

Low-Complexity Block Coordinate Descend Based Multiuser Detection for Uplink Grant-Free NOMA

Pengyu Gao, Zilong Liu, *Senior Member, IEEE*, Pei Xiao, *Senior Member, IEEE*, Chuan Heng Foh, *Senior Member, IEEE*, and Jing Zhang

Abstract—Grant-free non-orthogonal multiple access (NOMA) scheme is considered as a promising candidate for the enabling of massive connectivity and reduced signalling overhead for Internet of Things (IoT) applications in massive machine-type communication (mMTC) networks. Exploiting the inherent nature of sporadic transmissions in the grant-free NOMA systems, compressed sensing based multiuser detection (CS-MUD) has been deemed as a powerful solution to user activity detection (UAD) and data detection (DD). In this paper, block coordinate descend (BCD) method is employed in CS-MUD to reduce the computational complexity. We propose two modified BCD based algorithms, called enhanced BCD (EBCD) and complexity reduction enhanced BCD (CR-EBCD), respectively. To be specific, by incorporating a novel candidate set pruning mechanism into the original BCD framework, our proposed EBCD algorithm achieves remarkable CS-MUD performance improvement. In addition, the proposed CR-EBCD algorithm further ameliorates the proposed EBCD by eliminating the redundant matrix multiplications during the iteration process. As a consequence, compared with the proposed EBCD algorithm, our proposed CR-EBCD algorithm enjoys two orders of magnitude complexity saving without any CS-MUD performance degradation, rendering it a viable solution for future mMTC scenarios. Extensive simulation results demonstrate the bound-approaching performance as well as ultra-low computational complexity.

Index Terms—Massive machine type communication (mMTC), grant-free, non-orthogonal multiple access (NOMA), compressed sensing (CS), block coordinate descend (BCD), multi-user detection (MUD), user activity detection (UAD), data detection (DD).

I. INTRODUCTION

Massive machine-type communication (mMTC) is one of the three major use cases in the fifth generation (5G) mobile networks, which enables ubiquitous connectivity for various Internet-of-Things (IoT) applications [1]-[3]. Different from conventional human-centric communications, the design of an mMTC network needs to consider certain inherent features, such as small data size, sporadic traffic, low power consumption, etc. [4], [5].

Towards massive connectivity, the emerging non-orthogonal multiple access (NOMA) technique [6]-[8] is deemed as a promising candidate to accommodate a massive number of IoT devices. Unlike the classical orthogonal multiple access (OMA) techniques [6] where the number of supported devices is at most the number of available orthogonal resources,

NOMA can concurrently support more communication links by sharing common resources in a non-orthogonal fashion. Thus, NOMA permits a higher spectrum efficiency compared with OMA, albeit at the expense of increased computational complexity at the receiver [7]. In addition, NOMA can be generally divided into two main categories, code-domain NOMA [6] and power-domain NOMA [8], respectively. In this paper, we consider sequence based code-domain NOMA, where massive users share the same time-frequency resources, but employ user-specific, non-orthogonal spreading sequences. Another challenging problem for mMTC is how each IoT device can rapidly access a channel resource. The long-term evolution (LTE) standard adopts a grant based uplink random access (RA) protocol with a four-phase handshaking procedure [9]. When IoT devices transmit small packets to the base station (BS) sporadically, however, the complicated signaling interactions between the BS and IoT devices result in potentially excessive signalling overhead and large access latency.

As a promising solution to the aforementioned problems, uplink grant-free NOMA communication is carried out based on non-orthogonal resource allocation in the code-domain, where each resource unit may be shared by more than one user [10]. Hence, it not only enables massive connectivity, but also allows active users transmit freely without waiting the grant of a communication network [10]-[13]. However, since the uplink channel resources allocated to the users in uplink grant-free NOMA are non-orthogonal, complicated multiple users detection may be needed at the BS. Note that in power-domain NOMA, unequal power allocation is adopted to separate different users. By contrast, code-domain NOMA allows equal power allocation to different users and the decoding does not rely on successive interference cancellation (SIC). Instead, in uplink grant-free NOMA, compressed sensing (CS) theory [14]-[17] is employed to jointly carry out user activity detection and data detection by exploiting the sporadic nature of mMTC transmissions. Specifically, CS is a powerful signal processing technique to reconstruct sparse signals from far fewer samples than that required by the Nyquist-Shannon sampling theorem [15]. So far, CS has been widely applied in wireless communications [18], such as channel estimation [19] and spectrum sensing [20].

