20), i.e.,

$$\mathcal{C}_{MC}^* = \arg \max_{t=1,2,\dots,T} d_{\min,t}^{\mathcal{C}_{MC}}. \quad (21)$$

Lemma 3: By substituting $\kappa \rightarrow \infty$ (AWGN) and $\kappa = 0$ (Rayleigh fading) in (20), the equivalent distance metrics $d_{\min,t}^{\mathcal{C}_{MC}}$ are obtained as

$$d_{E,\min,t}^{\mathcal{C}_{MC}} = \arg \min_{i \neq l} \|c_i^{(t)} - c_l^{(t)}\|, \quad (22)$$

and

$$d_{P,\min,t}^{\mathcal{C}_{MC}} = \arg \min_{i \neq l} \prod_{n \in \rho(c_i, c_l)} |c_{n,i}^{(t)} - c_{n,l}^{(t)}|^2, \quad (23)$$

respectively, where $\rho(c_i, c_l)$ denotes the set of indices in which $c_{n,i} \neq c_{n,l}$. Clearly, the MED and MPD are the corresponding permutation criteria for AWGN and Rayleigh fading channels, respectively [23].

For small constellation size, e.g., $M \leq 8$ and $N \leq 2$, one can find the optimal permutation by exhaustive search with reasonable computational complexity. However, for large constellation size of \mathcal{C}_{MC} , it becomes intractable due to the huge number of permutation pairs. To overcome such a difficulty, the binary switching algorithm (BSA) [42] can be modified to find the sub-optimal permutation pattern. Fig. 2(b) shows the permuted constellation corresponding to the LP-GAM in Fig. 2(a) for a 2-dimensional \mathcal{C}_{MC} . Taking \mathcal{C}_{MC} with $N = 2, M = 4, T = 3$ as an example, the second and third points are overlapped at the first dimension; however, they are separated at the second dimension with the permutation.

C. Sparse codebook construction

After the design of the \mathcal{C}_{MC} , phase rotations are applied to generate multiple sparse codebooks. In addition, aiming at increasing the power diversity between users and improving the minimum distance properties of the superimposed codewords, we introduce energy scaling to each codebook dimension [22]. Specifically, the j th user's codebook with non-zero elements is generated by $\mathcal{C}_j = \mathbf{E}_j \mathbf{R}_j \mathcal{C}_{MC}$, where \mathbf{R}_j and \mathbf{E}_j denote the phase rotation matrix and power scaling matrix for the j th user, respectively. Note that the phase rotation matrix and power scaling matrix can be combined together to form a constellation operator matrix, denoted as $\mathbf{\Lambda}_j = \mathbf{E}_j \mathbf{R}_j$. For example, for a 2-dimensional \mathcal{C}_{MC} , \mathbf{E}_j and \mathbf{R}_j are given as

$$\mathbf{E}_j = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix}, \mathbf{R}_j = \begin{bmatrix} e^{j\theta_1} & 0 \\ 0 & e^{j\theta_2} \end{bmatrix}, \mathbf{\Lambda}_j = \begin{bmatrix} z_1 & 0 \\ 0 & z_2 \end{bmatrix}, \quad (24)$$

where $z_i = E_i e^{j\theta_i}, \forall E_i > 0, 1 \leq i \leq N$. Based on \mathbf{V}_j which has been introduced in Section II, the j th user's codebook is generated by $\mathcal{X}_j = \mathbf{V}_j \mathbf{\Lambda}_j \mathcal{C}_{MC}$. We further combine the constellation operation matrix $\mathbf{\Lambda}_j$ and mapping matrix \mathbf{V}_j together, i.e., $\mathbf{z}_{N \times J}^j = \mathbf{V}_j \mathbf{\Lambda}_j \mathbf{I}_K$, where \mathbf{I}_K denotes a column vector of K 1's. Hence, J codebooks can be represented by the signature matrix $\mathbf{Z}_{N \times J} = [\mathbf{z}_{N \times J}^1, \dots, \mathbf{z}_{N \times J}^J]$.

Owing to the sparsity of the factor graph matrix, the number of users superimposed on one resource node is d_f which is less than the number of users. Therefore, d_f distinct rotation angles and scaling factors of each dimension should be optimized to distinguish the superimposed codewords. Different from the Latin rectangular construction, we consider the signature matrix $\mathbf{Z}_{K \times J}$ for (4×6) and (5×10) SCMA systems [22] as follows:

$$\mathbf{Z}_{4 \times 6} = \begin{bmatrix} 0 & z_1 & z_2 & 0 & z_3 & 0 \\ z_1 & 0 & z_2 & 0 & 0 & z_3 \\ 0 & z_3 & 0 & z_2 & 0 & z_1 \\ z_3 & 0 & 0 & z_2 & z_1 & 0 \end{bmatrix}, \quad (25)$$

$$\mathbf{Z}_{5 \times 10} = \begin{bmatrix} z_1 & z_2 & z_3 & z_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ z_4 & 0 & 0 & 0 & z_1 & z_2 & z_3 & 0 & 0 & 0 \\ 0 & z_3 & 0 & 0 & z_4 & 0 & 0 & z_1 & z_2 & 0 \\ 0 & 0 & z_2 & 0 & 0 & z_3 & 0 & z_4 & 0 & z_1 \\ 0 & 0 & 0 & z_1 & 0 & 0 & z_2 & 0 & z_3 & z_4 \end{bmatrix}. \quad (26)$$

Observing the power diversity, row-wise or column wise, with $|z_i| \neq |z_j|, \forall 1 \leq i < j \leq d_f$, may help improve the Δ_{\min} . Power imbalance among different users is also introduced into $\mathbf{Z}_{4 \times 6}$ and $\mathbf{Z}_{5 \times 10}$ by $|z_1| + |z_3| \neq 2|z_2|$ and $|z_1| + |z_4| \neq |z_2| + |z_3|$, respectively. The phase rotation angles $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_{d_f}]^T$ and the energy factors $\mathbf{E} = [E_1, E_2, \dots, E_{d_f}]^T$ are the parameters to be optimized, where the degree of freedom for $\boldsymbol{\theta}$ and \mathbf{E} is $d_f - 1$. The other two parameters to be jointly optimized are the ρ and φ of the LP-GAM. Hence, the number of parameters to be optimized for the proposed codebooks is $2(d_f - 1) + 1$ and $2(d_f - 1) + 2$ for $M = 4$ and $M > 4$, respectively.

Let $\mathcal{X}_{\lambda, M, T}$ be the M -ary codebook set with T projected numbers to be optimized for the SCMA system with the overloading $\lambda \in \{150\%, 200\%\}$. As discussed in Subsection III-B, the design goal is to maximize the minimum distance metric Δ_{\min} which is defined in (29). Hence, based on the proposed multi-dimensional codebook construction scheme, the problem of \mathcal{P}_1 in (29) is reformulated as

$$\mathcal{P}_2: \max_{\mathbf{E}, \boldsymbol{\theta}, \rho, \varphi} \Delta_{\min}(\mathcal{X}_{\lambda, M, T}) \quad (27)$$

$$\text{Subject to} \quad \sum_{i=1}^{d_f} E_i = \frac{MJ}{K}, E_i > 0, \quad (27a)$$

$$0 \leq \theta_i \leq \pi, \quad (27b)$$

$$-1 < \rho \leq T, \quad (27c)$$

$$0 \leq \varphi \leq \pi/2, \quad (27d)$$

$$\forall i = 1, 2, \dots, d_f. \quad (27e)$$

Unfortunately, it is quite difficult, if not impossible, to directly solve \mathcal{P}_2 due to the non-convex and high computation complexity of $\Delta_{\min}(\mathcal{X}_{\lambda, M, T})$. To obtain Δ_{\min} , it requires calculate $M^J (M^J - 1)/2$ mutual distances between the superimposed codewords in Φ_{MJ} . For the SCMA system with $M = 4$ and $J = 6$, the computational complexity of

$\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$ is moderate. However, as the increase of K , J and M , it brings overwhelming computational complexity. Hence, for a highly overloaded SCMA system or high rate ($M \geq 8$) codebooks, we transform \mathcal{P}_2 to a feasible problem with affordable complexity by considering the distinct features of LPCBs.

