Downlink SCMA Codebook Design With Low Error Rate by Maximizing Minimum Euclidean Distance of Superimposed Codewords

Chinwei Huang , *Student Member, IEEE*, Borching Su , *Member, IEEE*, Tingyi Lin,
and Yenming Huang , *Member, IEEE*

Abstract—Sparse code multiple access (SCMA), as a codebook-based non-orthogonal multiple access (NOMA) technique, has received research attention in recent years. The codebook design problem for SCMA has also been studied to some extent since codebook choices are highly related to the system's error rate performance. In this paper, we approach the SCMA codebook design problem by formulating an optimization problem to maximize the minimum Euclidean distance (MED) of superimposed codewords under power constraints. While SCMA codebooks with a larger MED are expected to obtain a better BER performance, no optimal SCMA codebook in terms of MED maximization, to the authors' best knowledge, has been reported in the SCMA literature yet. In this paper, a new iterative algorithm based on alternating maximization with exact penalty is proposed for the MED maximization problem. The proposed algorithm, when supplied with appropriate initial points and parameters, achieves a set of codebooks of all users whose MED is larger than any previously reported results. A Lagrange dual problem is derived which provides an upper bound of MED of any set of codebooks. Even though there is still a nonzero gap between the achieved MED and the upper bound given by the dual problem, simulation results demonstrate clear advantages in error rate performances of the proposed set of codebooks over all existing ones not only in AWGN channels but also in some downlink scenarios that fit in 5G/NR applications, making it a good codebook candidate thereof. The proposed set of SCMA codebooks, however, are not shown to outperform existing ones in uplink channels or in the case where non-consecutive OFDMA subcarriers are used. The correctness and accuracy of error curves in the simulation results are further confirmed by the coincidences with the theoretical upper bounds of error rates derived for any given set of codebooks.

Index Terms—5G, mMTC, non-orthogonal multiple access (NOMA), sparse code multiple access (SCMA), optimization, minimum Euclidean distance (MED), semidefinite relaxation (SDR), alternating maximization, exact penalty.

I. INTRODUCTION

IN the fifth generation (5G) wireless communications and beyond, to enable the massive connectivity and high spectral efficiency for the Internet of Things (IoT) and the Factories of the Future (FoF), non-orthogonal multiple access (NOMA) [1] is considered an important multiple access scheme due to its extended spatial efficiency as opposed to the conventional orthogonal multiple access (OMA), such as orthogonal frequency-division multiplexing (OFDM). Among the many existing schemes in NOMA [2], SCMA is regarded as a promising multiple access scheme [3]–[5]. Sparse code multiple access (SCMA) [6] is one kind of code-domain NOMA, which distinguishes multiple users with the aid of codewords [7]. Incoming bits are directly mapped to multi-dimensional codewords of some set of SCMA codebooks, so the codebook design dominates the performance of the SCMA-based NOMA system [8].

Codebook design for SCMA has been studied extensively in the past few years [9]–[27]. The overall design goal is to find a set of codebooks that result in a good performance, in terms of low error rate or large spectral efficiency, in the scenarios of AWGN, uplink, and downlink fading channels. One of the major approaches for this goal is to find codebooks that have a large minimum Euclidean distance (MED) [10]–[12], [15], [17], [26]. The basic rationale behind the approaches of maximizing MED is that a codebook with a larger MED usually results in better error rate performance and we choose the approach of MED maximization in this paper. In this regard, pioneering works including Yu *et al.* [25], and many others [13], [16], [17], [22] considered multi-stage design approaches by first constructing a mother constellation (MC) and then letting every user apply the mother constellation with different rotation and permutation operations and occupy different resources. Under this multi-stage design approach, the maximization of MED of mother constellation has been considered as an important issue and has been studied to some extents [10], [12], [15]. While this approach has rather a simple complexity in the optimization problem, the fact that codebooks of all users are tied to a fixed mother constellation implicitly impose extra and probably unnecessary constraints to the choice of codebooks, and may lead to a suboptimal codebook design solution.

More recently, the idea of MED maximization is studied with a newer definition of MED, namely, the MED of superimposed codewords [11], [17], [26]. Many previously reported codebook design methods [11], [13], [17], [25]–[27] have used MED of superimposed codewords as one design KPI. However, few of the previous works have directly maximized the MED of

superimposed codewords with only power constraints and the reason may be the overwhelming complexity while dealing with this non-convex optimization problem. Although an MED maximization problem has been formulated in [11], the algorithm proposed therein does not guarantee to obtain the optimal point. In fact, obtaining the set of codebooks with the maximal MED is still an open question today, nor has an upper bound of the maximal MED been known yet. For convenience, hereafter we use the term "MED" as the MED of superimposed codewords, rather than of the mother constellation, throughout this paper.

In this work, we propose to approach the SCMA codebook design problem by maximizing the MED of the designed set of codebooks. The major contributions of the paper are summarized as follows: 1) A new method is proposed to deal with the non-convex optimization problem based on the exact penalty technique [42] with an alternating maximization [41] approach. 2) The aforementioned method achieves a codebook design that has a larger MED than any previously reported design, which also shows the best error rate performances among all existing codebooks in some cases, including the one which fits in the downlink scenario of 5G/NR applications. 3) A theoretical upper bound for MED of any possible codebook designs, that was not known before, is obtained by deriving a Lagrange dual problem of the main problem.

The rest of the paper is organized as follows. Section II describes the downlink SCMA system model based on OFDMA. In Section III, the MED maximization problem is formulated and a corresponding algorithm is proposed. In Section IV, the dual problem of the MED maximization problem is derived. The numerical results are given in Section V. Finally, conclusions are made in Section VI.

A. Notations

Boldfaced lower case letters such as $\mathbf{x}$ represent column vectors, boldfaced upper case letters such as $\mathbf{X}$ represent matrices, and italic letters such as x, X represent scalars. Superscripts as in $\mathbf{X}^T, \mathbf{X}^H$, and $\mathbf{X}^{-1}$ denote the transpose, transpose-conjugate, and inverse operators, respectively, of a matrix. The binary set $\{0, 1\}$ is denoted by $\mathbb{B}$. Given any positive integer n, $\mathcal{Z}_n$ stands for the set $\{1, 2, \ldots, n\}$. The n-dimensional complex, binary, and integer vector spaces are expressed as $\mathbb{C}^n, \mathbb{B}^n$, and $\mathbb{Z}^n$, respectively. The $(n \times \nu)$-dimensional complex, binary, and integer matrix spaces are expressed as $\mathbb{C}^{n \times \nu}, \mathbb{B}^{n \times \nu}$, and $\mathbb{Z}^{n \times \nu}$, respectively. The set of all $n \times n$ Hermitian matrices is denoted by $\mathbb{H}^n$ and the set of all positive semidefinite matrices is denoted by $\mathbb{H}^n_+$. The notation $\mathbf{A} \preceq \mathbf{B}$ means $\mathbf{B} - \mathbf{A} \in \mathbb{H}^n_+$. Let $\mathbf{1}_n, \mathbf{0}_n$, $\mathbf{e}^{(n)}_\nu$ be the all-one vector, all-zero vector, and ν-th standard unit vector, respectively, of dimension n. Let $\mathbf{I}_n$ be the $n \times n$ identity matrix. Operators $\| \cdot \|_p, \mathrm{tr}(\cdot), \mathrm{vec}(\cdot), \circ, \otimes$, and $\times$ denote ℓ_p-norm, trace, vectorization, Hadamard product, Kronecker product, and Cartesian, respectively. For some events A, B, the probability of A and the conditional probability of A given B are denoted by $\mathrm{Pr}\{A\}$ and $\mathrm{Pr}\{A|B\}$, respectively. For some random variables X and Y, the conditional probability density function of X given Y is denoted by $f(X|Y)$. The expectation of X is denoted by $\mathrm{E}\{X\}$. Throughout the paper, we adopt one-based

indexing. For some vector $\mathbf{x}$ and matrix $\mathbf{X}$, the i-th entry of $\mathbf{x}$ and the (i, j)-th entry of $\mathbf{X}$ are denoted by $[\mathbf{x}]_i$ and $[\mathbf{X}]_{i,j}$, respectively.

II. SYSTEM MODEL

A. Downlink SCMA System Based on OFDMA

We consider downlink SCMA transmission on top of an underlying orthogonal frequency division multiple access (OFDMA) system since SCMA multiplexed symbols need to be transmitted over orthogonal resource elements [28]. The block diagram of such a downlink SCMA system is shown in Fig. 1 where J, K, and N_B represent the number of users, the number of resources, and the FFT size in the underlying OFDMA systems, respectively. The message carrying the data bits of the b-th block transmission (i.e., the b-th OFDM symbol) for the j-th user is encoded and mapped into K-dimensional symbols $\mathbf{s}_j[b]$ by the j-th specific SCMA encoder, and then the sum of these codewords, $\mathbf{s}[b]$, is transmitted over K orthogonal resources, which are K consecutive subcarriers starting from the i_{sub}-th subcarrier of the b-th OFDM block. Other subcarriers of the OFDM block may contain data from other users. The result of the IFFT operation, $\mathbf{x}_{N_B}[b]$, is further added a cyclic prefix of length N_{CP} to obtain $\mathbf{x}[b]$ before being sent to the channel with a finite impulse response (FIR), characterized by $h[n]$, whose order is upper bounded by $L - 1$. Following the standard OFDMA receiver, at the output of the FFT operation, excluding subcarriers containing data from other users, the receiver observes the signal of the b-th block transmission as

$$\mathbf{r}[b] = \mathrm{diag}(\mathbf{h}^f_{sub})\mathbf{s}[b] + \mathbf{n}[b] \tag{1}$$

where $\mathbf{h}^f_{sub} = [[\mathbf{h}^f]_{i_{sub}}[\mathbf{h}^f]_{i_{sub}+1} \cdots [\mathbf{h}^f]_{i_{sub}+K-1}]^T \in \mathbb{C}^K$ is the subvector of the frequency-domain channel gain vector $\mathbf{h}^f \in \mathbb{C}^{N_B}$, $\mathbf{s}[b] = \sum_{j=1}^J \mathbf{s}_j[b] \in \mathbb{C}^K$ is the transmitted superimposed codeword of the b-th block transmission, and $\mathbf{n}[b] \sim \mathcal{CN}(\mathbf{0}_K, N_0\mathbf{I}_K)$ is the additive white Gaussian noise. The channel gain vector $\mathbf{h}^f$ is the discrete Fourier transform of vector $[\mathbf{h}^T \mathbf{0}^T_{N_B-L}]^T \in \mathbb{C}^{N_B}$, where $\mathbf{h} = [h[0]h[1] \cdots h[L-1]]^T \in \mathbb{C}^L$ is the channel vector which represents the channel impulse response as FIR filter of length L.

In 5G/NR applications, multipath channels are extensively considered [29], [31] and therefore the orthogonality of resources in SCMA systems needs to be ensured by an OMA technique, usually as OFDMA [28], [30]. When the popular OMA technique, OFDMA, is considered for the provision of orthogonal resources, consecutive subcarriers are usually chosen to form a resource block rather than separate subcarriers. (The interested readers can refer to Section 4.4.4 in [31].) Based on these facts, in order to fit the scenario of 5G/NR applications where SCMA is widely discussed [30], we, therefore, consider the SCMA system based on OFDMA with consecutive subcarriers chosen as orthogonal resources as what we set in the last paragraph.
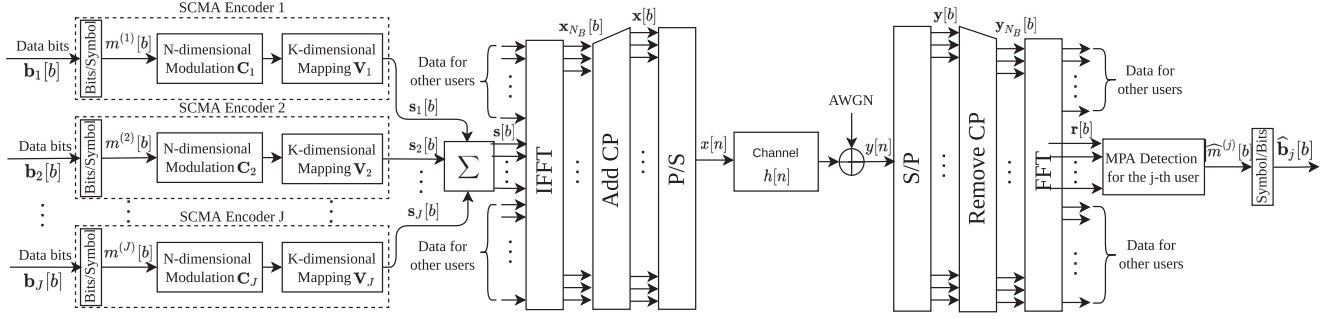
Fig. 1. Downlink SCMA system model based on OFDMA.