By leveraging the feature of sporadic communications in mMTC, numerous CS-based multiuser detections (CS-MUD) have been studied in recent years [21]-[29]. Among these works, [21]-[23] only focused on single time-slot transmission model with no consideration of the temporal correlation. To further improve the performance of CS-MUD, frame-wise joint sparsity has been exploited, where the inactive and active users remain their individual transmission status during a whole frame [24]-[29]. In this case, CS-MUD at the BS can be

P. Gao, P. Xiao and C. Foh are with 5G and 6G Innovation centre, Institute for Communication Systems (ICS) of University of Surrey, Guildford, GU2 7XH, UK (e-mail: p.gao@surrey.ac.uk; p.xiao@surrey.ac.uk; c.foh@surrey.ac.uk).

Z. Liu is with the School of Computer Science and Electronics Engineering, University of Essex, Colchester, CO4 3SQ, UK (e-mail: zilong.liu@essex.ac.uk).

J. Zhang is with Institute of space-based information systems, China Academy of Electronic and Information Technology (CAEIT), No. 11, Shuangyuan Road, Beijing 100041, China (e-mail: zhangjing-7@chinagci.com).

generalized into a multiple measurement vector (MMV) problem. In [24], a modified orthogonal matching pursuit (OMP) algorithm [16], called iterative order recursive least square (IORS), was developed for CS-MUD. To take full advantage of the frame-wise joint sparsity, Du *et al.* in [25] integrated the block sparsity structure into the classical subspace pursuit (SP) algorithm [17] and proposed the threshold-aided block sparsity adaptive subspace pursuit (TA-BSASP) that can approach the oracle performance with no need of the user sparsity level as the prior knowledge. Furthermore, subspace matching pursuit based symbol correcting mechanism (SMP-SCM) in [26] has shown improved symbol error rate (SER) over oracle least square (LS) [11] at the expense of computational cost. However, a salient drawback of these greedy-based algorithms is the computationally prohibitive operations of the matrix multiplications and inversions, especially for large-scale problems in mMTC scenarios. In parallel with greedy-based algorithms, many existing CS-MUD schemes have been developed based on the Bayesian approach [27]-[29]. The authors in [27] utilized approximate message passing (AMP) and expectation maximization (EM) to carry out MUD with the aid of the prior information of the transmitted symbols. Moreover, [28] introduced a maximum a posteriori probability (MAP) based active user detector by exchanging the extrinsic information. Mei *et al.* in [29] developed an orthogonal approximate message passing (OAMP) based algorithm by leveraging the prior knowledge of the discrete constellation symbols and structured transmission sparsity. However, the Bayesian-based methods may suffer from sophisticated probabilistic iterations and the inaccurate Gaussian approximation at the receiver side, precluding them from efficient practical implementations.

This paper presents the design of a novel CS-MUD detector which enjoys an excellent MUD performance and low computational complexity. For this purpose, we first transform the MUD problem into a group least absolute shrinkage and selection operator (LASSO) problem and then address it globally by the block coordinate descent (BCD) based algorithms [30]. Particularly, in the conventional BCD framework, the formulated non-convex sparse recovery problem can be decomposed into a sequence of small-scale subproblems after exploiting LASSO based regularization [31]. Subsequently, the variables in each subproblem can be optimized according to the closed-form solution sequentially with variables from other subproblems keeping fixed. In uplink grant-free NOMA systems, the updating rules of the BCD method indicates that all signals, regardless of active or inactive users, are calculated and then further reused to update others signals. In particular, we observe that estimating the signals of inactive users indeed is to reconstruct the noise. In this case, severe inter-user interference may be inevitable, making active users indistinguishable. Motivated by this, we first develop an enhanced BCD (EBCD) algorithm which can identify a majority of inactive users in advance, thus helping significantly reducing inter-user interference. Next, we further improve the proposed EBCD algorithm for computational complexity saving. The resultant algorithm is called complexity-reduction enhanced BCD (CR-EBCD), which shares the same MUD performance as that of EBCD, whilst having significantly reduced computational complexity. Without the need of knowing the user sparsity level as priori information, our proposed two BCD-based algorithms enable the identification of active users

adaptively upon the energy-based threshold. Specifically, our major contributions are summarized as follows:

- **Proposed EBCD Algorithm:** In the proposed EBCD algorithm, we introduce a candidate set pruning mechanism to improve the CS-MUD performance. More explicitly, different from the original BCD method where all potential users are involved in data updating in each iteration, our proposed EBCD algorithm discards inactive users according to the signal power at the early iterations. Furthermore, the corresponding signals of those discarded users are set to zeros for direct interference mitigation. Therefore, by utilizing the pruning candidate, the CS-MUD performance is significantly improved compared to that of the original BCD algorithm. In addition, with the decreasing number of users in the candidate set, the computational overhead is also reduced.
- **Proposed CR-EBCD Algorithm:** As a further extension of the proposed EBCD, our proposed CR-EBCD dramatically simplifies the iteration process by removing the redundant matrix multiplications. Consequently, the calculations in the proposed CR-EBCD are reduced into the simple calculation of vector multiplications, which avoids complicated large-scale matrix multiplications. Both algorithms are able to achieve the same bound-approaching performance with the aid of the candidate set pruning mechanism, but the proposed CR-EBCD can also enjoy about two orders of magnitude computational complexity reduction. Therefore, our proposed CR-EBCD strikes an excellent tradeoff between the detection performance and computational complexity.

It is important to point out that the core idea of the previous studies [32]-[33] focusing on low-complexity CS-MUD design in uplink grant-free NOMA systems is to substitute the complicated matrix inversion operations by simple gradient descend. However, since they follow greedy-based algorithms, calculating the correlation at each iteration inevitably leads to high computational cost due to the large-scale matrix multiplications. By contrast, the simple operations between vectors in our proposed BCD-based algorithms not only avoid the matrix inversion operations, but require no matrix multiplications.

The rest of the paper is organized as follows. In Section II, the system model and problem formulation are presented. Next, the original BCD algorithm, our proposed EBCD and CR-EBCD algorithms are detailed sequentially in Section III. Then, Section IV provides the computational complexity analysis. Simulation results are given in Section V to verify the effectiveness of our proposed algorithms. Finally, Section VI concludes the paper.

Notations: Boldface capital and lowercase symbols represent matrices and column vectors, respectively. The $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ operations represent transpose, Hermitian transpose and inverse, respectively. $(\cdot)^\dagger$ denotes pseudo-inverse operation. $\|\cdot\|_p$ is the l_p -norm operation. Besides, $\text{diag}(\cdot)$ denotes a diagonal matrix with the elements on its diagonal. \odot stands for the Hadamard product. \mathbf{I}_N means the identity matrix of size $N \times N$. Additionally, $\max(\Omega, k)$ represents selecting the k largest elements from the set Ω . Furthermore, for a matrix \mathbf{Z} , $\mathbf{Z}[\Lambda]$ refers to the sub-matrix of \mathbf{Z} that only contains those columns indexed by the set Λ . $\{1, 2, \dots, K\} \setminus \Lambda$ represents that the set contains $\{1, 2, \dots, K\}$, but excludes the elements in set Λ . Finally, $\mathcal{CN}(0, \sigma^2)$ represents the complex Gaussian

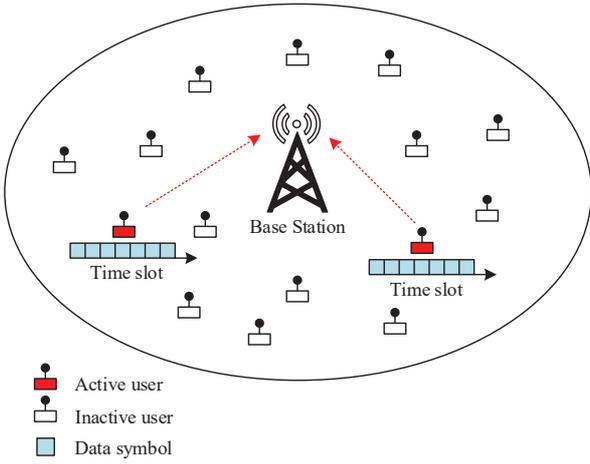


Fig. 1. System model of an uplink grant-free NOMA system in one transmission frame. Due to sporadic nature of mMTC transmissions, only a small fraction of users are active while others keep silent. Additionally, the transmission status of active or inactive users remains unchanged during an entire frame.

distribution with zero mean and variance σ^2 .