1) *The case of $T = \sqrt[M]{M}$:* Let $\mathcal{S}_{\text{sum}}^k$ denote the superimposed constellation at the k th resource node after removing the overlapped constellation in \mathcal{C}_{MC} . Namely, all constellation points in $\mathcal{S}_{\text{sum}}^k$ are unique and $|\mathcal{S}_{\text{sum}}^k| = T^{d_f}$. According to (25) and (26), $\mathcal{S}_{\text{sum}}^k$ can be obtained by $\mathcal{S}_{\text{sum}}^{(k)} = \{z_1 a_1 + z_2 a_2 + \dots + z_{d_f} a_{d_f} \mid \forall a_i \in \mathcal{A}_T, i = 1, 2, \dots, d_f\}$. For simplicity, and since each resource node has the same superimposed constellation, the superscript (k) is omitted. Then, we introduce the following Lemma:

Lemma 4: For the $\mathcal{X}_{\lambda,M,T}$ where $T = \sqrt[M]{M}$, solving the problem of \mathcal{P}_2 by maximizing $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$ is equivalent to address the following problem:

$$\begin{aligned} \mathcal{P}_{2-1} : \quad & \max_{\mathbf{E}, \theta, \rho, \varphi} \quad \text{MED}(\mathcal{S}_{\text{sum}}) \\ \text{Subject to} \quad & T = \sqrt[M]{M}, T \in \mathbb{Z}, \\ & (27a), (27b), (27c), (27d), (27e), \end{aligned} \quad (28)$$

where

$$\text{MED}(\mathcal{S}_{\text{sum}}) \triangleq \min \{|s_m - s_n|^2 \mid \forall s_m, s_n \in \mathcal{S}_{\text{sum}}, \forall m, n \in Z_{T^{d_f}}, m \neq n\}. \quad (29)$$

Proof: Refer to Appendix C.

Since T^{d_f} is relatively small, the complexity of calculating $\text{MED}(\mathcal{S}_{\text{sum}})$ is negligible.

2) *For the case when $T > \sqrt[M]{M}$:* We propose to maximize the lower bound of $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$. To proceed, let us first define

$$d_{1,\min}^{\mathcal{X}_{\lambda,M,T}} \triangleq \min \left\{ \sum_{k=1}^K d_{1,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2 \mid \forall \mathbf{w}, \hat{\mathbf{w}} \in \Phi_{M^J}, \mathbf{w} \neq \hat{\mathbf{w}} \right\}, \quad (30)$$

and

$$\begin{aligned} d_{2,\min}^{\mathcal{X}_{\lambda,M,T}} & \triangleq \min \left\{ \sum_{k=1}^K d_{2,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2 \mid \forall \mathbf{w}, \hat{\mathbf{w}} \in \Phi_{M^J}, \mathbf{w} \neq \hat{\mathbf{w}} \right\}, \\ & \stackrel{(i)}{=} \min_{1 \leq j \leq J} \left\{ \sum_{k=1}^K 4N_0 \ln \left(1 + \frac{|x_{j,i,k} - x_{j,l,k}|^2}{4N_0(1+\kappa)} \right) \right. \\ & \quad \left. \mid \forall x_{j,i,k}, x_{j,l,k} \in \mathcal{X}_j, \forall i, l \in Z_M, i \neq l \right\}, \end{aligned} \quad (31)$$

where $d_{1,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2$ and $d_{2,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2$ are given in (14). Step (i) holds since the minimum distance metric between superimposed codewords is achieved when the SEE occurs. Obviously, for the codebook $\mathcal{X}_{\lambda,M,T}$,

$$\Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T}) \triangleq d_{1,\min}^{\mathcal{X}_{\lambda,M,T}} + d_{2,\min}^{\mathcal{X}_{\lambda,M,T}}, \quad (32)$$

is a lower bound of $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$. Note that $\Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T})$ and $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$ may not achieve the minimum value at the same solution, however, the value of $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$ can still be improved by maximizing $\Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T})$. Moreover, upon taking into account of lemma 1 in Subsection III-B, one can see that improving the $d_{1,\min}^{\mathcal{X}_{\lambda,M,T}}$ and $d_{2,\min}^{\mathcal{X}_{\lambda,M,T}}$ can also reduce the probability of MEEs and SEE, respectively.

Algorithm 1 Modified BSA for SCMA codebook.

Input: Optimized codebook \mathcal{X}_j

Output: A locally optimum bit-to-symbol mapping \mathbf{z} for \mathcal{X}_j

- 1: **Initialize:** Randomly choose an index vector \mathbf{z} for codebook \mathcal{X}_j and sort \mathcal{X}_j out based on the random index vector, and set $\mathbf{z}^* = \mathbf{z}$
 - 2: **for** $I = 1 : I_{\max}$ **do**
 - 3: Obtain $\Xi(\mathbf{x}_{j,i}), i = 1, 2, \dots, M$ and Π_{ini} by updating the cost of $\mathcal{X}_j(\mathbf{z}^*)$ using (34)
 - 4: Sort $\Xi(\mathbf{x}_{j,i})$ in decreasing order and obtain the new index vector \mathbf{z}
 - 5: **for** $i = 1 : M$ **do**
 - 6: **for** $l = 1 : M$ **do**
 - 7: **if** $\mathbf{z}(i) = l$ **then next** l
 - 8: Switch the $\mathbf{z}(i)$ -th codeword with l -th codeword and update the index vector $\bar{\mathbf{z}}$
 - 9: Calculate the total cost Π_{sw} using (34)
 - 10: **if** $\Pi_{\text{sw}} \leq \Pi_{\text{ini}}$ **then**
 - 11: Retain the switch, and update \mathbf{z}^* and Π_{ini} with $\bar{\mathbf{z}}$ and Π_{sw} , respectively
 - 12: **else**
 - 13: Switch back the two symbols
 - 14: **end if**
 - 15: **end for**
 - 16: **end for**
 - 17: **end for**
-

In fact, for large or small value of κ , $\Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T})$ is a tight lower bound of $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$. Hence, instead of directly maximizing $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$ in \mathcal{P}_2 , we propose to address its lower bound, i.e.,

$$\begin{aligned} \mathcal{P}_{2-2} : \quad & \max_{\mathbf{E}, \theta, \rho, \varphi} \quad \Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T}) \\ \text{Subject to} \quad & T > \sqrt[M]{M}, T \in \mathbb{Z}, \\ & (27a), (27b), (27c), (27d), (27e). \end{aligned} \quad (33)$$

Note that the computational complexity of $d_{2,\min}^{\mathcal{X}_{\lambda,M,T}}$ is negligible since it only involves the calculation of \mathcal{X}_j . Hence, the computational complexity of $\Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T})$ is largely reduced compared to that of $\Delta_{\min}(\mathcal{X}_{\lambda,M,T})$.

For the sparse codebook optimization problem, it is designed with the highly non-linear objective function and linear constraints. Such problem can be efficiently solved with an optimization solver, e.g. MATLAB Global Optimization toolbox. In particular, we employ the *fmincon* solver to address the problem (28) and (33). The maximum iteration number for $\lambda = 150\%$ and $\lambda = 200\%$ SCMA systems are 30 and 25, respectively. Although the complexity of $\Delta_{\text{LB}}(\mathcal{X}_{\lambda,M,T})$ has been significantly reduced, the calculation of $d_{1,\min}^{\mathcal{X}_{\lambda,M,T}}$ is still a challenging problem for $M \geq 16$ or $d_f = 4$. To solve this problem, a sub-optimal Monte Carlo search method is adopted. The distance metric $d_{1,\min}^{\mathcal{X}_{\lambda,M,T}}$ is first calculated between Q elements, which is randomly selected from the superimposed codewords. Then, the above process is repeated with t_{\max} times to estimate the $d_{1,\min}^{\mathcal{X}_{\lambda,M,T}}$.

Algorithm 2 Construction of LPCBs.

Input: $J, K, \mathbf{V}_j, M, T, \kappa, N_0, \mathcal{P}_{M,T}$ and $\mathcal{T}_{M,T}$

Initialize: $E_i, \theta_i, i = 1, 2, \dots, d_f, \rho, \varphi$

2: **Step 1** : For the given T , generate \mathcal{A}_T based GAM point.

Step 2 : For the given T and M , repeat the elements in \mathcal{A}_T to obtain $\mathcal{A}_{M,T}$ according to $\mathcal{P}_{M,T}$ and $\mathcal{T}_{M,T}$.

4: **Step 3** : Perform a permutation of $\mathcal{A}_{M,T}$ according to the criteria given in (20) to obtain the N -dimensional MC, i.e., \mathcal{C}_{MC} .

Step 4 : Generate the j th user's codebook by $\mathcal{X}_j = \mathbf{V}_j \mathcal{A}_j \mathcal{C}_{MC}$.

6: **Step 5** : Jointly optimize the MC and Λ_j by addressing the \mathcal{P}_{2-1} or \mathcal{P}_{2-2} .