B. SCMA Encoder

An SCMA encoder for the j-th user can be regarded as a function defined as

$$f_j : \mathbb{B}^{\log_2 M} \to \mathcal{S}_j \tag{2}$$

where $\mathcal{S}_j \subset \mathbb{C}^K$ is the codebook of the j-th user with cardinality $|\mathcal{S}_j| = M$, i.e., $\mathcal{S}_j$ contains M codewords. We require M to be a power of two so that each codeword in $\mathcal{S}_j$ represents $\log_2 M$ bits of information. For notational convenience, we say that $m^{(j)} \in \mathcal{Z}_M$ is an *SCMA symbol* according to the vector of data bits $\mathbf{b}_j \in \mathbb{B}^{\log_2 M}$ from the j-th user: $m^{(j)} = 1 + \sum_{i=1}^{\log_2 M} 2^{i-1} [\mathbf{b}_j]_i$. Each SCMA symbol $m^{(j)}$ maps to a K-dimensional complex codeword $\mathbf{s}_j \in \mathcal{S}_j$, which is a sparse vector with N non-zero entries, and $K > N$. Following [11], [33], the j-th user's codebook is chosen as

$$\mathcal{S}_j = \{ \mathbf{V}_j \mathbf{C}_j \mathbf{e}_m^{(M)} | m \in \mathcal{Z}_M \} \tag{3}$$

where $\mathbf{C}_j \in \mathbb{C}^{N \times M}$ and $\mathbf{V}_j \in \mathbb{B}^{K \times N}$ are the constellation matrix and the mapping matrix of the j-th user, respectively. The mapping matrix $\mathbf{V}_j$ is obtained by removing $K - N$ columns from $\mathbf{I}_K$. Now, the codeword in $\mathcal{S}_j$ selected by the j-th user for the b-th block transmission can be expressed as

$$\mathbf{s}_j[b] = f_j(\mathbf{b}_j[b]) = \mathbf{V}_j \mathbf{C}_j \mathbf{e}_{m^{(j)}[b]}^{(M)} \in \mathbb{C}^K \tag{4}$$

where $\mathbf{b}_j[b] \in \mathbb{B}^{\log_2 M}$ is the vector of the given data bits and $m^{(j)}[b] \in \mathcal{Z}_M$ is the corresponding SCMA symbol. All codewords chosen by the J users will be summed up together to form a *superimposed codeword* before being assigned to orthogonal resources:

$$\mathbf{s}[b] = \sum_{j=1}^{J} \mathbf{s}_j[b].$$

A superimposed codeword is determined by all SCMA symbols from J users, collectively a vector in $\mathcal{Z}_M^J$ which we refer to as the *multiplexed symbol*. There are in total M^J multiplexed symbols, and we denote the k-th multiplexed symbol, $k \in \mathcal{Z}_{M^J}$, by $\mathbf{m}_k$ defined by

$$\mathbf{m}_k = [k_1, k_2, \ldots, k_J] \in \mathcal{Z}_M^J$$

where $k_1, \ldots, k_J \in \mathcal{Z}_M$ are the unique integers that satisfy

$$k = 1 + \sum_{j=1}^{J} (k_j - 1) M^{j-1} \in \mathcal{Z}_{M^J}. \tag{5}$$

Note that the set $\mathcal{Z}_M^J \subset \mathbb{Z}^J$ stands for the Cartesian product of J identical sets, $\mathcal{Z}_M$. The set of all SCMA superimposed codewords can be expressed as

$$\mathcal{S} = \left\{ \sum_{j=1}^{J} \mathbf{V}_j \mathbf{C}_j \mathbf{e}_{k_j}^{(M)} \middle| k \in \mathcal{Z}_{M^J} \right\}, \tag{6}$$

where $k \in \mathcal{Z}_{M^J}$ and $k_j \in \mathcal{Z}_M$ follow the relationship defined in (5).

Since each of the J users has a distinct mapping matrix $\mathbf{V}_j$, it is obvious that $J \leq \binom{K}{N}$ must hold for any given K and N. Furthermore, a loading factor is defined as $\lambda = J/K$ [13] that directly determines the spectral efficiency of SCMA. It must satisfy $\lambda > 1$ for SCMA to be more spectral efficient than conventional OMA, and it is well known that this implies $K \geq 4$, $2 \leq N \leq K - 2$, and $J > K$ [13]. In order to achieve the maximum sparsity, $N = 2$ is often chosen. In this paper, we choose to study the simplest case by taking $K = 4$, $N = 2$, and $M = 4$, which is also the most studied case in the literature. Since $J \leq \binom{K}{N}$, we take $J = 6$ and set the mapping matrices $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_J$ as follows:

$$\mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{V}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{V}_4 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{V}_5 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{V}_6 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{7}$$

Moreover, we refer to $\{\mathcal{S}_j\}_{j=1}^J$, the set of all codebooks, by the *collection of codebooks* [18], or simply the *codebook collection* throughout the paper. A codebook collection is completely determined by the J constellation matrices $\mathbf{C}_j$, $j = 1, \ldots, J$ since the mapping matrices $\mathbf{V}_j$ are fixed as in (7).

C. SCMA Decoder

1) MAP Detection: Supposing that the channel estimation is perfect and the codebook collection are available for the receiver, the detection of SCMA can be regarded as a problem of traditional multi-user detection, which can be solved by joint maximum a posteriori (MAP) [34]. Then, given some received signal $\mathbf{r}$ of some block transmission (i.e., $\mathbf{r}[b] = \mathbf{r}$ for some b), the detected multiplexed symbol, denoted by $\widehat{\mathbf{m}}$, will be

$$\widehat{\mathbf{m}} = \arg \max_{\mathbf{m} \in \mathcal{Z}_M^J} \Pr\{\mathbf{m}|\mathbf{r}\}. \tag{8}$$

The j-th user's symbol can also be detected by maximizing its marginal a posteriori probability as follows [34], [35].

$$\widehat{m}^{(j)} = \arg \max_{m \in \mathcal{Z}_M} \sum_{\mathbf{m} \in \mathcal{Z}_M^J, \ [\mathbf{m}]_j = m} \Pr\{\mathbf{m}|\mathbf{r}\} \tag{9}$$

$$= \arg \max_{m \in \mathcal{Z}_M} \sum_{\mathbf{m} \in \mathcal{Z}_M^J, \ [\mathbf{m}]_j = m} \Pr\{\mathbf{m}\} \prod_{k=1}^{K} f([\mathbf{r}]_k|\mathbf{m}). \tag{10}$$

Given the result of the detection above, we have $\widehat{m}^{(j)}[b] = \widehat{m}^{(j)}$ and $\widehat{\mathbf{b}}_j[b] = \widehat{\mathbf{b}}_j$ for all $j \in \mathcal{Z}_J$ in Fig. 1 if $\mathbf{r}[b] = \mathbf{r}$.

2) Message Passing Algorithm: Solving problems (8), (9), or (10) with brute-force has exponential complexity. Thanks to the sparsity of the codewords, the solution of this problem can be approximated by an iterative decoding algorithm, message passing algorithm (MPA), which updates the extrinsic information of function nodes (FNs) and variable nodes (VNs) along the edges in the factor graph and has moderate complexity [34], [35].

III. MAXIMIZATION OF MINIMUM EUCLIDEAN DISTANCE

A codebook collection with a large MED tends to have a smaller detection error when a joint MAP detector (8) is applied [11], [27] since the joint MAP detector will always choose the superimposed codeword closest to the received signal due to its highest posterior probability in (8). Furthermore, since the popular MPA detector usually performs very close to the joint MAP detector [6], [13], [36], [37], it makes sense to choose a codebook collection with an MED that is as large as possible.

In this section, we formulate the SCMA codebook design problem as an optimization problem that maximizes the minimum Euclidean distance (MED) under some power constraints. We follow the definitions of MED from [11], [26], that is, the MED of any two superimposed codewords of a codebook collection.

Note that in some references, MED is taken as a reasonable KPI only for the AWGN channel [17]. However, in the scenario of 5G/NR applications [31], the considered system uses consecutive OFDM subcarriers as orthogonal resources. It is believed that the adjacent OFDM tones tend to be nearly identical over fading channels [28] and thus the received signal over some fading channel will be similar to the one over AWGN channel because putting identical gain on each resource is playing the same role as amplifying the noise on each resource by the

reciprocal of the gain. Therefore, we believe that MED can be a reasonable design criterion even when frequency-selective Rayleigh fading channel is considered.

A. Problem Formulation

1) Minimum Euclidean Distance: For any given codebook collection determined by the matrices $\mathbf{V}_j, \mathbf{C}_j, j \in \mathcal{Z}_J$, we first define the square of Euclidean distance of the k-th and l-th possible superimposed codewords as

$$d_{kl} = \left\| \sum_{j=1}^{J} \mathbf{V}_j \mathbf{C}_j \left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right) \right\|_2^2 \tag{11}$$

where $k, l \in \mathcal{Z}_{M^J}$ and $k_j, l_j \in \mathcal{Z}_M, \forall j \in \mathcal{Z}_J$ are defined according to the same convention as in (5). Then, the minimum Euclidean distance (MED) $d_{\min}$ is defined as

$$d_{\min} = \min_{\substack{k,l \in \mathcal{Z}_{M^J} \\ k \neq l}} \sqrt{d_{kl}}. \tag{12}$$

Note that there are totally $\binom{M^J}{2}$ possible pairs of superimposed codewords.

2) MED Maximization Problem: We aim to maximize the MED (12) subject to the power constraint. Therefore, the problem is formulated as [11]

$$\operatorname*{maximize}_{\mathbf{C} \in \mathbb{C}^{N \times MJ}, t \in \mathbb{R}} \quad t \tag{13a}$$

$$\text{subject to} \quad d_{kl} \geq t, \ \forall k, l \in \mathcal{Z}_{M^J}, k \neq l \tag{13b}$$

$$\frac{1}{M} \operatorname{tr}(\mathbf{C}_j^H \mathbf{C}_j) = P, \forall j \in \mathcal{Z}_J \tag{13c}$$

where $\mathbf{C} \triangleq [\mathbf{C}_1 \mathbf{C}_2 \cdots \mathbf{C}_J]$ contains the constellation matrices for all users, t is an extra real-valued variable representing the square of MED, and P is the limit of each user's average power of transmitted codewords. Here we choose equality power constraints so that each user is ensured to achieve the same power limit. For convenience, and without loss of generality, we set $P = 1$ throughout the paper.

Problem (13) can be transformed into an equivalent problem in a QCQP form. Specifically, we define $\mathbf{x} = \operatorname{vec}(\mathbf{C}) \in \mathbb{C}^{n_x}$ with $n_x = NMJ$, and reformulate Problem (13) as [11]

$$\operatorname*{maximize}_{\mathbf{x} \in \mathbb{C}^{n_x}, t \in \mathbb{R}} \quad t \tag{14a}$$

$$\text{subject to} \quad \mathbf{x}^H \mathbf{A}_i \mathbf{x} \geq t, \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}} \tag{14b}$$

$$\mathbf{x}^H \mathbf{B}_j \mathbf{x} = MP, \ \forall j \in \mathcal{Z}_J \tag{14c}$$

where $\mathbf{A}_i$ and $\mathbf{B}_j$ are distance constraint and power constraint matrix, respectively. The matrices $\mathbf{A}_i$ and $\mathbf{B}_j$ have closed-form expressions as below. Note that they are real, symmetric, very sparse, and with nonzero entries limited to only values -1 and 1. The matrix $\mathbf{A}_i$ is

$$\mathbf{A}_i = \mathbf{A}_{k,l}$$
$$= \sum_{j=1}^{J} \sum_{q=1}^{J} \left((\mathbf{e}_q^{(J)})^T \otimes \mathbf{I}_{MN} \right)^T \mathbf{K}_{k,l}^{(q,j)} \left((\mathbf{e}_j^{(J)})^T \otimes \mathbf{I}_{MN} \right) \tag{15}$$

where

$$\mathbf{K}_{k,l}^{(q,j)} = \left((\mathbf{e}_{k_q}^{(M)} - \mathbf{e}_{l_q}^{(M)})(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)})^T \right) \otimes \left(\mathbf{V}_q^T \mathbf{V}_j \right),$$

and the matrix $\mathbf{B}_j$ is

$$\mathbf{B}_j = \text{diag}(\mathbf{e}_j^{(J)}) \otimes \mathbf{I}_{NM}. \tag{16}$$

The derivation of (15) is provided in Appendix A.