II. SYSTEM MODEL

Consider an uplink grant-free NOMA system, where users are static (or near static) and a BS communicates with a total of fixed K users [34], as illustrated in Fig. 1. Without loss of generality, we assume that the BS and all users are equipped with a single antenna.

In an arbitrary time interval, although the total number of users may be very large, it is reasonable to assume that only a small fraction of users are active due to the sporadic traffic nature of the mMTC transmissions. For active user k , the transmitted symbol x_k is selected from a complex-constellation set χ . Meanwhile, if user k is silent, x_k can be regarded as zero. Hence, the augmented complex-constellation set $\tilde{\chi}$ can be represented by $\tilde{\chi} \triangleq \{\chi \cup 0\}$. Then, data symbols are spread onto N subcarriers by user-specific spreading sequences. Subsequently, the signals from all the active users are superposed over the same N orthogonal resources at the receiver side. Particularly, to meet the demand of massive connectivity in mMTC scenarios, code-domain NOMA technique is considered and non-orthogonal spreading sequences are assigned to each user, indicating that the length of each spreading sequence is less than the total number of users in the systems, i.e., $N < K$.

At the BS, the frequency-domain received signal $\mathbf{y} \in \mathbb{C}^{N \times 1}$ can be expressed as

$$\mathbf{y} = \sum_{k=1}^K (\mathbf{h}_k \odot \mathbf{s}_k) x_k + \mathbf{v}, \quad (1)$$

where $\mathbf{h}_k = [h_{1,k}, h_{2,k}, \dots, h_{N,k}]^T \in \mathbb{C}^{N \times 1}$ is the channel coefficients between user k and the BS, whose elements follow the independent complex Gaussian distributions $h_{n,k} \sim \mathcal{CN}(0, 1)$. Here, as many previous works [24]-[29] did, perfect multi-user synchronization¹ and reliable channel estimation² are assumed at the BS. $\mathbf{s}_k = [s_{1,k}, s_{2,k}, \dots, s_{N,k}]^T \in \mathbb{C}^{N \times 1}$ is the unique spreading sequence of user k . In addition, x_k represents the transmitted constellation symbol from active user k to the BS. For the transmission power control, we have

$E[|x_k|^2] = 1$. Moreover, $\mathbf{v} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$ is the additive white Gaussian noise (AWGN).

Accordingly, (1) can be equivalently rewritten as

$$\mathbf{y} = \sum_{k=1}^K \text{diag}(\mathbf{h}_k) \mathbf{s}_k x_k + \mathbf{v} = \mathbf{G} \mathbf{x} + \mathbf{v}, \quad (2)$$

where $\mathbf{G} \in \mathbb{C}^{N \times K}$ denotes the equivalent channel matrix with elements $g_{n,k} = h_{n,k} s_{n,k}$. The transmitted signals of all K users are represented as $\mathbf{x} = [x_1, x_2, \dots, x_K]^T \in \mathbb{C}^{K \times 1}$.

In this paper, we consider the frame-wise joint sparsity model, where the transmission status of active and inactive users remain unchanged during an entire frame, as shown in Fig. 1. Assuming that a data frame contains J continuous time slots, the frame-wise joint sparsity can be formulated as

$$\text{supp}(\mathbf{x}^{(1)}) = \dots = \text{supp}(\mathbf{x}^{(J)}) = \Gamma, \quad (3)$$