Step 6 : Perform the bit labeling with **Algorithm 1** for the optimized codebook \mathcal{X}_j^* , and use the labeling metric given in (34).

D. Labeling design for each user

For each sparse codebook \mathcal{X}_j , the corresponding bit-to-symbol mapping, i.e., bit labeling, needs to be optimized. Similar to the case of permutation, the labeling rules should be designed to minimize the MC constrained ABER. Under Rician fading channels, the labeling metric is given as

$$\Pi(\xi_t) = \sum_{i=1}^{M-1} \underbrace{\sum_{l=1, l \neq i}^M N_{i,l}(\xi_j) \exp\left(\frac{-d_{t,i,l}}{4N_0}\right)}_{\Xi(\mathbf{x}_{j,i})}, \quad (34)$$

where $N_{i,l}(\xi_t)$ denotes the number of different labelling bits between $\mathbf{x}_{j,i}$ and $\mathbf{x}_{j,l}$ based on the considered labelling rule ξ_t , $\Xi(\mathbf{x}_{j,i})$ denotes the cost for the i th codeword of \mathcal{X}_j , and $d_{t,i,l}$ is given as

$$d_{t,i,l} = \sum_{k=1}^K \left\{ \frac{\kappa |x_{i,k}^{(t)} - x_{l,k}^{(t)}|^2}{1 + \kappa + \frac{|x_{i,k}^{(t)} - x_{l,k}^{(t)}|^2}{4N_0}} + 4N_0 \ln \left(1 + \frac{|x_{i,k}^{(t)} - x_{l,k}^{(t)}|^2}{4N_0(1 + \kappa)} \right) \right\}, \quad (35)$$

where $x_{i,k}^{(t)}$ represents the element of \mathbf{x}_i at the k th entry for the labelling rule ξ_t .

Lemma 4: Obviously, let $\kappa \rightarrow \infty$ and $\kappa = 0$ in (35), we can also obtain the equivalent labeling metric [23]

$$\Pi_A(\xi) = \sum_{i=1}^{M-1} \sum_{l=i+1}^M N_{i,l}(\xi) \exp\left(\frac{-\|\mathbf{x}_{j,i} - \mathbf{x}_{j,l}\|}{4N_0}\right), \quad (36)$$

and

$$\Pi_R(\xi) = \sum_{i=1}^{M-1} \sum_{l=i+1}^M \frac{N_{i,l}(\xi)}{\prod_{k' \in \rho(\mathbf{x}_{j,i}, \mathbf{x}_{j,l})} |x_{j,i,k'} - x_{j,l,k'}|^2}, \quad (37)$$

for AWGN and Rayleigh fading channels, respectively.

The optimal labeling assignment for \mathcal{X}_j is a computationally demanding task, especially for $M \geq 8$, hence the BSA is employed again as an effective means of obtaining locally optimal solutions. A modified BSA for SCMA codebooks is described in **Algorithm 1**. In the BSA based labeling design, the arbitrary initial index assignments is used as the input. Then, starting from the first codeword, we calculate

TABLE I: Comparison of the proposed codebooks in terms of Δ_{\min} , MED and I_t .

System setting (M, λ)	Codebooks	Δ_{\min}	MED	I_t (MPA)
(4, 150%)	A _{4,2}	1.16	1.1	1
	A _{4,3}	1.31	1.23	4
	Star-QAM [20]	0.98	0.9	
	GAM [17]	0.74	0.57	
	Huawei [43]	0.67	0.57	
(4, 200%)	A _{4,2}	1.05	0.96	1
	A _{4,3}	0.75	0.64	4
	Star-QAM [20]	0.56	0.48	6
	GAM [17]	0.47	0.43	
(8, 150%)	A _{8,3}	0.62	0.54	2
	A _{8,4}	0.75	0.67	4
	A _{8,5}	0.62	0.53	4
	A _{8,6}	0.51	0.46	5
	A _{8,7}	0.54	0.50	5
	Star-QAM [20]	0.5	0.45	4
	GAM [17]	0.51	0.47	6
(16, 150%)	A _{16,4}	0.32	0.26	1
	A _{16,9}	0.28	0.21	5
	Star-QAM [20]	0.2	0.15	7
	GAM [17]	0.2	0.16	7

the individual cost $\Xi(\mathbf{x}_{j,i})$ for all codewords and sort them out in descending order. Next, the algorithm swaps codeword index with the highest individual cost with another to find a lower total cost. **Algorithm 1** can be performed I_{\max} times with different initial conditions to prevent it from jumping into a local optimum. In particular, the I_{\max} values for $M = 4, 8, 16$ -ary codebooks are given $I_{\max} = 8, 24, 37$, respectively [42]. **Algorithm 1** can significantly reduce the computational complexity of searching labelling pattern. For each M -ary codebook, the computational complexity of exhaustive search can be approximated by $\mathcal{O}(M!)$. In contrast, the complexity of **Algorithm 1** can be estimated as $\mathcal{O}(I_{\max} M^2)$. For example, the reduction in complexity levels reaches 78% and 98% when considering $M = 8$ and $M = 16$, respectively.

To sum up, we have proposed an efficient codebook design with LP numbers in the downlink Rician fading channel and the overall construction procedures are given in **Algorithm 2**.

V. NUMERICAL RESULTS

In this section, we conduct numerical evaluations of the proposed LPCBs for the (4×6) and (5×10) SCMA systems characterized in (4) and (5), respectively. These two systems lead to overloading factors of $\lambda = 150\%$ and $\lambda = 200\%$, respectively. Let $\mathbf{A}_{M,T}$ denote the proposed LPCB of M -ary with T LP numbers. Since the proposed permutation, labeling and sparse codebook optimization criteria are applicable in the high E_b/N_0 regime, we consider $E_b/N_0 = 16$ dB, which is sufficiently large. In general, in an IoT networks with an LoS path, a large κ is assumed for many scenarios [32]–[34], [36], [37]. Hence, we set $\kappa = 20$ for the optimization of the proposed LPCBs. Moreover, considering the trade-off between

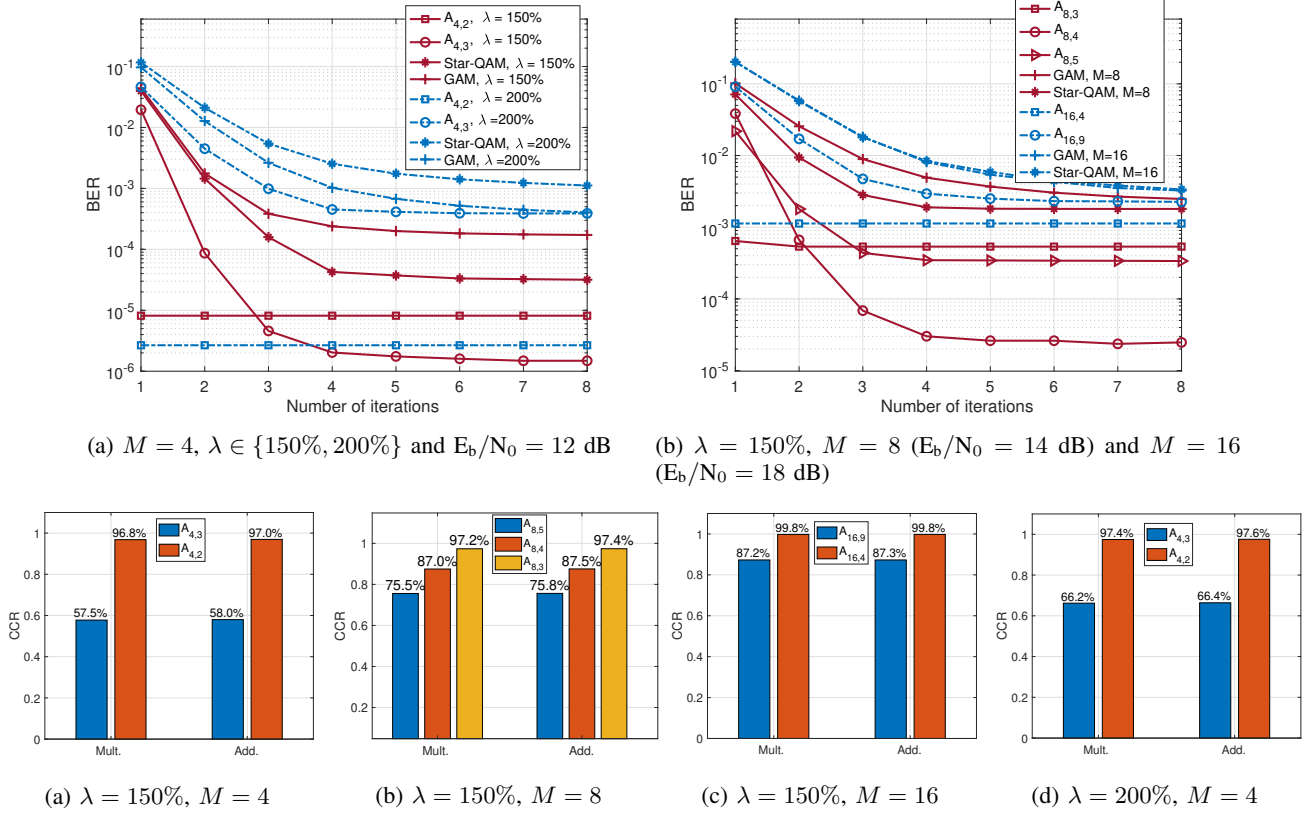


Fig. 4: CCR of the proposed LPCBs to Non-LPCBs with MPA.