Problem (14) is not convex, so we apply the technique of semidefinite relaxation: let $\mathbf{X} = \mathbf{x}\mathbf{x}^H \in \mathbb{H}_+^{n_x}$ and reformulate the problem as

$$\begin{aligned} \underset{\mathbf{X} \in \mathbb{H}_+^{n_x}, t \in \mathbb{R}}{\text{maximize}} \quad & t \end{aligned} \tag{17a}$$

$$\text{subject to} \quad \text{tr}(\mathbf{A}_i \mathbf{X}) \geq t \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}} \tag{17b}$$

$$\text{tr}(\mathbf{B}_j \mathbf{X}) = MP, \ \forall j \in \mathcal{Z}_J \tag{17c}$$

$$\text{rank}(\mathbf{X}) = 1 \tag{17d}$$

Note that problems (14) and (17) are equivalent since for any $\mathbf{X} \in \mathbb{H}_+^{n_x}$ that satisfies the rank constraint (17d), there always exists some $\mathbf{x} \in \mathbb{C}^{n_x}$ (subject to a unit-norm complex ambiguity) such that $\mathbf{X} = \mathbf{x}\mathbf{x}^H$.

B. Exact Penalty Approach and Biconvex Problem Formulation

Since the rank constraint (17d) is not a convex constraint, we can not solve it directly by the tools for solving convex optimization problems. Therefore, we propose a method based on the concept of alternating maximization and exact penalty approach mentioned in [41], [42] to obtain a rank-one solution of Problem (17). We first formulate a new problem based on Problem (17) as follows

$$\begin{aligned} \underset{\substack{\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{H}_+^{n_x}, \\ t_1, t_2 \in \mathbb{R}}}{\text{maximize}} \quad & t_1 + t_2 \end{aligned} \tag{18a}$$

$$\text{subject to} \quad \text{tr}(\mathbf{A}_i \mathbf{X}_1) \geq t_1, \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}} \tag{18b}$$

$$\text{tr}(\mathbf{A}_i \mathbf{X}_2) \geq t_2, \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}} \tag{18c}$$

$$\text{tr}(\mathbf{B}_j \mathbf{X}_1) = MP, \ \forall j \in \mathcal{Z}_J \tag{18d}$$

$$\text{tr}(\mathbf{B}_j \mathbf{X}_2) = MP, \ \forall j \in \mathcal{Z}_J \tag{18e}$$

$$\text{tr}(\mathbf{X}_1 \mathbf{X}_2) = \text{tr}(\mathbf{X}_1)\text{tr}(\mathbf{X}_2). \tag{18f}$$

The following theorem shows that Problem (18) is equivalent to Problem (17).

Theorem 1: If $\{\mathbf{X}_1^\star, \mathbf{X}_2^\star, t_1^\star, t_2^\star\}$ is an optimal point for (18), then $\mathbf{X}_1^\star = \mathbf{X}_2^\star$ and $t_1^\star = t_2^\star$ and $\{\mathbf{X}, t\} = \{\mathbf{X}_1^\star, t_1^\star\}$ is an optimal point of (17). Conversely, if $\{\mathbf{X}^\star, t^\star\}$ is an optimal point of (17), then, $\{\mathbf{X}_1, \mathbf{X}_2, t_1, t_2\} = \{\mathbf{X}^\star, \mathbf{X}^\star, t^\star, t^\star\}$ is an optimal point of (18).

Proof: By Theorem 1 in [43], if there are any $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{H}_+^{n_x}$ satisfying constraint (18f), the necessary and sufficient conditions will be both of rank one and $\mathbf{X}_1 = \alpha \mathbf{X}_2$ where α is a positive scalar. Suppose $\{\mathbf{X}_1^\star, \mathbf{X}_2^\star, t_1^\star, t_2^\star\}$ is an optimal point of Problem (18). Then, (18f) implies $\mathbf{X}_1^\star = \alpha \mathbf{X}_2^\star$ for some $\alpha > 0$. Since $\mathbf{X}_1^\star$ and $\mathbf{X}_2^\star$ satisfy constraints (18d) and (18e), we have

$$MP = \text{tr}(\mathbf{B}_j \mathbf{X}_1^\star) = \text{tr}(\mathbf{B}_j \alpha \mathbf{X}_2^\star) = \alpha \cdot \text{tr}(\mathbf{B}_j \mathbf{X}_2^\star) = \alpha MP,$$

which implies $\alpha = 1$ and $\mathbf{X}_1^\star = \mathbf{X}_2^\star$. It can also be shown that $t_1^\star = t_2^\star = \min_i \text{tr}(\mathbf{A}_i \mathbf{X}_1^\star)$ using (18b), (18c), and (18a). Then, we can show that $\{\mathbf{X}_1^\star, t_1^\star\}$ is an optimal point of (17) by contradiction: if $\{\tilde{\mathbf{X}}, \tilde{t}\}$ is some optimal point for (17) with $\tilde{t} > t_1^*$, then setting $\{\mathbf{X}_1, \mathbf{X}_2, t_1, t_2\} = \{\tilde{\mathbf{X}}, \tilde{\mathbf{X}}, \tilde{t}, \tilde{t}\}$ in (18) will result in a larger value in (18a) $(t_1 + t_2 = 2\tilde{t} > 2t_1^\star = t_1^\star + t_2^\star)$.

Conversely, if $\{\mathbf{X}^\star, t^\star\}$ is an optimal point of Problem (17), then $\{\mathbf{X}_1, \mathbf{X}_2, t_1, t_2\} = \{\mathbf{X}^\star, \mathbf{X}^\star, t^\star, t^\star\}$ can be shown to be an optimal point in (18) by contradiction as follows. If there is any other feasible point of Problem (18), say, $\{\mathbf{X}_1', \mathbf{X}_2', t_1', t_2'\}$, such that $t_1' + t_2' > t_1 + t_2$, then the point $\{\mathbf{X}', t'\} = \{\mathbf{X}_1', t_1'\} = \{\mathbf{X}_2', t_2'\}$, will results in a larger value in (17a) $(t' = (t_1' + t_2')/2 > (t_1 + t_2)/2 = t^\star)$. ∎

To deal with constraint (18f), we apply the exact penalty approach and introduce the non-positive penalty function

$$\text{tr}(\mathbf{X}_1 \mathbf{X}_2) - \text{tr}(\mathbf{X}_1)\text{tr}(\mathbf{X}_2)$$

whose value is zero if and only if (18f) holds [43], [46]. By the equality constraints (18d), (18e) and using (16), we have $\text{tr}(\mathbf{X}_1) = \text{tr}(\mathbf{X}_2) = JMP$, implying that $\text{tr}(\mathbf{X}_1)\text{tr}(\mathbf{X}_2)$ is a constant. Therefore, the penalty function can be further simplified as $\text{tr}(\mathbf{X}_1 \mathbf{X}_2)$ and then we formulate another problem as

$$\begin{aligned} \underset{\substack{\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{H}_+^{n_x}, \\ t_1, t_2 \in \mathbb{R}}}{\text{maximize}} \quad & t_1 + t_2 + w \cdot \text{tr}(\mathbf{X}_1 \mathbf{X}_2) \end{aligned} \tag{19a}$$

$$\text{subject to} \quad \text{tr}(\mathbf{A}_i \mathbf{X}_1) \geq t_1, \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}} \tag{19b}$$

$$\text{tr}(\mathbf{A}_i \mathbf{X}_2) \geq t_2, \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}} \tag{19c}$$

$$\text{tr}(\mathbf{B}_j \mathbf{X}_1) = MP, \ \forall j \in \mathcal{Z}_J \tag{19d}$$

$$\text{tr}(\mathbf{B}_j \mathbf{X}_2) = MP, \ \forall j \in \mathcal{Z}_J. \tag{19e}$$

where $w > 0$ is a positive weight of the simplified penalty function $\text{tr}(\mathbf{X}_1 \mathbf{X}_2)$, which makes the optimization problem tend to meet constraint (18f).

Problem (19) is a biconvex optimization problem [42], meaning that it is a convex problem in $\mathbf{X}_1$ and t_1 when $\mathbf{X}_2$ and t_2 are given constants, and vice versa. An iterative algorithm exploiting alternative maximization is presented in Algorithm 1, where $\varphi_{\max}$ is the maximum allowable number of iterations, and, noting that the weighting w can be chosen

---

Algorithm 1: Proposed Algorithm Based on Alternating Maximization and Exact Penalty Approach.

Input: $J, \mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_J, \mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_J, \varphi_{\max}$, $\{w_{1,\varphi}\}_{\varphi=1}^{\varphi_{\max}}, \{w_{2,\varphi}\}_{\varphi=1}^{\varphi_{\max}}$.

Output: $\tilde{\mathbf{x}}$

1: **Initialization**: Create $\mathbf{A}_i, \forall i \in \mathcal{Z}_{\binom{M^J}{2}}$ and $\mathbf{B}_j, \forall j \in \mathcal{Z}_J$ according to (15) and (16).

2: Initialize $\mathbf{X}_2^{(0)} = \mathbf{x}\mathbf{x}^H$, where $\mathbf{x} = \text{vec}([\mathbf{C}_1 \mathbf{C}_2 \cdots \mathbf{C}_J])$.

3: Set $t_2^{(0)} = \min_i \text{tr}(\mathbf{A}_i \mathbf{X}_2)$ and $\varphi = 0$.

4: **repeat**

5: Solve Problem (19) for $\{\mathbf{X}_1^{(\varphi+1)}, t_1^{(\varphi+1)}\}$ while fixing $\{\mathbf{X}_2, t_2\}$ as $\{\mathbf{X}_2^{(\varphi)}, t_2^{(\varphi)}\}$ with the weight being chosen as $w = w_{1,\varphi}$.

6: Solve Problem (19) for $\{\mathbf{X}_2^{(\varphi+1)}, t_2^{(\varphi+1)}\}$ while fixing $\{\mathbf{X}_1, t_1\}$ as $\{\mathbf{X}_1^{(\varphi+1)}, t_1^{(\varphi+1)}\}$ with the weight being chosen as $w = w_{2,\varphi}$.

7: $\varphi \leftarrow \varphi + 1$

8: **until** $|\text{tr}(\mathbf{X}_1^{(\varphi)})\text{tr}(\mathbf{X}_2^{(\varphi)}) - \text{tr}(\mathbf{X}_1^{(\varphi)}\mathbf{X}_2^{(\varphi)})| < 10^{-3}$ or $\varphi > \varphi_{\max}$

9: Perform singular value decomposition (SVD) on $\mathbf{X}_2^{(\varphi)}$: $\mathbf{X}_2^{(\varphi)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, with the singular values along the main diagonal of $\mathbf{\Sigma}$ in non-ascending order, i.e., $[\mathbf{\Sigma}]_{1,1} \geq [\mathbf{\Sigma}]_{2,2} \geq \cdots \geq [\mathbf{\Sigma}]_{n_x, n_x} \geq 0$.

10: Set $\sigma_1 = \left[\mathbf{\Sigma}\right]_{1,1}$ and $\sigma_2 = \left[\mathbf{\Sigma}\right]_{2,2}$.

11: **if** $\sigma_2/\sigma_1 \leq 10^{-4}$ **then**

12: Obtain $\tilde{\mathbf{x}} = \sqrt{\sigma_1} \cdot \mathbf{U}\mathbf{e}_1^{(n_x)}$.

13: **else**

14: Declare failure of convergence.

15: **end if**

---

different among various iterations, $w_{1,\varphi}$ and $w_{2,\varphi}$ represent the positive weights for the φ-th iteration.