the computation complexity and accuracy, we set $Q = 10000$ and $t_{\max} = 20$ for the cases of $M \geq 16$ or $\lambda = 200\%$. The details of the designed LPCBs are given in our GitHub project⁵ and some LPCBs are also presented in Appendix D.

The main state-of-the-art codebooks for comparison with the proposed ones are the GAM codebook [17] and the Star-QAM codebook [20]⁶. This is because both [17] and [20] achieve good BER performance in downlink channels. We first analyze and present the complexity reduction of our proposed LPCBs for MPA decoder. Then, the BER performances are compared with the benchmark codebooks in both uncoded and coded systems.

Table I compares the Δ_{\min} of the proposed codebooks, GAM and Star-QAM for the SCMA systems with overloading $\lambda = 150\%$ and $\lambda = 200\%$, respectively. In addition, since the MED is widely used as a performance metric, we also present the MED of different codebooks. Clearly, most of the proposed LPCBs enjoy larger Δ_{\min} and MED values. For example, the MEDs of proposed codebooks $A_{4,3}$ and $A_{4,2}$ of $\lambda = 150\%$ and $\lambda = 200\%$ SCMA systems, respectively, are significantly larger than that of other codebooks. It is also worth mentioning that, to our best knowledge, the proposed codebooks $A_{4,2}$ for $\lambda = 150\%$, $A_{8,4}$ and $A_{16,4}$ for $\lambda = 200\%$, enjoy the largest MED values among all existing codebooks. Table I also shows that a codebook with less projection number generally

results in larger Δ_{\min} . This is because with smaller T , the superimposed constellation generally owns better distance profile as it has less signal point, which facilitates the joint optimization of the MC and users' constellation operators.

A. Complexity reduction of MPA with the proposed LPCBs

In this subsection, we evaluate the convergence behaviors of MPA with the proposed LPCBs and then investigate the receiver complexities of the proposed LPCBs.

We first present the convergence behaviors of different codebooks with $\kappa \rightarrow \infty$ of MPA decoder in Fig. 3. The number of iterations I_t for decoding convergence of different codebooks are also summarized in Table I. It is noted that most of the proposed codebooks enjoy faster convergence rate than the Star-QAM codebook and GAM codebook⁷. With reduced projected numbers (T), less message propagation between function nodes and variable nodes are required during MPA iteration, thus leading to less iterations for convergence. In particular, the proposed codebooks $A_{4,2}$, $A_{8,3}$ and $A_{16,4}$ for $\lambda = 150\%$, and $A_{4,2}$ for $\lambda = 200\%$ only need about one iteration to converge. The decoding can be carried out in a one-shot for these codebooks and the decoding complexity is significantly reduced compared to other codebooks.

We next evaluate the computation complexity in terms of multiplication and addition of LPCBs in MPA, which can be

⁵<https://github.com/ethanlq/SCMA-codebook>

⁶We have also optimized the MCs that proposed in [28] with the proposed design approach. However, it is found that the resultant BER performance is far inferior to the GAM and Star-QAM codebooks in downlink channels.

⁷For other values of κ , we also observe the same result that the proposed LPCBs require less iterations for convergence.

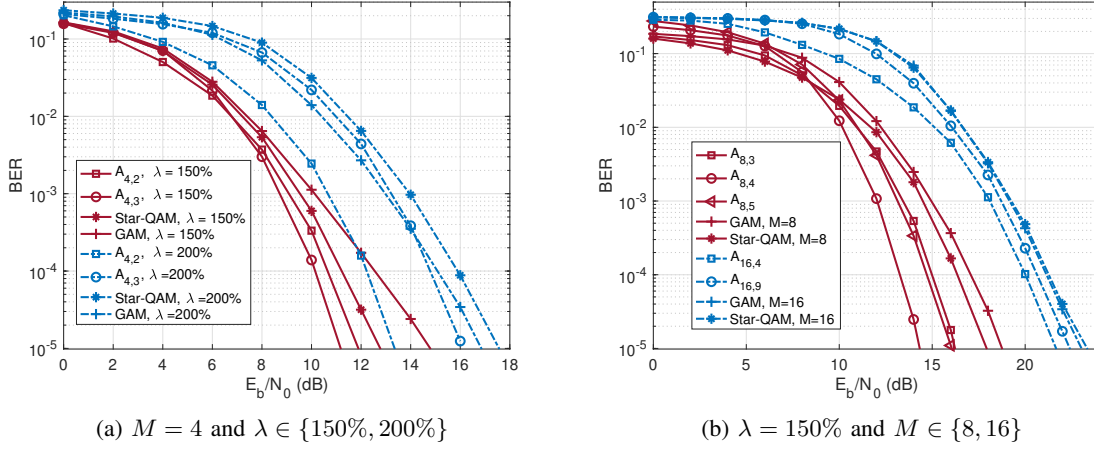


Fig. 5: BER performance of different codebooks with $\kappa \rightarrow \infty$.

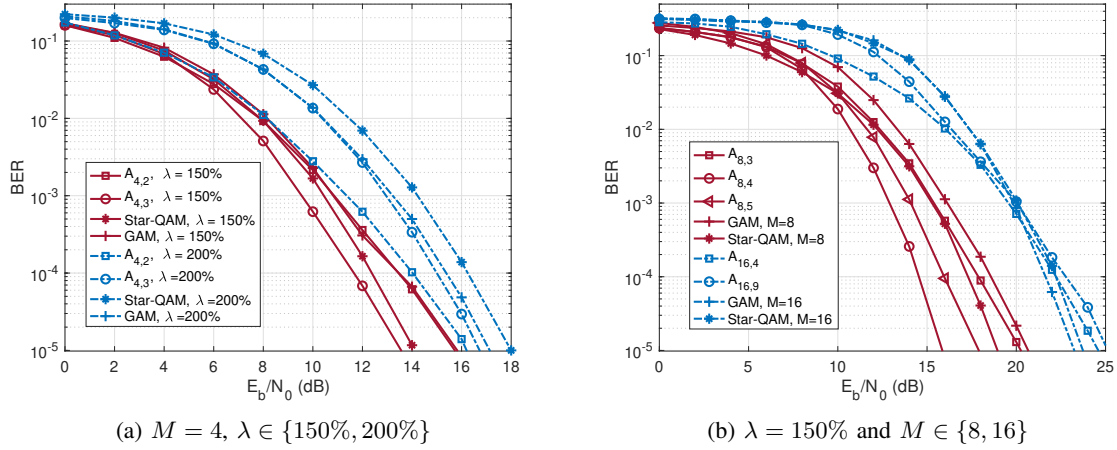


Fig. 6: BER performance of different codebooks with $\kappa = 15$.

approximated as [27], [44]:

$$\begin{aligned} N_m &= [(d_f + 3)T^{d_f}N + (N - 2)TN]JI_t + T(N - 1)J, \\ N_a &= [(d_f + 1)T^{d_f}N + (T - 1)N + (T^{d_f - 1} - 1)TN]JI_t, \end{aligned} \quad (38)$$

respectively. Clearly, the decoding complexity depends on T , I_t and the system overloading. To proceed, we further define the complexity reduction ratio (CRR) as follows [28]:

$$\text{CRR} \triangleq 1 - \frac{\text{Number of Calculations of LPCBs}}{\text{Number of Calculations of Non-LPCBs}}, \quad (39)$$

where the Non-LPCBs are the conventional codebooks [17], [19]–[23], [45], [46]. From (39), it is clear a LPCB with smaller T owns larger CRR. The CRRs of some representative LPCBs and non-LPCBs codebook are compared in Fig. 4, which shows the proposed LPCBs can greatly reduce the detection complexity, with more than about 60% complexity reduction being achieved compared to Non-LPCBs. The gain becomes more significant as the increase of M and λ . For example, the proposed codebooks $A_{8,3}$ and $A_{16,4}$ can reduce the decoding complexity by 97.4% and 99.8%, respectively.

B. Comparison of uncoded BER performance

In this subsection, we evaluate the uncoded BER performance of the proposed LPCBs. $\kappa \rightarrow \infty$ is the typical value for open environments in satellite IoT [32]. In the rural and suburban scenarios of satellite IoT, and UAV IoT networks, large κ is generally assumed [33], [34], [36]. Hence, we consider $\kappa \rightarrow \infty$ and $\kappa = 15$ for the uncoded systems, which are shown in Fig. 5 and Fig. 6, respectively. As can be seen in Fig. 5, the proposed codebook $A_{4,3}$ for $\lambda = 150\%$ achieves about 2 dB gain over the Star-QAM codebook and 4 dB gain over the GAM codebook at $\text{BER} = 10^{-5}$, respectively. Moreover, $A_{4,3}$ for $\lambda = 200\%$ also outperforms than the Star-QAM codebook and the GAM codebook. The excellent BER advantage can be observed for the proposed $A_{4,2}$ codebook for $\lambda = 200\%$ due to large MED, which achieves about 3 dB gain over the Star-QAM codebook and 4.5 dB gain over the GAM codebook at $\text{BER} = 10^{-5}$, respectively. For $M = 8, 16$ and $\lambda = 150\%$, we are interested in codebooks with less projected numbers, thus only $A_{8,3}$, $A_{8,4}$ and $A_{8,5}$ are compared to GAM and Star-QAM codebooks in Fig. 5(b). One can see that the proposed codebooks outperform the GAM and Star-QAM codebooks significantly. Among these codebooks, $A_{8,4}$

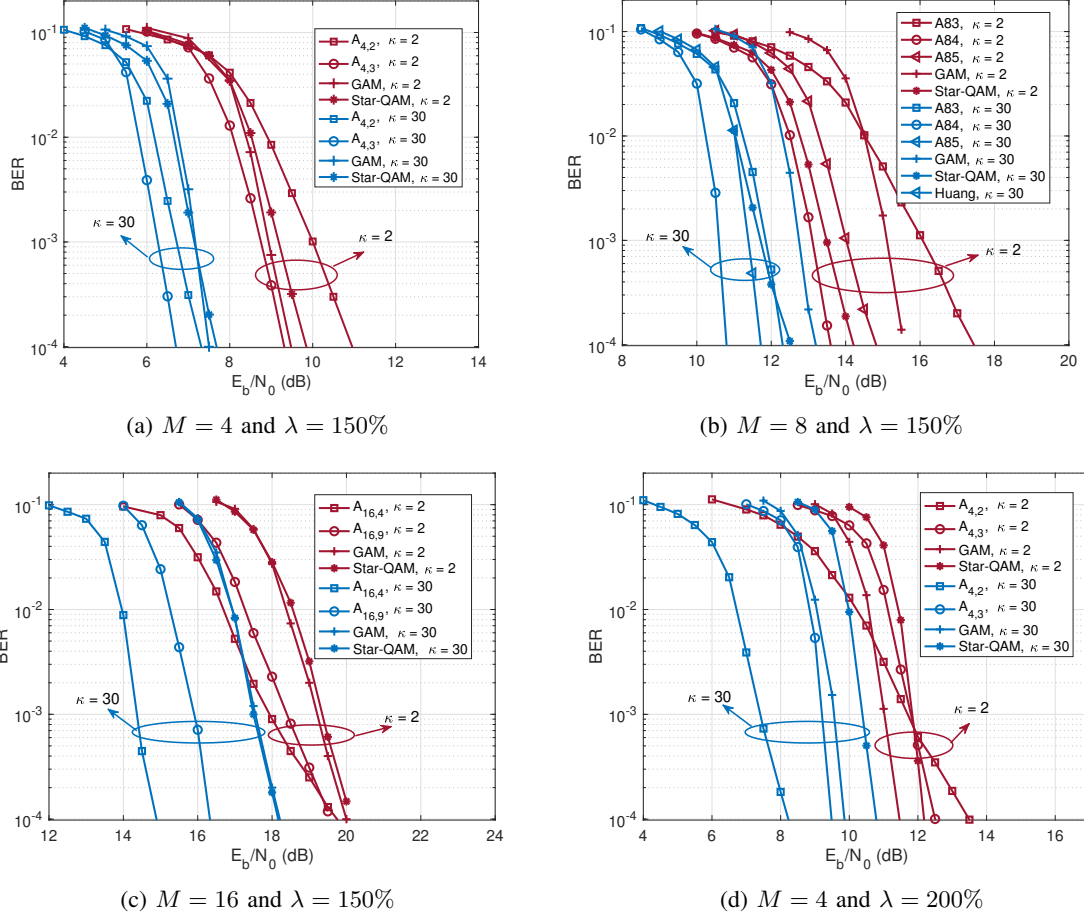


Fig. 7: LDPC coded BER performance comparison of different codebooks.

and $A_{16,4}$ achieve the best BER performance for $M = 8$ and $M = 16$, respectively. In particular, $A_{8,4}$ has a gain about 3.8 dB over the GAM codebook and 4.2 dB over the Star-QAM codebook, respectively.

From Fig. 6, we have the following main observations: 1) For the case of $\kappa = 15$, most of the proposed LPCBs achieve good BER performance. In particular, the codebooks $A_{4,3}$ ($\lambda = 150\%$), $A_{4,2}$ ($\lambda = 200\%$), $A_{8,4}$ still outperform the GAM and Star-QAM codebooks; 2) Compared with the results in Fig. 5, the performance of LPCBs with smallest LP numbers for $\kappa = 15$, such as $A_{4,2}$ ($\lambda = 150\%$), $A_{8,3}$, $A_{16,4}$ are more likely affected by the channel fading due to the constellation overlapping.

C. Comparison of coded BER performance

In this subsection, we compare the coded BER performances of the proposed codebooks with the GAM and Star-QAM codebooks. In some scenarios, such as dense urban and urban areas with an LoS path, κ is relatively small [32]–[34]. Hence, we set κ with a wide range. Considering the short-packet nature of IoT networks, we apply the 5G NR LDPC codes with small block length [4], as specified in TS38.212 [47]. The detailed simulation settings are given in Table II and the simulated BER performance are depicted in Fig. 7.

In Fig. 7(a), the proposed codebook $A_{4,2}$ and $A_{4,3}$ show better performance than the GAM and Star-QAM codebooks

TABLE II: Simulation Parameters

Parameters	Values
Transmission	Downlink
Channel model	Rician fading channel, $\kappa = 2, 30$
SCMA setting	$\lambda = 150\%$ and $\lambda = 200\%$
Channel coding	5G NR LDPC codes with rate of 0.8462 and block length of 260
Codebooks	The same codebooks presented in Fig. 5 and Fig. 6
Receiver	Turbo-MPA: 1 MPA iteration for $A_{4,2}$ ($\lambda = 150\%$), $A_{4,2}$ ($\lambda = 200\%$), $A_{8,3}$ and $A_{16,4}$, and 4 MPA iterations for other codebooks, 5 BICM iterations

for $\kappa = 30$. In particular, the $A_{4,3}$ outperforms Star-QAM codebook by 1 dB gain at $\text{BER} = 10^{-4}$. When there exists severe fading, i.e., $\kappa = 2$, the error performance of $A_{4,2}$ deteriorates, however, the proposed $A_{4,3}$ still achieves similar BER performance with the Star-QAM codebook and performs slightly better in the low E_b/N_0 range. By observing the slopes of the curves, the GAM codebook enjoys a larger diversity gain than the low projected codebooks. This is because some constellation points in the LPCBs are overlapped in order to reduce the decoding complexity, thus the diversity between these constellation pairs may be reduced. However, we show that by properly designing the MC and maximising the coding gain, most of the proposed LPCBs can still achieve good BER

performance for $\kappa = 2$.