Unfortunately, Algorithm 1 is not guaranteed to always converge within a satisfactory number of iterations. As will be elaborated in Section V, we observe that if the weight is kept constant during all iterations, the algorithm is more prone not to converge. A possible solution to this problem, accordingly to our empirical experiments, is to manually change the weights $\{w_{1,\varphi}\}, \{w_{2,\varphi}\}$. We found that if the weights are set as a sequence that gradually increases, the algorithm tends to converge to a solution with a large MED.

IV. DUAL PROBLEM OF MED MAXIMIZATION PROBLEM

In this section, we derive the Lagrange dual problem associated with the primal problem (14). Unlike the primal problem, which is non-convex, the dual problem is always a convex problem that is easier to solve. And the optimal value of a dual problem serves as an upper bound of the optimal value of the primal problem. Noting that Problems (13) and (14) are the two equivalent forms of the primal problem, one can derive the dual problem for each of the forms. Here, we choose to derive the

Lagrange dual of Problem (14) since its QCQP form makes the dual problem derivation much easier than Problem (13). First of all, the Lagrangian of Problem (14) is

$$
\mathcal{L}(\{\lambda_i\}, \{\mu_j\}, \mathbf{x}, t)
$$

$$
= t + \sum_{i=1}^{\binom{M^J}{2}} \lambda_i(\mathbf{x}^H \mathbf{A}_i \mathbf{x} - t) + \sum_{j=1}^{J} \mu_j(MP - \mathbf{x}^H \mathbf{B}_j \mathbf{x})
$$

$$
= \mathbf{x}^H \left(\sum_{i=1}^{\binom{M^J}{2}} \lambda_i \mathbf{A}_i - \sum_{j=1}^{J} \mu_j \mathbf{B}_j \right) \mathbf{x} + \left(1 - \sum_{i=1}^{\binom{M^J}{2}} \lambda_i \right) t
$$

$$
+ \sum_{j=1}^{J} \mu_j MP, \tag{20}
$$

where we introduced Lagrange dual variables λ_i and μ_j associated with constraints (14b) and (14c), respectively. Then, the Lagrange dual function $g(\{\lambda_i\}, \{\mu_j\})$, defined as

$$
g(\{\lambda_i\}, \{\mu_j\}) = \sup_{\mathbf{x}, t} \mathcal{L}(\{\lambda_i\}, \{\mu_j\}, \mathbf{x}, t), \tag{21}
$$

is unbounded above if any eigenvalue of $\sum_{i=1}^{\binom{M^J}{2}} \lambda_i \mathbf{A}_i - \sum_{j=1}^{J} \mu_j \mathbf{B}_j$ is greater than zero or $\sum_{i=1}^{\binom{M^J}{2}} \lambda_i \neq 1$. Otherwise, the dual function is

$$
g(\{\lambda_i\}, \{\mu_j\}) = \sum_{j=1}^{J} \mu_j MP. \tag{22}
$$

Therefore, the Lagrange dual problem is found to be

$$
\underset{\{\lambda_i\}, \{\mu_j\}}{\text{minimize}} \qquad \sum_{j=1}^{J} \mu_j MP \tag{23a}
$$

$$
\text{subject to} \qquad \sum_{i=1}^{\binom{M^J}{2}} \lambda_i \mathbf{A}_i \preceq \sum_{j=1}^{J} \mu_j \mathbf{B}_j \tag{23b}
$$

$$
\sum_{i=1}^{\binom{M^J}{2}} \lambda_i = 1 \tag{23c}
$$

$$
\lambda_i \geq 0, \ \forall i \in \mathcal{Z}_{\binom{M^J}{2}}. \tag{23d}
$$

It is well known that weak duality [38] dictates that the optimal value of the dual problem (23) is an upper bound of the optimal value of the primal problem (14), as also that of (13) since problems (13) and (14) are equivalent. In fact, as we will find later in Section V, the optimal values of primal problem (14) and dual problem (23) will coincide, at least for the case of $J = 3$, suggesting that strong duality holds for this case.

V. SIMULATION RESULTS

In this section, we conduct numerical simulations to verify the proposed methods presented in Section III and compare their performances with existing methods [9], [10], [16]–[18], and also the MED upper bound derived in Section IV. Throughout

all simulations, we set the number of resources as $K = 4$, the cardinality of codebooks as $M = 4$, the constellation sizes as $N = 2$, and 15 times message passing iterations. The definitions of SER and BER are shown as follows.

Definition The symbol error rate (SER) is defined as

$$P_{e,s} = \frac{1}{J} \sum_{j=1}^{J} \sum_{m=1}^{M} \Pr\{m^{(j)} = m\} \Pr\{\widehat{m}^{(j)} \neq m \mid m^{(j)} = m\} \tag{24}$$

where, for all $j \in \mathcal{Z}_J$, $m^{(j)}$ and $\widehat{m}^{(j)}$ follow the definitions in Section II, and $\Pr\{m^{(j)} = m\}$ is assumed to be $1/M$ for all $m \in \mathcal{Z}_M$.

Definition: The bit error rate (BER) is defined as

$$P_{e,b}$$

$$= \frac{1}{J \log_2 M} \sum_{j=1}^{J} \sum_{i=1}^{\log_2 M} \left(\Pr\{[\mathbf{b}_j]_i = 0\} \Pr\{[\widehat{\mathbf{b}}_j]_i = 1 | [\mathbf{b}_j]_i = 0\} \right.$$

$$\left. + \Pr\{[\mathbf{b}_j]_i = 1\} \Pr\{[\widehat{\mathbf{b}}_j]_i = 0 | [\mathbf{b}_j]_i = 1\} \right) \tag{25}$$

where, for all $j \in \mathcal{Z}_J$, $\mathbf{b}_j, \widehat{\mathbf{b}}_j$ follow the definitions in Section II, and $\Pr\{[\mathbf{b}_j]_i = 0\}$ and $\Pr\{[\mathbf{b}_j]_i = 1\}$ are assumed to be $1/2$ for all $i \in \mathcal{Z}_{\log_2 M}$.

Notations E_s and E_b, representing the average energy of all users' symbols and data bits, respectively, are defined as

$$E_s = \frac{1}{JM} \sum_{j=1}^{J} \sum_{m=1}^{M} \left\| \mathbf{V}_j \mathbf{C}_j \mathbf{e}_m^{(M)} \right\|_2^2 = \frac{1}{JM} \sum_{j=1}^{J} \mathrm{tr}(\mathbf{C}_j^H \mathbf{C}_j) \tag{26}$$

and

$$E_b = \frac{1}{\log_2 M} E_s. \tag{27}$$

The normalized minimum Euclidean distance is defined as

$$\hat{d}_{\min} = \frac{d_{\min}}{\sqrt{E_s}} \tag{28}$$

where E_s is also the average power of each codeword of each user by the definition in (26).

A. The Case With Three Users ($J = 3$)

We first consider the simple case where only three users are allowed to share the $K = 4$ orthogonal resources, i.e., $J = 3$. It will be demonstrated that, in this simple case, it is possible to use Algorithm 1 to obtain a codebook collection whose MED reaches the theoretical upper bound determined by (23). Mapping matrices $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ in (7) are chosen and this choice makes sure the number of collisions is as small as 2 [11]. Algorithm 1 is applied to solve Problem (19) with the initial point $\mathbf{X}_2 = \mathbf{x}\mathbf{x}^H$ where all elements of $\mathbf{x}$ (i.e., all entries of $\{\mathbf{C}_j\}_{j=1}^3$) are independently generated by a random variable uniformly distributed between 0 and 1. In this relatively simple case, Algorithm 1 is found to be converging and leading to an optimal codebook collection whose optimal value coincides with that of the dual problem (23). We used CVX, a MATLAB-based
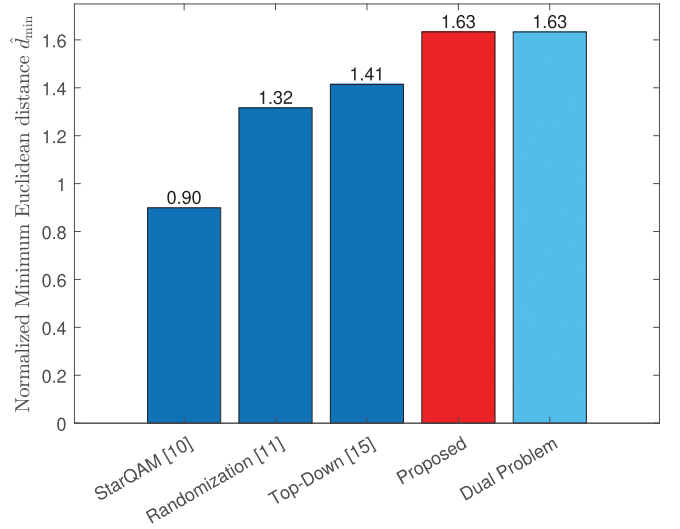

Fig. 2. Minimum Euclidean distance comparison for $J = 3$.

modeling system for convex optimization [39], for Steps 5 and 6 of Algorithm 1 in each iteration. As a result, we observed that, in about 10 iterations, the algorithm converged. We also used CVX to solve the dual problem (23) to get the upper bound of optimal value. Note that the dual problem (23) is always convex, so it can be solved with just a single CVX instance.

Fig. 2 shows the MED comparison of the proposed algorithms and various previously reported methods in terms of MED, along with the bound given by the dual problem (23). As indicated in the figure, the proposed method achieves the largest value among all methods, including the randomization method [11] with $L_{rand} = 10^6$ [1], the starQAM codebook collection [10], and the Top-Down codebook collection [15]. We notice that the MED of the proposed method is the same as the optimal value of the dual problem (23). As stated in Section IV, it is sufficient to show that the codebook collection we proposed is a set of optimal codebooks that achieve the maximum MED.

B. The Case With Six Users ($J = 6$)

We now consider the case when the number of users is $J = 6$. As mentioned in Section III, a large number of distance constraints greatly increase the consumption of memory and computational complexity. It is extremely inefficient for the CVX tool to handle a large number of more than 8 million constraints in constructing the problem settings alone, not to mention solving it. As an alternative approach, we resort to directly using SDPT3 [40], the default solver of the CVX tool. We observed that the SDPT3 solver, without the extra burden caused by the problem-constructing routines of CVX, is capable of returning correct results within an acceptable time duration. Moreover, it is worthy to note that the sparse properties of the matrices $\mathbf{A}_i, \mathbf{B}_j$ with nonzero entries being ± 1, may have also expedited the computation of the SDPT3 solver [40]. We used

---

[1] The codebook collection obtained here do not necessarily coincide with the one reported in [11], due to the random nature of Gaussian randomization algorithm.

this new approach to execute Algorithm 1 and tried to solve Problem (14). Although it may not always converge within $\varphi_{\max}$ iterations, we usually can obtain an $\mathbf{X}$ with the largest eigenvalue dominating all the other eigenvalues (i.e., the ratio of the second largest eigenvalue to the largest one $\sigma_2/\sigma_1 \leq 10^{-4}$) and thus it is fair enough to consider $\mathbf{X}$ as $\mathbf{x}\mathbf{x}^H$ for retrieving $\mathbf{x}$ by Step 12 in Algorithm 1.

Since Algorithm 1 needs the constellation matrices $\{\mathbf{C}_j\}_{j=1}^6$ of some codebook collection for initialization, we chose two codebook collections for AWGN channel proposed in [17] and [18] since they have relatively good error rate performances comparing to other codebook collections over AWGN channels. We first test Algorithm 1 using the codebook collection for AWGN channel proposed in [17] (referred to as "Chen's AWGN codebook collection") for initialization, and we chose weight $w = 0.1$. The algorithm converged in 18 iterations with a total computation time of 26 hours. The MED of the resultant codebook collection is 1.17, which is already greater than the MEDs of all previously reported codebook collections (The detailed codebook collection is shown in Appendix B). Then, we tried Algorithm 1 using the codebook collection for AWGN channel proposed in [18] (referred to as "Deka's AWGN codebook collection")[2]. We at first tried the fixed weight $w = 0.3$ but it ran for over 80 iterations, which takes almost a whole week of computing with the CPU being AMD Ryzen 9 3900X, and did not converge. Hoping for the convergence of Algorithm 1, we manually chose weight w between 0.1 and 0.35 in the process of the iterations in Algorithm 1 and it converged with MED being 1.30 (The detailed codebook collection is shown in Appendix C). More specifically, we chose:

$$
w_{1,\varphi} = \begin{cases} 0.1 & 1 \leq \varphi \leq 6 \\ 0.15 & 7 \leq \varphi \leq 27 \\ 0.2 & 28 \leq \varphi \leq 35 \\ 0.25 & 36 \leq \varphi \leq 59 \\ 0.3 & 60 \leq \varphi \leq 71 \\ 0.35 & 72 \end{cases}, w_{2,\varphi} = \begin{cases} 0.1 & 1 \leq \varphi \leq 6 \\ 0.15 & 7 \leq \varphi \leq 26 \\ 0.2 & 27 \leq \varphi \leq 34 \\ 0.25 & 35 \leq \varphi \leq 58 \\ 0.3 & 59 \leq \varphi \leq 71 \\ 0.35 & 72 \end{cases}.
$$

It took 72 iterations and 114 hours to obtain this codebook collection.