For the 8-ary and 16-ary codebooks shown in Fig. 7(b) and Fig. 7(c), respectively, BER trends similar to that of the 4-ary codebook are observed. The proposed codebooks outperform the GAM and Star-QAM codebooks for $\kappa = 30$, whereas the error performance of the proposed codebooks with least projection numbers deteriorate for the case of $\kappa = 2$. It is interesting to see the proposed codebook $A_{8,4}$ and $A_{16,4}$ achieve the best coded BER performance for both $\kappa = 30$ and $\kappa = 2$ among all the codebooks owing to the well optimized bit labeling and large coding gain. For example, $A_{8,4}$ achieves about 3.6 dB gain over the GAM codebook and 1.5 dB gain over the Star-QAM codebook for $\kappa = 30$, and achieves about 2 dB gain over the GAM codebook and 1 dB gain over the Star-QAM codebook for $\kappa = 2$ at $\text{BER} = 10^{-4}$, respectively.

As shown in Fig. 7(d), for the SCMA system with $\lambda = 200\%$, the proposed codebook $A_{4,2}$ achieves about 2 dB gain compared to the GAM with $\kappa = 30$ due to the very large coding gain. The coded BER curve of $A_{4,2}$ displays slower slope than that of GAM and Star-QAM codebooks, implying $A_{4,2}$ enjoys smaller diversity gain than that of GAM and Star-QAM codebooks. This is reasonable since the diversity gain of $A_{4,2}$ is sacrificed due to the constellation overlapping for complexity reduction. However, the coding gain for the proposed $A_{4,2}$ is significantly improved. This can be seen by the E_b/N_0 point at the BER of 10^{-1} . The detailed discussion on the coding gain and diversity gain over Rician fading channels is given in Appendix A. The superiority of the proposed codebooks mainly lies in the huge reduction of complexity, especially in largely overloaded system.

VI. CONCLUSION

In this paper, we have introduced a novel class of SCMA codebooks that can significantly reduce the decoding complexity while achieving good BER performance for downlink IoT networks. We have analyzed the corresponding pairwise error probability of SCMA transmissions under a generalized Rician fading channel model and derived the codebook design criteria. To reduce the decoding complexity, we have proposed to construct the MC with LP-GAM by allowing several overlapped constellation points in each dimension. In addition, we have developed an efficient codebook design approach, including permutation, bit labeling, sparse codebook construction and parameter optimization. Numerical results demonstrated the benefits of proposed codebook in terms of complexity reduction and BER performance improvements in different IoT environments. In particular, some codebooks (e.g., $A_{4,2}$, $A_{4,3}$, $A_{8,4}$ and $A_{16,4}$ for $\lambda = 150\%$ system and $A_{4,2}$ for $\lambda = 200\%$ system) achieve significant performance improvements for typical κ values in open, rural and suburban environments, and significant complexity reduction compared to the existing codebooks.

APPENDIX A PROOF OF THE LEMMA 1

1). For $(1 + \kappa) \gg \frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0}$, we have

$$d_{1,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k) = \frac{\kappa \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{\kappa + 1}, d_{2,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k) = 0, \quad (40)$$

and thus $d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2 = \frac{\kappa \delta_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}}{1 + \kappa}$. In this special case, (13) reduces to

$$\Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}}\} \leq \exp\left(-\frac{\kappa \delta_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}}{4(1 + \kappa)N_0}\right). \quad (41)$$

Then, based on (15) and (41), we know that the ABER is dominated by the MED between \mathbf{w} and $\hat{\mathbf{w}}$ in Φ_{M^J} , which corresponding to the MEEs that all the errors occurred with multiple users. Therefore, an important codebook design criteria in this case is to maximize the MED of the superimposed codewords, which has been widely employed for SCMA systems operating in the AWGN channel [22], [26], [46]. Note that the MED can also obtained by letting $\kappa \rightarrow \infty$ in $\Delta_{\min}(\mathcal{X})$.

2). In the case when $(1 + \kappa) \ll \frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0}$, we have

$$d_{1,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k) = 4N_0\kappa, d_{2,\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2(k) = 4N_0 \ln\left(-\frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0(1 + \kappa)}\right), \quad (42)$$

and thus $d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2$ is the sum of the logarithms of the element-wise distance. Substituting (42) and (14) into (13), we obtain the PEP for this special case as

$$\Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}}\} \leq \prod_{k \in D(\mathbf{w} \rightarrow \hat{\mathbf{w}})} \left(\frac{\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)}{4N_0(1 + \kappa)}\right)^{-1} \exp(-\kappa). \quad (43)$$

where $D(\mathbf{w} \rightarrow \hat{\mathbf{w}})$ denotes the set of indices in which $\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k) \neq 0$. Note that one can approximate the right-hand side of (43) as

$$\Pr\{\mathbf{w} \rightarrow \hat{\mathbf{w}}\} \leq G_c(\mathbf{w} \rightarrow \hat{\mathbf{w}}) \left(\frac{1}{4N_0(1 + \kappa)}\right)^{-G_d(\mathbf{w} \rightarrow \hat{\mathbf{w}})}, \quad (44)$$

where $G_d(\mathbf{w} \rightarrow \hat{\mathbf{w}})$ is the cardinality of set $D(\mathbf{w} \rightarrow \hat{\mathbf{w}})$ and

$$G_c(\mathbf{w} \rightarrow \hat{\mathbf{w}}) = \prod_{k \in D(\mathbf{w} \rightarrow \hat{\mathbf{w}})} \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)^{-1} \exp(-\kappa). \quad (45)$$

Let $G_d \triangleq \min_{\mathbf{w} \neq \hat{\mathbf{w}}} G_d(\mathbf{w} \rightarrow \hat{\mathbf{w}})$ and $G_c \triangleq \min_{\mathbf{w} \neq \hat{\mathbf{w}}} G_c(\mathbf{w} \rightarrow \hat{\mathbf{w}})$, where G_d and G_c are the so-called diversity gain and coding gain [40]. For $\kappa \rightarrow 0$, from (44) we observe the PEP decays with the slope of $(1/(\kappa + 1)N_0)^{-G_d}$ and the system achieves the diversity order of G_d . The ABER is mainly dominated by the SEE that all the codewords are detected correctly except for a codeword of only one user. In other words, for the pairs \mathbf{w} and $\hat{\mathbf{w}}$ such that $G_d(\mathbf{w} \rightarrow \hat{\mathbf{w}}) = G_d$, the term $\prod_{k \in D(\mathbf{w} \rightarrow \hat{\mathbf{w}})} \tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k)^{-1}$ in (45) equals to the MPD of the sparse codebook. The MPD of the codebook is given by

$$\text{MPD}(\mathcal{X}) = \min_{1 \leq j \leq J} \min_{i \neq l} \prod_{k \in \rho(\mathbf{x}_i^{(j)}, \mathbf{x}_l^{(j)})} |x_{k,i}^{(j)} - x_{k,l}^{(j)}|^2, \quad (46)$$

where $x_{k,i}^{(j)}$ is the j th user's i th codeword at k th entry and $\rho(\mathbf{x}_i^{(j)}, \mathbf{x}_l^{(j)})$ denotes the set of indices in which $x_{k,i}^{(j)} \neq x_{k,l}^{(j)}$. Similarly, in the case when $G_d(\mathbf{w} \rightarrow \hat{\mathbf{w}}) = G_d$ and let $\kappa = 0$ (Rayleigh fading channel), $\Delta_{\min}(\mathcal{X})$ reduces to

$$\begin{aligned} \Delta_{\min}^{\kappa=0}(\mathcal{X}) &= \min_{1 \leq j \leq J} \min_{i \neq l} \sum_{k=1}^K 4N_0 \ln \left(1 + \frac{|x_{k,i} - x_{k,l}|^2}{4N_0}\right) \\ &\approx 4N_0 \ln(\text{MPD}(\mathcal{X})(4N_0)^{-G_d}). \end{aligned} \quad (47)$$

For $\kappa = 0$, (47) indicates that to improve the ABER, it is

idea to improve the G_d first and then maximize the $\text{MPD}(\mathcal{X})$ [40].

This completes the proof of Lemma 1.

APPENDIX B

PROOF OF THE LEMMA 2

For an M -ary \mathcal{C}_{MC} , the primary design rule is that the resultant MC should be uniquely separable. Namely, giving arbitrary two vectors $\bar{\mathbf{c}}_i$ and $\bar{\mathbf{c}}_l$ in \mathcal{C}_{MC} , we have $\bar{\mathbf{c}}_i \neq \bar{\mathbf{c}}_l$ with $i \neq l$. Hence, for the basic constellation with T distinct points, the N -dimensional constellation with maximum size can be constructed is T^N , where construction is given by the Cartesian product over the basic constellation. For example, consider $T = 2$ and $N = 2$ and assume that the designed basic constellation $\mathcal{A}_2 = \{-a, +a\} \in \mathbb{C}$, obviously, the unique decodable 2-dimensional constellation with maximum size can be generated as $\mathcal{A}_2 \times \mathcal{A}_2 \rightarrow \mathcal{C}_{MC}$, i.e.,

$$\mathcal{C}_{MC} = \begin{bmatrix} -a & +a & -a & +a \\ -a & -a & +a & +a \end{bmatrix}. \quad (48)$$

Therefore, for the given M and N , the LP number should satisfy $\lceil \sqrt[N]{M} \rceil \leq T \leq M$, and for the case of $T = \sqrt[N]{M} \in \mathbb{Z}$, the construction of \mathcal{C}_{MC} is given by the Cartesian product of \mathcal{A}_T .