The dual problem (23) is also solved via SDPT3 with $J = 6$ and the resulting upper bound is 1.63. The comparison of the normalized MEDs of different codebook collections is shown in Fig. 3 [3]. Although Deka's AWGN codebook collection (labeled as "Deka 2020 (AWGN) [18]") have a smaller MED than Chen's AWGN codebook collection (labeled as "Chen 2020 (AWGN) [17]"), the former results in a codebook collection with an even larger MED. We found that this may have been due to that only a very small number of pairs of superimposed codewords achieve the MED for Deka's AWGN codebook collection.

As the codebook collection obtained by Algorithm 1 with Deka's AWGN codebook collection [18] for initialization has

[2]Note that the "Chen's AWGN codebook collection" [17] and "Deka's AWGN codebook collection" [18] for initialization we used here are scaled to meet power constraint (13c).

[3]Note that all codebook collections we used for comparison hereafter are scaled such that the power of the user with maximum average power meet power constraint (13c) (i.e., $\max_{j \in \mathcal{Z}_J} \frac{1}{M}\mathrm{tr}(\mathbf{C}_j^H\mathbf{C}_j) = P$ for all codebook collections).
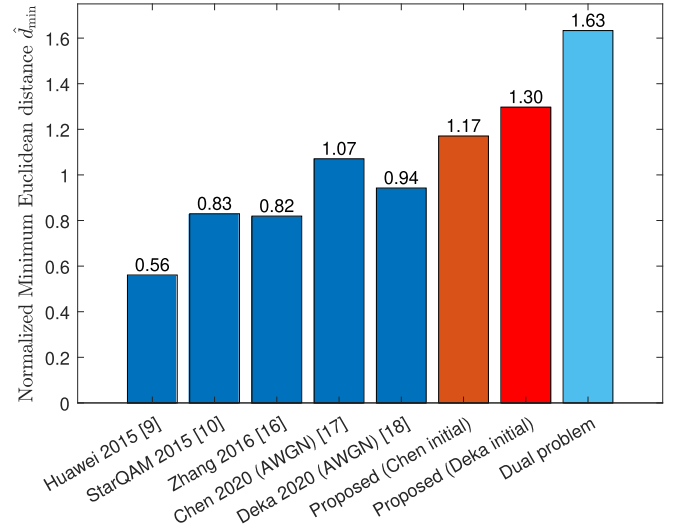

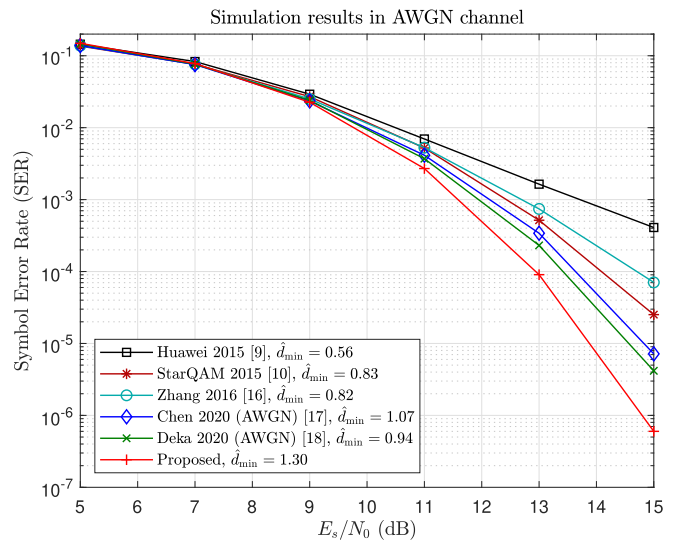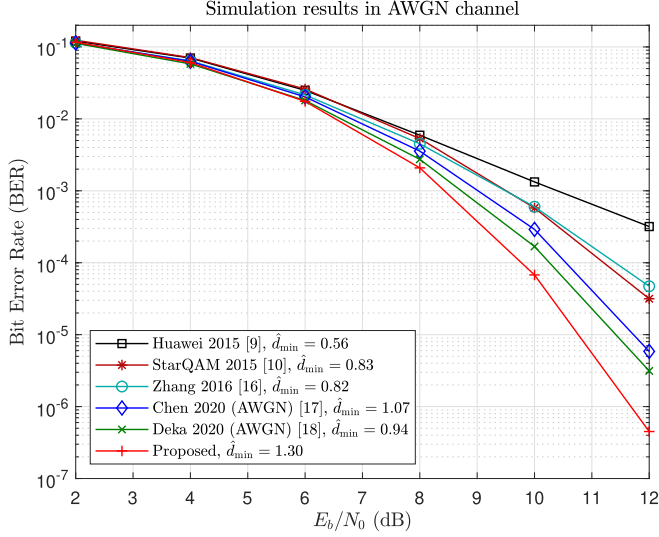Fig. 3.   Minimum Euclidean distance comparison for $J = 6$.

Fig. 4.   SER performance comparison for $J = 6$ over AWGN channel.

the largest MED, 1.30, it is expected that it will attain relatively better error rate performances than other codebook collections and therefore we use it as the proposed codebook collection in the following numerical results.

The parameter setting is shown as follows. The FFT size N_B is 256. The channel length L is 18 and cyclic-prefix length N_{CP} is 17. The subcarrier indices of the K-dimensional SCMA signal is set to 127, 128, 129, 130 within the 256 subcarriers of an OFDM symbol (i.e. $i_{sub} = 127$ in (1)). The other signals loaded on the remaining subcarriers are set as independent and identically distributed (i.i.d.) signals with distribution being $\mathcal{CN}(0, JE_s/K)$. We adopted the MPA algorithm described in Section III-B in reference [34] for the detection. The results, corresponding to the AWGN channel and Rayleigh fading channel, are discussed in the following parts respectively.

1) AWGN Channel: For the simulation of the downlink SCMA system based on OFDMA over AWGN channel, the results, as shown in Fig. 4, and 5, demonstrate that the

Fig. 5. BER performance comparison for $J = 6$ over AWGN channel.

Fig. 6. BER performance comparison for $J = 6$ over Rayleigh fading channel (including cases with poor channel).

Fig. 7. BER performance comparison for $J = 6$ over Rayleigh fading channel (excluding cases with poor channel).

proposed codebook collection obtained by Algorithm 1 with Deka's AWGN codebook collection [18] for initialization indeed achieves the best SER and BER performances since it has the largest MED. Specifically, there are gains of both about 0.7 dB at SER $= 10^{-5}$ and BER $= 10^{-5}$ over the best existing codebook collection [18].

Moreover, since the downlink SCMA system based on OFDMA over AWGN channel has an all-one channel gain for all subchannels, and the additive noises after the FFT operation are still uncorrelated Gaussian noises with zero means and N_0 variances, this case is actually the same case with AWGN channel considered in other works [10]–[18].

2) Rayleigh Fading Channel: The Rayleigh fading channel here is set to be a frequency-selective channel, which fits in the scenario with multipath channel in 5G/NR applications (e.g. TDL-A/C, EPA) [29] and is, for convenience, specifically set as

$$\mathbf{h} \sim \mathcal{CN}(0, \text{diag}(\sigma_{h[0]}^2, \sigma_{h[1]}^2, \dots, \sigma_{h[L-1]}^2)) \quad (29)$$

where $[\sigma_{h[0]} \sigma_{h[1]} \cdots \sigma_{h[L-1]}]^T = \mathbf{h}_\sigma / \|\mathbf{h}_\sigma\|_2$, and $\mathbf{h}_\sigma \in \mathbb{R}^L$ is the vector whose elements are linearly spaced between 0 dB and -48 dB. With the setting above, the simulation results are shown in Fig. 6. We notice that most of the codebook collections have similar performances on bit error rate, but the codebook collection proposed by Chen *et al.* [17] for downlink Rayleigh fading channel (labeled as "Chen 2020 (downlink) [17]") has worse performances.

Although the proposed codebook collection does not outperform other ones in BER for the case mentioned above, we notice that E_b/N_0 needs to be about 26 dB for BER $= 10^{-3}$, and thus E_b/N_0 is expected to be unrealistically high for a better BER, such as 10^{-5} or 10^{-6}. The bad performance curves for all methods may have been mostly contributed by deep fade channels in some Monte Carlo trials. In the 5G/NR applications, however, we believe it is reasonable to assume all users that share the SCMA resources would possess sufficiently good channel quality on these subcarriers since the base station, with the knowledge of channel state information reported from a user
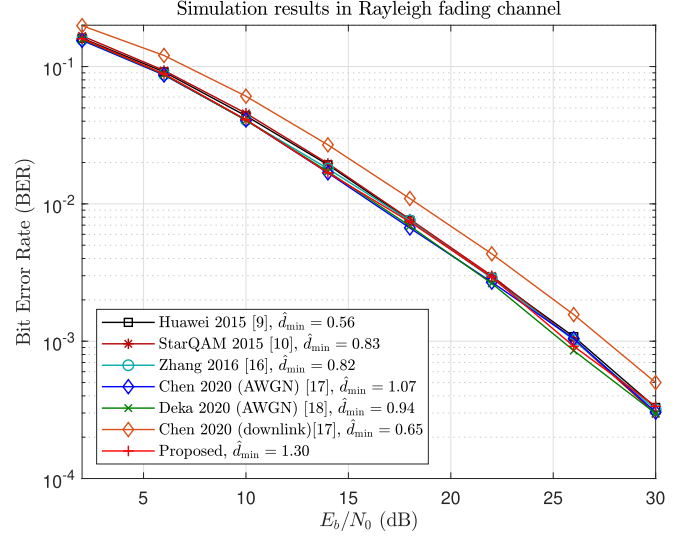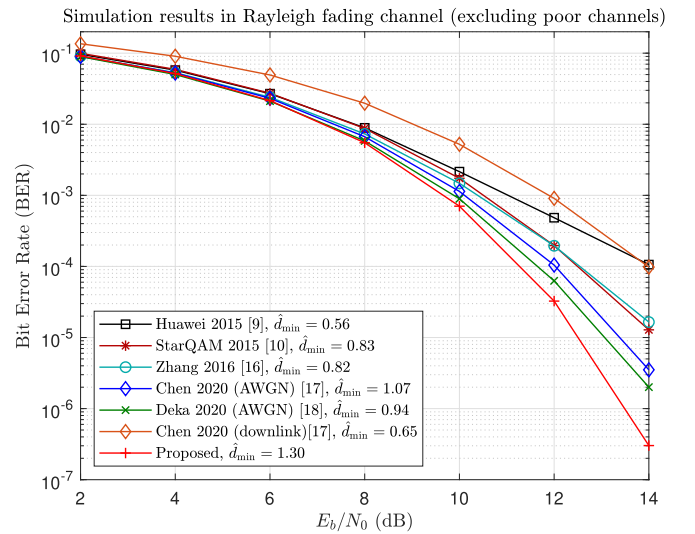
(e.g., see Section V of [32]), is likely to assign users to the resources with good channel quality. With this assumption in mind, we conduct the performance analysis again by excluding the 40 percent poorest channels $\mathbf{h}_{sub}^f$. The results, as shown in Fig. 7, demonstrate that the proposed codebook collection outperforms all the other codebook collections on bit error rate. Specifically, there is a gain of about 0.6 dB at BER $= 10^{-5}$ over the best existing codebook collection [18].

C. Simulation Results Under Other Channel Models

In this subsection, we conduct simulations under some channel models that have been considered in [17]–[19], whose system models are slightly different from the one we described in Section II-A. Specifically, they consider the downlink and uplink
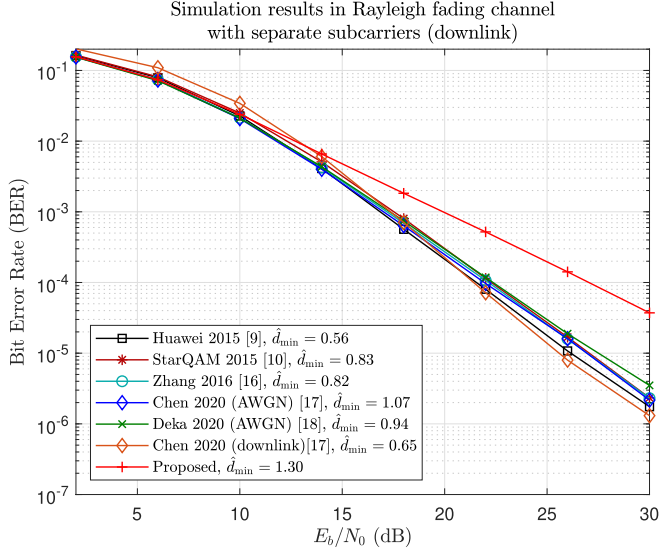
Fig. 8. BER performance comparison for $J = 6$ over Rayleigh fading channel with separate subcarriers (downlink).

Fig. 9. BER performance comparison for $J = 6$ over Rayleigh fading channel with consecutive subcarriers (uplink).

SCMA systems over OFDMA with separate subcarriers. Even though the 5G/NR standards have adopted resource blocks that are composed of consecutive subcarriers rather than separate ones [31], we investigate these two cases for a more comprehensive comparison of the proposed codebook collection obtained in the previous subsection (as shown in Appendix C) with existing ones.

1) Downlink Rayleigh Fading Channel With Non-Consecutive Subcarriers: In [17], the authors considered this type of channel as an equivalent SISO fast Rayleigh channel since the statistics of channels of non-adjacent subcarriers separated by a frequency gap greater than the coherence bandwidth could be considered to be uncorrelated. To simulate this scenario, we slightly modified the system model depicted in Fig. 1 and make $\mathbf{s}_1[b]$ to $\mathbf{s}_J[b]$ occupy non-consecutive subcarriers, so the received signals at the corresponding subcarriers can be expressed as

$$\mathbf{r}[b] = \text{diag}(\mathbf{h}^f_{\text{sub,sep}})\mathbf{s}[b] + \mathbf{n}[b] \qquad (30)$$

where $\mathbf{h}^f_{\text{sub,sep}} = [[\mathbf{h}^f]_{i_1}[\mathbf{h}^f]_{i_2}\cdots[\mathbf{h}^f]_{i_K}]^T \in \mathbb{C}^K$ is the sub-vector of the frequency-domain channel gain vector $\mathbf{h}^f \in \mathbb{C}^{N_B}$ at subcarriers indexes i_1 through i_K that are non-consecutive or even separated apart. We choose $(i_1, i_2, i_3, i_4) = (32, 96, 160, 224)$ in our simulation and the results are shown in Fig. 8. The results match the corresponding results shown in [17][4], reconfirming the equivalence of the SISO fast Rayleigh channel considered therein and the Rayleigh fading channel using widely separated subcarriers under OFDMA. We observe that the proposed codebook collection, compared to others, does not work well perhaps because the product distance [17] of the codebook collection is not optimized.

2) Uplink Rayleigh Fading Channel: For the uplink SCMA system based on OFDMA, it can be regarded as that the transmitter of each user is transmitting the OFDM signals with its SCMA signals, say, $\mathbf{s}_j[b]$, loaded on the chosen subcarriers, and the receiver receives the J synchronized signals from the J transmitters and process the received signal in the same way as in the downlink SCMA system based on OFDMA. We first consider the case with consecutive subcarriers and the corresponding received signal can be expressed as

$$\mathbf{r}[b] = \sum_{j=1}^{J} \text{diag}(\mathbf{h}^f_{j,sub})\mathbf{s}_j[b] + \mathbf{n}[b], \qquad (31)$$

where $\mathbf{h}^f_{j,sub} = [[\mathbf{h}^f_j]_{i_{sub}}[\mathbf{h}^f_j]_{i_{sub}+1}\cdots[\mathbf{h}^f_j]_{i_{sub}+K-1}]^T \in \mathbb{C}^K$ is the subvector of $\mathbf{h}^f_j$, the frequency-domain channel gain vector for the j-th user. Vector $\mathbf{h}^f_j$ is the discrete Fourier transform of vector $[\mathbf{h}^T_j \ \mathbf{0}^T_{N_B-L}]^T \in \mathbb{C}^{N_B}$, where $\mathbf{h}_j$ shares the same distribution as the $\mathbf{h}$ in (29). We assume that $\mathbf{h}^f_{j_1}$ and $\mathbf{h}^f_{j_2}$ are statistically uncorrelated whenever $j_1 \neq j_2$. With the setting above, the simulation results are shown in Fig. 9[5]. The proposed codebook collection does not outperform other ones. For the uplink case with non-consecutive subcarriers, as the readers can imagine, the proposed codebook collection will have an even worse performance. The relatively bad BER performances of the proposed codebook collection indicate the fact that MED may not be a sufficient KPI for the uplink case, which reconfirms the claim of weak correlation of MED and BER performances in the uplink case in [17].

D. Comparison of Theoretical Results and Simulation Results

To certify the error rate performances reported in the previous simulation plots, we compare all error rate curves of the cases in

[4]Note that the shift in E_b/N_0 between our results and the corresponding ones in [17] is contributed by the different setting of E_s.
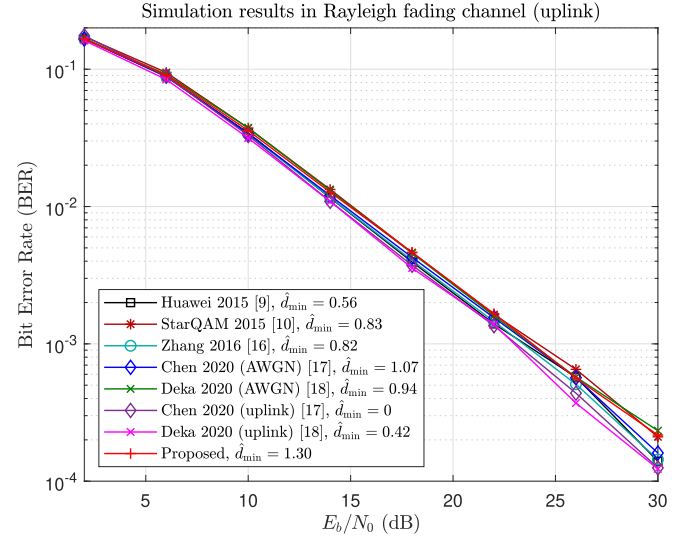
[5]The codebook collections for uplink systems proposed in [17], [18] are labeled as "Chen 2020 (uplink) [17]," "Deka 2020 (uplink) [18]," respectively.

AWGN channel with theoretical bounds in this subsection. The theoretical bounds are derived based on the maximum likelihood (ML) multi-user detection [36], [37], which is equivalent to joint MAP detection due to the equally likely input symbols. Moreover, because of the similar error rate performance in MPA detection and joint MAP detection [6], [13], [36], [37], it is reasonable to compare the MPA-based simulation results and the bounds of the error rates of the MAP detector. Hence, we present these bounds in the following paragraph for the comparison with the simulation results.

For the SER of the MAP detector, denoted by $P_{e,s}^{\mathrm{MAP}}$, an upper bound can be found by simply taking the average over J users of the upper bound of the SER of each single user derived by Bao *et al.* (Eq. (38) in [36]).

$$P_{e,s} \approx P_{e,s}^{\mathrm{MAP}} \leq \frac{1}{M^J \cdot J} \mathbf{1}_{M^J}^T (\mathbf{Q} \circ \mathbf{D}_s) \mathbf{1}_{M^J} \qquad (32)$$

where the matrix $\mathbf{Q} \in \mathbb{R}^{M^J \times M^J}$ is defined with $[\mathbf{Q}]_{k,l} = Q(\sqrt{\frac{d_{kl}}{2N_0}})$, d_{kl} was defined in (11), the matrix $\mathbf{D}_s \in \mathbb{Z}^{M^J \times M^J}$ is defined with $[\mathbf{D}_s]_{k,l} = d_{s,kl}$, and $d_{s,kl}$ is the Hamming distance between the multiplexed symbols $\mathbf{m}_k, \mathbf{m}_l$, i.e., $d_{s,kl} = \sum_{j=1}^{J} |k_j - l_j|_0$ with $|\cdot|_0$ being the indicator function of nonzero values. For the BER of the MAP detector, denoted by $P_{e,b}^{\mathrm{MAP}}$, an upper bound is found to be

$$P_{e,b} \approx P_{e,b}^{\mathrm{MAP}} \leq \frac{1}{M^J \cdot J\log_2 M} \mathbf{1}_{M^J}^T (\mathbf{Q} \circ \mathbf{D}_b) \mathbf{1}_{M^J} \qquad (33)$$

where the matrix $\mathbf{D}_b \in \mathbb{Z}^{M^J \times M^J}$ is defined with $[\mathbf{D}_b]_{k,l} = d_{b,kl}$, and $d_{b,kl}$ is the Hamming distance between the bit patterns loaded on the multiplexed symbols $\mathbf{m}_k, \mathbf{m}_l$, i.e., $d_{b,kl} = \sum_{j=1}^{J} \|\mathbf{b}^{[k_j]} - \mathbf{b}^{[l_j]}\|_0$ with $\|\cdot\|_0$ being the ℓ^0 norm in [45], and $\mathbf{b}^{[k_j]}, \mathbf{b}^{[l_j]} \in \mathbb{B}^{\log_2 M}$ denote the bits corresponding to symbol k_j, l_j, respectively, according to the convention in Section II-B. The derivation of (33) can be done by the definitions of SER, BER, and (32).

In the following comparisons shown in Fig. 10, and 11, we will find that all simulation curves are matching the theoretical upper bounds (32), (33) in high SNR regions within acceptable margins. These results not only double-checked the correctness of the bounds, but also secured all performance advantages of the proposed codebook collection that we have seen in Fig. 4, 5, and 7. The bounds are rather loose in low SNR regions, but they go tighter as SNR goes higher. Sometimes it is observed that simulation results even have slightly larger error rates than the bounds, and we believe this is because the MPA is still worse than MAP. Specifically, we notice MPA detection may not always take the closest superimposed codeword to the received signal $\mathbf{r}[b]$ as the detected one as MAP does, and this may have resulted in the slightly larger error rates than the theoretical bounds.

Based on the tightness of the upper bounds and the simulation results discussed above, we can further predict the SER, and BER of some codebook collections in the high SNR region as shown in Fig. 12. It can be observed that the proposed codebook collection still has the best error rate performances.