This completes the proof of Lemma 2.

APPENDIX C

PROOF OF THE LEMMA 4

We first show that Φ_{M^J} can be expressed as the Cartesian product of \mathcal{S}_{sum} . To proceed, we rewrite the constellation superposition set as

$$\Phi_{M^J} = \{\mathbf{w} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_J | \forall \mathbf{x}_j \in \mathcal{X}_j, j = 1, 2, \dots, J\}, \quad (49)$$

where the mapping rule is given by $g_\Phi : \mathcal{X}_1 + \cdots + \mathcal{X}_J \rightarrow \Phi_{M^J}$. In fact, the constellation superposition process is characterized by a Cartesian product mapping. Similarly, we have $g_S : z_1 \mathcal{A}_T + \cdots + z_{d_f} \mathcal{A}_T \rightarrow \mathcal{S}_{\text{sum}}$. The lemma 2 reveals that for $T = \sqrt[N]{M}$, the mapping $f_{MC} : \mathcal{A}_T \times \cdots \times \mathcal{A}_T \rightarrow \mathcal{C}_{MC}$ is a Cartesian product. Accordingly, the mapping $f_{C_j} : z_1^{(j)} \mathcal{A}_T \times \cdots \times z_N^{(j)} \mathcal{A}_T \rightarrow \mathcal{C}_j$ is also a Cartesian product, where $z_n^{(j)}$ is the j th user's constellation operator. Based on f_{C_j} , we can rewrite g_Φ as

$$g_\Phi : \mathbf{V}_1 f_{C_j} + \cdots + \mathbf{V}_J f_{C_j} \rightarrow \Phi_{M^J}. \quad (50)$$

(50) indicates that Φ_{M^J} is obtained by two steps: 1) Mapping the 1-dimension LP-GAM \mathcal{A}_T to \mathcal{C}_j ; 2) Constellation superposition. Since the two steps are both characterized by the Cartesian product, the changes in the order of the two steps will not affect the result of Φ_{M^J} . Namely, Φ_{M^J} can also be obtained by: 1) Constellation superposition based on LP-GAM, i.e., $g_S : z_1 \mathcal{A}_T + \cdots + z_{d_f} \mathcal{A}_T \rightarrow \mathcal{S}_{\text{sum}}$; 2) Obtain Φ_{M^J} by performing Cartesian product over \mathcal{S}_{sum} , i.e.,

$$f_\Phi : \mathcal{S}_{\text{sum}}^{(1)} \times \cdots \times \mathcal{S}_{\text{sum}}^{(K)} \rightarrow \Phi_{M^J}. \quad (51)$$

Next, we show that maximize $\text{MED}(\mathcal{S}_{\text{sum}})$ is equivalent to maximize $\Delta_{\min}(\mathcal{X}_{\lambda, M, T})$. Let \mathbf{w} and $\hat{\mathbf{w}}$ be two superimposed

codeword vectors, and define

$$G \triangleq \min_{\mathbf{w}, \hat{\mathbf{w}}, \mathbf{w} \neq \hat{\mathbf{w}}} \underbrace{\sum_{k=1}^K \text{Ind}(\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k))}_{G(\mathbf{w} \rightarrow \hat{\mathbf{w}})}, \quad (52)$$

where $\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k) = |w(k) - \hat{w}(k)|^2$, and $w(k)$ is the k th entry of \mathbf{w} . Obviously, since Φ_{M^J} can be generated by the Cartesian product of \mathcal{S}_{sum} , we have $G = 1$ and the MED of Φ_{M^J} equals to $\text{MED}(\mathcal{S}_{\text{sum}})$. Assume \mathbf{w}_m and \mathbf{w}_n are the two vectors that achieve the MED value, and $\tau_{\mathbf{w}_m \rightarrow \mathbf{w}_n}(k') \neq 0$. Accordingly, we obtain $\tau_{\mathbf{w}_m \rightarrow \mathbf{w}_n}(k') = \text{MED}(\mathcal{S}_{\text{sum}})$. This also indicates that for arbitrary two vectors \mathbf{w} and $\hat{\mathbf{w}}$, $\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k) = 0$ or $\tau_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}(k) \geq \text{MED}(\mathcal{S}_{\text{sum}})$, $k = 1, 2, \dots, K$.

Based on (14), $d_{\mathbf{w} \rightarrow \hat{\mathbf{w}}}^2$ is an increasing function of $G(\mathbf{w} \rightarrow \hat{\mathbf{w}})$ and $\tau_{\mathbf{w}_m \rightarrow \mathbf{w}_n}(k)$. Hence, (29) achieves the minimum value at \mathbf{w}_m and \mathbf{w}_n , where $G = 1$ and $\tau_{\mathbf{w}_m \rightarrow \mathbf{w}_n}(k') = \text{MED}(\mathcal{S}_{\text{sum}})$. Therefore, improving $\text{MED}(\mathcal{S}_{\text{sum}})$ is equivalent to increasing $\Delta_{\min}(\mathcal{X}_{\lambda, M, T})$.

This completes the proof of Lemma 4.

APPENDIX D

THE PROPOSED LPCBS

The proposed LPCBs of $A_{4,3}$ ($\lambda = 150\%$) is presented, and the other LPCBs are at our GitHub project: <https://github.com/ethanlq/SCMA-codebook>.

$$\begin{aligned} \mathcal{X}_1 &= \begin{bmatrix} 0 & 0.0850 + 1.0324i & 0 & 0 \\ 0 & 0 & 0 & 1.0841 + 0.0000i \\ 0 & 0 & 0 & -1.0841 + 0.0000i \\ 0 & -0.0850 - 1.0324i & 0 & 0 \end{bmatrix}^T, \\ \mathcal{X}_2 &= \begin{bmatrix} 0.0850 + 1.0324i & 0 & 0 & 0 \\ 0 & 0 & 1.0841 + 0.0000i & 0 \\ 0 & 0 & -1.0841 + 0.0000i & 0 \\ -0.0850 - 1.0324i & 0 & 0 & 0 \end{bmatrix}^T, \\ \mathcal{X}_3 &= \begin{bmatrix} -0.7156 + 0.4894i & 0 & 0 & 0 \\ 0 & -0.7156 + 0.4894i & 0 & 0 \\ 0 & 0 & 0.7156 - 0.4894i & 0 \\ 0.7156 - 0.4894i & 0 & 0 & 0 \end{bmatrix}^T, \\ \mathcal{X}_4 &= \begin{bmatrix} 0 & 0 & -0.7156 + 0.4894i & 0 \\ 0 & 0 & 0 & -0.7156 + 0.4894i \\ 0 & 0 & 0.7156 - 0.4894i & 0 \\ 0 & 0 & 0.7156 - 0.4894i & 0 \end{bmatrix}^T, \\ \mathcal{X}_5 &= \begin{bmatrix} 1.0841 + 0.0000i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0850 + 1.0324i \\ 0 & 0 & 0 & -0.0850 - 1.0324i \\ -1.0841 + 0.0000i & 0 & 0 & 0 \end{bmatrix}^T, \\ \mathcal{X}_6 &= \begin{bmatrix} 0 & 1.0841 + 0.0000i & 0 & 0 \\ 0 & 0 & 0.0850 + 1.0324i & 0 \\ 0 & 0 & -0.0850 - 1.0324i & 0 \\ 0 & -1.0841 + 0.0000i & 0 & 0 \end{bmatrix}^T. \end{aligned}$$

REFERENCES

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys & Tut.*, vol. 17, no. 4, pp. 2347–2376, June. 2015.
- [2] L. Yu, Z. Liu, M. Wen, D. Cai, S. Dang, Y. Wang, and P. Xiao, "Sparse code multiple access for 6G wireless communication networks: Recent advances and future directions," *IEEE Commun. Stand. Mag.*, vol. 5, no. 2, pp. 92–99, Jun. 2021.