Moreover, some remarks about the relation of Euclidean distances and error rates are made. It is observed that the upper
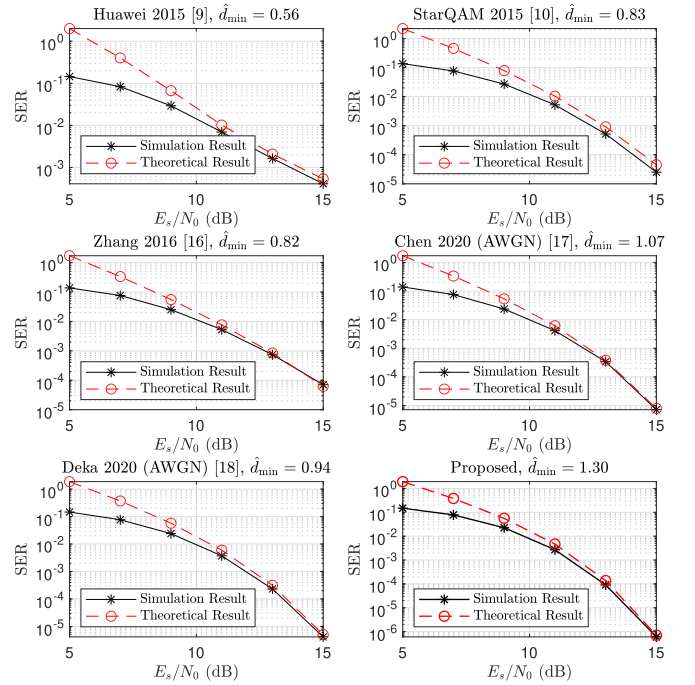

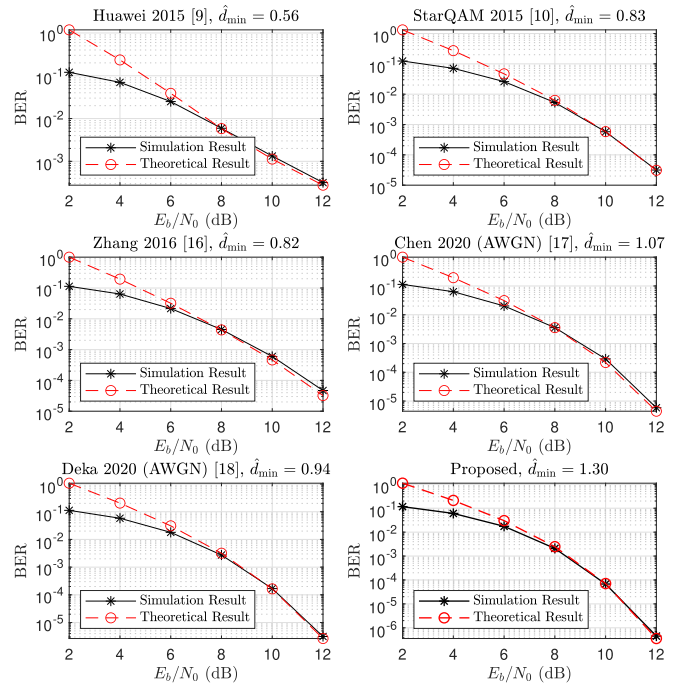Fig. 10. Comparison of theoretical and simulation results of SER over AWGN channel.

Fig. 11. Comparison of theoretical and simulation results of BER over AWGN channel.

bounds, (32), (33), are both proportional to some weighted sums of all elements in $\mathbf{Q}$, so it is desirable to minimize the entries of $\mathbf{Q}$, which depend on the Euclidean distances, for a low error rate. Since the largest entry in $\mathbf{Q}$ is $Q(\frac{d_{\min}}{\sqrt{2N_0}})$, which will dominate the contributions to error rate upper bound formulas (i.e.,
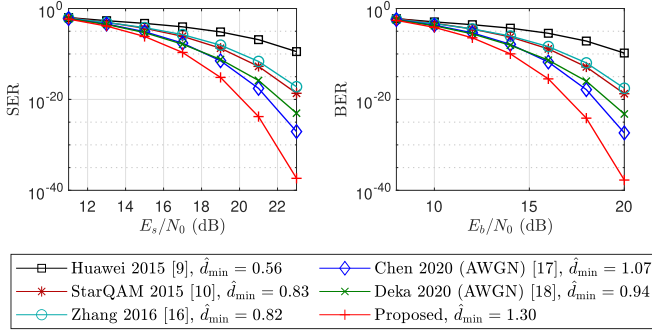
Fig. 12. Performance prediction comparisons over AWGN channel.

$Q(\frac{d_{\min}}{\sqrt{2N_0}}) \gg Q(\sqrt{\frac{d_{kL}}{2N_0}})$ for most $k, l \in \mathcal{Z}_{M^J}, k \neq l$ when the SNR goes to infinity[6] (i.e., $N_0 \to 0$), our approach of maximizing $d_{\min}$ for the codebook collection design problem is thus justified based on the tightness of these bounds. At least, it suggests that the codebook collection designed will have clear advantages in high-SNR regions, as we have shown in this section.

VI. CONCLUSION

In this paper, a new method for downlink SCMA codebook design, based on maximizing the minimum Euclidean distance (MED) of superimposed codewords, is proposed. An iterative algorithm based on alternating maximization is applied by reformulating the MED-maximization problem into a biconvex form with an exact penalty function. With appropriate choices of the initial codebook collection and weighting coefficients, the proposed algorithm has successfully produced a codebook collection with an MED greater than any existing codebook collection with a large margin, for the most popular six-user four-resource case. A Lagrange dual problem of the MED-maximizing problem was derived and solved, resulting in a theoretical MED upper bound of any SCMA codebook collections that were unknown before. Although the codebook collection reported in this article has achieved an MED that is much closer to the upper bound than any previously reported codebook collections, the fact that there is still a nonzero gap between the upper bound and the largest MED suggests there is still room for codebook improvement in the future.

Simulation results demonstrate clear advantages of the obtained largest-MED codebook collection in terms of error-rate performance over all available reported codebook collections. The performance advantages are not only seen in AWGN channels but also in frequency-selective fading channels assuming downlink SCMA resources are allocated from an underlying OFDMA system with resource elements with sufficiently good channel quality, which fits in the downlink scenario of 5G/NR applications, making it a promising codebook candidate in such scenarios. The proposed codebook collection, however, does not show a performance advantage over existing ones in uplink scenarios or scenarios where non-consecutive subcarriers are used as the underlying orthogonal resources. The validity of all

[6]This can be seen by the fact that $\lim\limits_{x \to \infty} Q(ax)/Q(x) = 0$ for all $a > 1$.

simulation curves is further verified by comparing them with theoretical error rate bounds.

In the future, it is still desirable to further reduce the duality gap between the attained MED and the upper bound obtained from the dual problem: either a codebook collection with an even larger MED is to be found, or a smaller upper bound of MED is to be derived, or both. The choices of the initial codebook collection and weighting coefficients appear to play important roles in the proposed biconvex algorithm. It is therefore desirable to try out other possible combinations of these parameters to find even a better codebook collection.

APPENDIX A
DERIVATION OF THE DISTANCE MATRICES IN (15)

First note that the difference of some pair of superimposed codewords corresponding to some two multiplexed symbols $\mathbf{m}_k, \mathbf{m}_l$ is

$$\sum_{j=1}^{J} \mathbf{V}_j \mathbf{C}_j \left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right) = \sum_{j=1}^{J} \text{vec} \left(\mathbf{V}_j \mathbf{C}_j \left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right) \right)$$

$$= \sum_{j=1}^{J} \left[\left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right)^T \otimes \mathbf{V}_j \right] \mathbf{x}_j$$

where $\mathbf{x}_j = \text{vec}(\mathbf{C}_j)$ is actually the j-th MN-block of $\mathbf{x}$. Therefore,

$$\mathbf{x}_j = \left((\mathbf{e}_j^{(J)})^T \otimes \mathbf{I}_{MN} \right) \mathbf{x}. \tag{34}$$

Then the square of the Euclidean distance of this pair will be

$$\left\| \sum_{j=1}^{J} \left[\left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right)^T \otimes \mathbf{V}_j \right] \mathbf{x}_j \right\|_2^2$$

$$= \sum_{q=1}^{J} \mathbf{x}_q^H \left[\left(\mathbf{e}_{k_q}^{(M)} - \mathbf{e}_{l_q}^{(M)} \right)^T \otimes \mathbf{V}_q \right]^H$$

$$\cdot \sum_{j=1}^{J} \left[\left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right)^T \otimes \mathbf{V}_j \right] \mathbf{x}_j$$

$$= \sum_{q=1}^{J} \sum_{j=1}^{J} \mathbf{x}_q^H \left(\left(\mathbf{e}_{k_q}^{(M)} - \mathbf{e}_{l_q}^{(M)} \right) \otimes \mathbf{V}_q^T \right)$$

$$\cdot \left(\left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right)^T \otimes \mathbf{V}_j \right) \mathbf{x}_j$$

$$= \sum_{q=1}^{J} \sum_{j=1}^{J} \left[\mathbf{x}_q^H \left(\left(\mathbf{e}_{k_q}^{(M)} - \mathbf{e}_{l_q}^{(M)} \right) \left(\mathbf{e}_{k_j}^{(M)} - \mathbf{e}_{l_j}^{(M)} \right)^T \right) \right.$$

$$\left. \otimes \left(\mathbf{V}_q^T \mathbf{V}_j \right) \mathbf{x}_j \right].$$

In the end, by substituting (34) for $\mathbf{x}_j$ for all $j \in \mathcal{Z}_J$, we have the expression of $\mathbf{x}^H \mathbf{A}_i \mathbf{x}$ as shown in Section III-A2. Note that the index $i \in \mathcal{Z}_{\binom{M^J}{2}}$ here can be mapped freely to the $\binom{M^J}{2}$ pairs

of multiplexed symbols as long as the mapping is confirmed to be one-to-one.

APPENDIX B
THE CODEBOOK COLLECTION OBTAINED BY ALGORITHM 1 WITH CHEN'S AWGN CODEBOOK COLLECTION [17] FOR INITIALIZATION

$\mathcal{S}_1 - \mathcal{S}_6$, shown as follows.

$$\mathcal{S}_1 = \left\{ \begin{bmatrix} -0.9643 - 0.0000i \\ -0.0698 - 0.2553i \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -0.2562 - 0.0664i \\ 0.4795 + 0.8367i \\ 0 \\ 0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0.2562 + 0.0664i \\ -0.4795 - 0.8367i \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.9643 - 0.0000i \\ 0.0698 + 0.2553i \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{S}_2 = \left\{ \begin{bmatrix} 0 \\ 0 \\ -0.3519 + 0.8978i \\ -0.2121 - 0.1584i \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -0.0317 + 0.2628i \\ 0.6031 + 0.7525i \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0 \\ 0 \\ 0.0317 - 0.2628i \\ -0.6031 - 0.7525i \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0.3519 - 0.8978i \\ 0.2121 + 0.1584i \end{bmatrix} \right\}$$

$$\mathcal{S}_3 = \left\{ \begin{bmatrix} -0.2821 + 0.8866i \\ 0 \\ -0.3653 + 0.0320i \\ 0 \end{bmatrix}, \begin{bmatrix} -0.1631 + 0.3284i \\ 0 \\ 0.9284 + 0.0609i \\ 0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0.1631 - 0.3284i \\ 0 \\ -0.9284 - 0.0609i \\ 0 \end{bmatrix}, \begin{bmatrix} 0.2821 - 0.8866i \\ 0 \\ 0.3653 - 0.0320i \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{S}_4 = \left\{ \begin{bmatrix} 0 \\ -0.9095 + 0.1960i \\ 0 \\ -0.1543 + 0.3326i \end{bmatrix}, \begin{bmatrix} 0 \\ -0.3660 + 0.0218i \\ 0 \\ 0.5154 - 0.7746i \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0 \\ 0.3660 - 0.0218i \\ 0 \\ -0.5154 + 0.7746i \end{bmatrix}, \begin{bmatrix} 0 \\ 0.9095 - 0.1960i \\ 0 \\ 0.1543 - 0.3326i \end{bmatrix} \right\}$$

$$\mathcal{S}_5 = \left\{ \begin{bmatrix} -0.4802 - 0.8223i \\ 0 \\ 0 \\ -0.2843 - 0.1112i \end{bmatrix}, \begin{bmatrix} -0.2646 - 0.1522i \\ 0 \\ 0 \\ 0.9420 - 0.1395i \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0.2646 + 0.1522i \\ 0 \\ 0 \\ -0.9420 + 0.1395i \end{bmatrix}, \begin{bmatrix} 0.4802 + 0.8223i \\ 0 \\ 0 \\ 0.2843 + 0.1112i \end{bmatrix} \right\}$$

$$\mathcal{S}_6 = \left\{ \begin{bmatrix} 0 \\ -0.4747 + 0.8255i \\ -0.0452 - 0.3019i \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -0.0006 + 0.3052i \\ 0.5903 + 0.7472i \\ 0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0 \\ 0.0006 - 0.3052i \\ -0.5903 - 0.7472i \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.4747 - 0.8255i \\ 0.0452 + 0.3019i \\ 0 \end{bmatrix} \right\}$$

APPENDIX C
THE PROPOSED CODEBOOK COLLECTION

$\mathcal{S}_1 - \mathcal{S}_6$, shown as follows.

$$\mathcal{S}_1 = \left\{ \begin{bmatrix} -0.4969 - 0.0000i \\ 0.2516 + 0.8044i \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -0.5790 - 0.0043i \\ -0.7819 - 0.3102i \\ 0 \\ 0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0.5790 + 0.0043i \\ 0.7819 + 0.3102i \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.4969 + 0.0000i \\ -0.2516 - 0.8044i \\ 0 \\ 0 \end{bmatrix} \right\}$$

$$\mathcal{S}_2 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0.3036 - 0.2955i \\ 0.6048 + 0.7008i \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0.7086 - 0.5300i \\ -0.3028 - 0.2982i \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} 0 \\ 0 \\ -0.7086 + 0.5300i \\ 0.3028 + 0.2982i \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -0.3036 + 0.2955i \\ -0.6048 - 0.7008i \end{bmatrix} \right\}$$

$$\mathcal{S}_3 = \left\{ \begin{bmatrix} 0.1787 - 0.7137i \\ 0 \\ -0.5264 - 0.0696i \\ 0 \end{bmatrix}, \begin{bmatrix} 0.8996 - 0.3570i \\ 0 \\ 0.4898 - 0.0103i \\ 0 \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} -0.8996 + 0.3570i \\ 0 \\ -0.4898 + 0.0103i \\ 0 \end{bmatrix}, \begin{bmatrix} -0.1787 + 0.7137i \\ 0 \\ 0.5264 + 0.0696i \\ 0 \end{bmatrix} \right\}$$

$$S_4 = \left\{ \begin{bmatrix} 0 \\ -0.5491+0.0077i \\ 0 \\ 0.3049-0.8077i \end{bmatrix}, \begin{bmatrix} 0 \\ 0.5034+0.0712i \\ 0 \\ 0.7996-0.2351i \end{bmatrix}, \right.$$

$$\begin{bmatrix} 0 \\ -0.5034-0.0712i \\ 0 \\ -0.7996+0.2351i \end{bmatrix}, \left. \begin{bmatrix} 0 \\ 0.5491-0.0077i \\ 0 \\ -0.3049+0.8077i \end{bmatrix} \right\}$$

$$S_5 = \left\{ \begin{bmatrix} 0.2927+0.8221i \\ 0 \\ 0 \\ -0.5292-0.0245i \end{bmatrix}, \begin{bmatrix} -0.7798-0.2581i \\ 0 \\ 0 \\ -0.5316-0.0226i \end{bmatrix}, \right.$$

$$\begin{bmatrix} 0.7798+0.2581i \\ 0 \\ 0 \\ 0.5316+0.0226i \end{bmatrix}, \left. \begin{bmatrix} -0.2927-0.8221i \\ 0 \\ 0 \\ 0.5292+0.0245i \end{bmatrix} \right\}$$

$$S_6 = \left\{ \begin{bmatrix} 0 \\ 0.4950-0.4726i \\ 0.4539+0.5691i \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.5690-0.5140i \\ -0.4126-0.4934i \\ 0 \end{bmatrix}, \right.$$

$$\begin{bmatrix} 0 \\ -0.5690+0.5140i \\ 0.4126+0.4934i \\ 0 \end{bmatrix}, \left. \begin{bmatrix} 0 \\ -0.4950+0.4726i \\ -0.4539-0.5691i \\ 0 \end{bmatrix} \right\}$$

REFERENCES

[1] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, pp. 1501–1505, Dec. 2014.
[2] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I., and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.
[3] Y. Wu, C. Wang, Y. Chen, and A. Bayesteh, "Sparse code multiple access for 5G radio transmission," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, 2017, pp. 1–6.
[4] K. Xiao, B. Xia, Z. Chen, B. Xiao, D. Chen, and S. Ma, "On capacity-based codebook design and advanced decoding for sparse code multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3834–3849, Jun. 2018.
[5] 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. 8, pp. 3160–3174, Aug. 2018.
[6] N. Hosein and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2013, pp. 332–336.
[7] Q. Wang, R. Zhang, L.-L. Yang, and L. Hanzo, "Non-orthogonal multiple access: A unified perspective," *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 10–16, Apr. 2018.
[8] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA Codebook design," in *Proc. IEEE 80th Veh. Technol. Conf.*, 2014, pp. 1–5.
[9] 1st 5G Algorithm Innovation Competition, 2019. [Online]. Available: http://www.innovateasia.com/5g/images/pdf/1st%205G%20Algorithm%20Innovation%20Competition-ENV1.0%20-%20SCMA.pdf

[10] L. Yu, X. Lei, P. Fan, and D. Chen, "An optimized design of SCMA codebook based on star-QAM signaling constellations," in *Proc. IEEE Int. Conf. Wireless Commun. Signal Process.*, 2015, pp. 1–5.
[11] J. Peng, W. Chen, B. Bai, X. Guo, and C. Sun, "Joint optimization of constellation with mapping matrix for SCMA codebook design," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 264–268, Mar. 2017.
[12] M. Alam and Q. Zhang, "Designing optimum mother constellation and codebooks for SCMA," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.
[13] Y. Zhou, Q. Yu, W. Meng, and C. Li, "SCMA codebook design based on constellation rotation," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.
[14] S. Sharma, K. Deka, V. Bhatia, and A. Gupta, "SCMA codebook based on optimization of mutual information and shaping gain," in *Proc. IEEE Globecom Workshops*, 2018, pp. 1–6.
[15] H. Yan, H. Zhao, Z. Lv, and H. Yang, "A top-down SCMA codebook design scheme based on lattice theory," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2016, pp. 1–5.
[16] S. Zhang et al., "A capacity-based codebook design method for sparse code multiple access systems," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process.*, 2016, pp. 1–5.
[17] Y. Chen and J. Chen, "On the design of near-optimal sparse code multiple access codebooks," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2950–2962, May 2020.
[18] K. Deka, M. Priyadarsini, S. Sharma, and B. Beferull-Lozano, "Design of SCMA codebooks using differential evolution," in *Proc. IEEE Int. Conf. Commun. Workshops*, Dublin, Ireland, 2020, pp. 1–7.
[19] 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, Feb. 2019.
[20] C. Jiang and Y. Wang, "An uplink SCMA codebook design combining probabilistic shaping and geometric shaping," *IEEE Access*, vol. 8, pp. 76 726–76 736, 2020.
[21] S. M. Hasan, K. Mahata, and M. M. Hyder, "Sparse code multiple access codebook design using equiangular tight frames," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, 2020, pp. 1–6.
[22] 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 83rd Veh. Technol. Conf.*, 2016, pp. 1–5.
[23] K. Lai, J. Lei, L. Wen, and G. Chen, "Codeword position index modulation design for sparse code multiple access system," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 273–13 288, Nov. 2020.
[24] M. Vameghestahbanati, I. D. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Multidimensional constellations for uplink SCMA systems—A comparative study," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2169–2194, Apr. 2019.
[25] 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. 10 543–10 553, Nov. 2018.
[26] Q. Wang, T. Li, R. Feng, and C. Yang, "An efficient large resource-user scale SCMA codebook design method," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1787–1790, Oct. 2019.
[27] S. Lou, C. Gong, Q. Gao, and Z. Xu, "SCMA with low complexity symmetric codebook design for visible light communication," in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, USA, 2018, pp. 1–6.
[28] H. Nikopour et al., "SCMA for downlink multiple access of 5G wireless networks," in *Proc. IEEE Glob. Commun. Conf.*, Austin, TX, USA, 2014, pp. 3940–3945.
[29] 3GPP, "Study on new radio access technology physical layer aspects," Tech. Rep. TR 38.802, V14.2.0, Sep. 2017.
[30] 3GPP, "Study on non-orthogonal multiple access (NOMA) for NR," Tech. Rep. TR 38.812, V16.0.0, Dec. 2018.
[31] 3GPP, "Physical channels and modulation," Tech. Spec. TS 38.211, V16.2.0, Jul. 2020.
[32] 3GPP, "Physical layer procedures for data," Tech. Spec. TS 38.214, V15.3.0, Oct. 2018.
[33] Y. Huang, R. Yang, and B. Su, "Integrating sparse code multiple access with circularly pulse-shaped OFDM waveform for 5G and the factories of the future," in *Proc. Eur. Conf. Netw. Commun.*, 2019, pp. 20–24.
[34] Y. Lin, Y. Liu, and Y. Siu, "Low complexity message passing algorithm for SCMA system," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2466–2469, Dec. 2016.

[35] 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.

[36] J. Bao, Z. Ma, G. K. Karagiannidis, M. Xiao, and Z. Zhu, "Joint multiuser detection of multidimensional constellations over fading channels," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 161–172, Jan. 2017.

[37] Z. Ma and J. Bao, "Sparse code multiple access (SCMA)," in *Multiple Access Techniques for 5G Wireless Networks and Beyond*, Cham, Switzerland: Springer, 2019, pp. 369–416.

[38] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.2," Jan. 2020. [Online] Available: http://cvxr.com/cvx

[40] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "On the implementation and usage of SDPT3 — A Matlab software package for semidefinite-quadratic-linear programming, version 4.0," in *Handbook on Semidefinite*, Conic and Polynomial Optimization, Boston, MA, USA: Springer, 2012, pp. 715–754.

[41] C. L. Byrne, "Alternating minimization as sequential unconstrained minimization: A survey," *J. Optim. Theory Appl.*, vol. 156, no. 3, pp. 554–566, Mar. 2013.

[42] Ö. T. Demir and T. E. Tuncer, "Multicast beamforming with antenna selection using exact penalty approach," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, Brisbane, Australia,, 2015, pp. 2489–2493.

[43] Ö. T. Demir and T. E. Tuncer, "Alternating maximization algorithm for the broadcast beamforming," in *Proc. 22nd Eur. Signal Process. Conf.*, Lisbon, 2014, pp. 1915–1919.

[44] S. Haykin, *Communication Systems*. New York, NY, USA: Wiley, 2001, pp. 332–335.

[45] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[46] R. A. Horn and C. R. Johnson, "Singular value inequalities," in *Topics in Matrix Analysis*, Cambridge, U.K.: Cambridge Univ. Press, 1991, pp. 134–238.

Borching Su (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering and communication engineering from National Taiwan University, Taipei, Taiwan, in 1999 and 2001, respectively, and the Ph.D. degree in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2008. He was with NextWave Broadband, Inc., San Diego, CA, USA, in 2008. In 2009, he joined National Taiwan University where he is currently an Associate Professor. His current research interests include signal processing for communications, particularly waveform designs for next-generation communication systems.

Tingyi Lin received the B.S. degree in electronic engineering from Chang Gung University, Taoyuan, Taiwan, in 2018, and the M.S. degree in communication engineering from Nation Taiwan University (NTU), Taipei, Taiwan, in 2020. From 2019 to 2020, he experienced the project of EU TW H2020 Converged Wireless Access for Reliable 5G MTC for Factories of the Future. He is currently with RealTek Semiconductor Corp., researching wireless network system.

Yenming Huang (Member, IEEE) received the Ph.D. degree in Communication Engineering from National Taiwan University, Taipei, Taiwan, in January 2019. He is currently a Principal Engineer with the Algorithm Design Department, Tron Future Tech Inc., Hsinchu, Taiwan, and is an Adjunct Assistant Professor with the Graduate Institute of Communication Engineering, National Taiwan University. In the subsequent six months, he was a Postdoctoral Researcher with the Graduate Institute of Communication Engineering, National Taiwan University to execute the EU-TW Clear5G Project. Yenming Huang had many experiences in 4G LTE-A and 5G New Radio (NR) standardization and actively participated in 3GPP RAN1 meetings from 2013 to 2016. He also had much experience in industry-academy cooperation project execution, associated with MediaTek, ASUSTeK. His research interests include signal processing and optimization for communications, including multi-antenna multi-carrier transceiver systems, phased array radar systems, digital calibration for hardware non-ideal effects, next-generation telecommunication key technologies, and signal and data processing algorithms.

Chinwei Huang (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering and communication engineering from National Taiwan University, Taipei, Taiwan, in 2019 and 2021. He is currently a Senior Engineer with Algorithm Design Department at Tron Future Tech Inc., Hsinchu, Taiwan. His research interests include signal processing and optimization for wireless communications and radar systems, including codebook design in sparse code multiple access, filterbank transceiver design, and radar waveform design.