- [3] 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. Surveys Tuts*, vol. 19, no. 2, pp. 721–742, Oct. 2016.
- [4] Z. Liu and L.-L. Yang, "Sparse or dense: A comparative study of code-domain NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4768–4780, Aug. 2021.
- [5] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE VTC*, Vancouver, Canada, Dec. 2014, pp. 1–5.
- [6] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE J. Sel. Areas Commun.*, vol. 56, no. 4, pp. 1616–1626, Mar. 2008.
- [7] Q. Luo *et al.*, "An error rate comparison of power domain non-orthogonal multiple access and sparse code multiple access," *IEEE Open J. Commun. Soc.*, vol. 2, no. 4, pp. 500–511, Mar. 2021.
- [8] J. Zeng, T. Lv, Z. Lin, R. P. Liu, J. Mei, W. Ni, and Y. J. Guo, "Achieving ultrareliable and low-latency communications in IoT by FD-SCMA," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 363–378, Oct. 2020.
- [9] S. Moon, H.-S. Lee, and J.-W. Lee, "SARA: Sparse code multiple access-applied random access for IoT devices," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3160–3174, May 2018.
- [10] L. Miuccio, D. Panno, and S. Riolo, "Joint control of random access and dynamic uplink resource dimensioning for massive MTC in 5G NR based on SCMA," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5042–5063, Feb. 2020.
- [11] K. Lai, J. Lei, Y. Deng, L. Wen, G. Chen, and W. Liu, "Analyzing uplink grant-free sparse code multiple access system in massive IoT networks," *IEEE Internet Things J.*, pp. 1–1, Sep. 2021.
- [12] J. Zhang, Q. He, Z. Yu, B. Bai, and M. Zhu, "Low-complexity coherent iterative receiver for SCMA-based LEO satellite communications," in *2019 IEEE GLOBECOM*, 2019, pp. 1–6.
- [13] J. Wang, C. Jiang, and L. Kuang, "Iterative NOMA detection for multiple access in satellite high-mobility communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1101–1113, Apr. 2022.
- [14] Z. Zhang, Y. Li, C. Huang, Q. Guo, L. Liu, C. Yuen, and Y. L. Guan, "User activity detection and channel estimation for grant-free random access in LEO satellite-enabled internet of things," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8811–8825, Sep. 2020.
- [15] P. Li, G. Cui, and W. Wang, "Asynchronous flipped grant-free SCMA for satellite-based Internet of things communication networks," *Applied Sciences*, vol. 9, no. 2, p. 335, Jan. 2019.
- [16] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE PIMRC*, London, U.K., Sep. 2013, pp. 332–336.
- [17] Z. Mheich, L. Wen, P. Xiao, and A. Maaref, "Design of SCMA codebooks based on golden angle modulation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1501–1509, Dec. 2018.
- [18] D. Cai, P. Fan, X. Lei, Y. Liu, and D. Chen, "Multi-dimensional SCMA codebook design based on constellation rotation and interleaving," in *Proc. IEEE VTC*, Nanjing, China, Jul. 2016, pp. 1–5.
- [19] K. Deka, M. Priyadarsini, S. Sharma, and B. Beferull-Lozano, "Design of SCMA codebooks using differential evolution," in *Proc. IEEE ICC*, Dublin, Ireland, Jul. 2020, pp. 1–7.
- [20] L. Yu, P. Fan, D. Cai, and Z. Ma, "Design and analysis of SCMA codebook based on Star-QAM signaling constellations," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10543–10553, Aug. 2018.
- [21] X. Zhang, D. Zhang, L. Yang, G. Han, H.-H. Chen, and D. Zhang, "SCMA codebook design based on uniquely decomposable constellation groups," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4828–4842, Mar. 2021.
- [22] X. Li, Z. Gao, Y. Gui, Z. Liu, P. Xiao, and L. Yu, "Design of power-imbalanced SCMA codebook," *IEEE Tran. Veh. Technol.*, vol. 71, no. 2, pp. 2140–2145, Feb. 2022.
- [23] Y.-M. Chen and J.-W. Chen, "On the design of near-optimal sparse code multiple access codebooks," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2950–2962, Feb. 2020.
- [24] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 502–518, Mar. 1996.
- [25] H. Wen, Z. Liu, Q. Luo, C. Shi, and P. Xiao, "Designing enhanced multi-dimensional constellations for code-domain NOMA," *IEEE Wireless Commun. Lett.*, pp. 1–1, Jul. 2022.
- [26] C. Huang, B. Su, T. Lin, and Y. Huang, "Downlink SCMA codebook design with low error rate by maximizing minimum euclidean distance of superimposed codewords," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5231–5245, Mar. 2022.
- [27] A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, and J. Ma, "Low complexity techniques for SCMA detection," in *Proc. IEEE GLOBECOM*, San Diego, CA, USA, Feb. 2015, pp. 1–6.
- [28] J. Bao, Z. Ma, M. Xiao, T. A. Tsiftsis, and Z. Zhu, "Bit-interleaved coded SCMA with iterative multiuser detection: Multidimensional constellations design," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5292–5304, Dec. 2017.
- [29] O. Hahm, E. Baccelli, H. Petersen, and N. Tsiftes, "Operating systems for low-end devices in the Internet of things: A survey," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 720–734, Dec. 2016.
- [30] J. Zhang, J. Fan, J. Zhang, D. W. K. Ng, Q. Sun, and B. Ai, "Performance analysis and optimization of noma-based cell-free massive MIMO for IoT," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9625–9639, 2022.
- [31] G. Yu, X. Chen, and D. W. K. Ng, "Low-cost design of massive access for cellular internet of things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8008–8020, 2019.
- [32] 3GPP TS 38.811, Rel. 15, "5G NR, study on new radio (NR) to support non-terrestrial networks," Jul. 2018. [Online]. Available: <https://www.3gpp.org/ftp/Specs/archive/38series/38.811>
- [33] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel-recording, statistics, and channel model," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 375–386, May 1991.
- [34] B. Vucetic and J. Du, "Channel modeling and simulation in satellite mobile communication systems," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 8, pp. 1209–1218, Oct. 1992.
- [35] Y. Gui, L. Zhu, and J. Liu, "SCMA secure communication scheme for satellite system based on distance spectrum," in *Proc. IEEE WCNC*, Nanjing, China: IEEE, 2021, pp. 1–6.
- [36] T. Z. H. Ernest, A. S. Madhukumar, R. P. Sirigina, and A. K. Krishna, "NOMA-aided UAV communications over correlated rician shadowed fading channels," *IEEE Trans. Signal Processing*, vol. 68, pp. 3103–3116, May 2020.
- [37] Y. Wang, Z. Gao, J. Zhang, X. Cao, D. Zheng, Y. Gao, D. W. K. Ng, and M. D. Renzo, "Trajectory design for UAV-based internet of things data collection: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3899–3912, Feb. 2022.
- [38] P. Larsson, "Golden angle modulation," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 98–101, May 2017.
- [39] Q. Luo, Z. Liu, G. Chen, Y. Ma, and P. Xiao, "A novel multi-task learning empowered codebook design for downlink SCMA networks," *IEEE Wireless Commun. Lett.*, 2022.
- [40] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Mar. 2003.
- [41] C. Tellambura, "Evaluation of the exact union bound for trellis-coded modulations over fading channels," *IEEE Trans. Commun.*, vol. 44, no. 12, pp. 1693–1699, Dec. 1996.
- [42] K. Zeger and A. Gersho, "Pseudo-gray coding," *IEEE Trans. commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.
- [43] Altera Innovate Asia website, Presentation, "1st 5G algorithm innovation competition-env1.0-SCMA." [Online]. Available: <http://www.innovateasia.com/5G/en/gp2.html>
- [44] L. Yang, Y. Liu, and Y. Siu, "Low complexity message passing algorithm for SCMA system," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2466–2469, Sep. 2016.
- [45] H. Yan, H. Zhao, Z. Lv, and H. Yang, "A top-down SCMA codebook design scheme based on lattice theory," in *Proc. IEEE PIMRC*, Valencia, Spain, Dec. 2016, pp. 1–5.
- [46] V. P. Klimentyev and A. B. Sergienko, "SCMA codebooks optimization based on genetic algorithm," in *Proc. 23th European Wireless Conf.*, Dresden, Germany, Aug. 2017, pp. 1–6.
- [47] 3GPP TS 38.212, Rel. 15, "5G NR, multiplexing and channel coding," Jul. 2018. [Online]. Available: <https://www.etsi.org/deliver/etsits/138200138299/138212/15.02.0060/ts138212v150200p.pdf>