where $\text{supp}(\mathbf{x}^{(j)}) = \{k \mid x_k^{(j)} \neq 0, 1 \leq k \leq K\}$, the support Γ can be described as the candidate set, which indicates the set of active users. $|\Gamma|$ denotes the total number of active users. Furthermore, the channel remains unchanged during a whole frame, i.e. $\mathbf{G}^{(1)} = \dots = \mathbf{G}^{(J)} = \mathbf{G}$. Due to the frame-wise joint sparsity and static channel, the received signals in frequency-domain during a certain frame can be rewritten as

$$\mathbf{Y} = \mathbf{G} \mathbf{X} + \mathbf{V} = \sum_{k=1}^K \mathbf{g}_k \mathbf{x}_k^T + \mathbf{V}, \quad (4)$$

where $\mathbf{Y} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(J)}] \in \mathbb{C}^{N \times J}$, $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(J)}] \in \mathbb{C}^{K \times J}$ and $\mathbf{V} = [\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(J)}] \in \mathbb{C}^{N \times J}$, respectively. In addition, $\mathbf{g}_k \in \mathbb{C}^{N \times 1}$ denotes the k -th column of \mathbf{G} , which represents the equivalent channel coefficients of user k . $\mathbf{x}_k^T \in \mathbb{C}^{1 \times J}$ denotes the k -th row of \mathbf{X} , representing the data symbols of user k transmitting over the consecutive J time slots.

The BS aims at recovering the transmitted signal \mathbf{X} based on the received signal \mathbf{Y} . It is noted that each column in \mathbf{X} is a sparse vector due to the sporadic transmission nature of mMTC data traffic. Moreover, with the frame-wise joint sparsity transmission model, \mathbf{X} is a sparse matrix consisting of many zero rows. Thus, the sparse construction problem in (3) can be regarded as a multiple measurement vector (MMV) based CS problem.

III. THE PROPOSED BCD BASED ALGORITHMS

In this section, we first introduce the original BCD algorithm for uplink grant-free NOMA systems. Subsequently, we present our proposed EBCD algorithm which is shown to achieve a near-oracle MUD performance and enjoy low computational complexity, thanks to the candidate set pruning mechanism. Finally, our proposed CR-EBCD algorithm for further computational complexity reduction is described.

¹This can be achieved by sending beacon signals periodically from the BS [35]. At the same time, all the uplink users constantly adjust their synchronization points by aligning with that of the beacon signals.

²We assume that channel estimation is carried out at the BS for all users (including those inactive ones) through a dedicated control channel. Furthermore, since we consider the mMTC scenario, users in the systems are mostly static with fixed locations [5]. Thus, it is reasonable that the channel information obtained by the BS can be utilized during a long data transmission period.

A. The Original BCD Algorithm

The objective of the BS is to solve the sparse recovery problem in (4), which can be equivalently formulated into an optimization problem as follows:

$$\begin{aligned} \min_{\{\mathbf{x}_k\}_{k=1}^K} & \left\| \mathbf{Y} - \sum_{k=1}^K \mathbf{g}_k \mathbf{x}_k^T \right\|_2^2, \\ \text{s.t.} & \sum_{k=1}^K \|\mathbf{x}_k^T\|_0 \leq K_{\max}^{\text{active}}, \end{aligned} \quad (5)$$

where \mathbf{g}_k is the k -th column of \mathbf{G} and K_{\max}^{active} denotes the maximum number of active users in the system. Since the optimization problem in (5) is NP-hard due to the non-convex sparsity constraint [36], it is thus computationally intractable. A straightforward solution is to carry out exhaustive search for the maximum likelihood (ML) detection [26]. However, this involves a total number of $(M+1)^K$ options, considering all possible active user combinations and the constellation symbol combinations, where M is the modulation cardinality. For instance, when quadrature phase shift keying (QPSK) modulation is adopted, $M = 4$. As such, the computational complexity of the exhaustive search in ML method is prohibitively high, which precludes it from practical implementation especially for large value of K in the future mMTC scenarios.

To tackle this bottleneck, the group LASSO technique can be employed, which provides the convex relaxation of such discontinuous non-convex constraint in (5). Based on the group LASSO framework, (5) can be approximated in an equivalent penalized form as

$$\min_{\{\mathbf{x}_k\}_{k=1}^K} \frac{1}{2} \left\| \mathbf{Y} - \sum_{k=1}^K \mathbf{g}_k \mathbf{x}_k^T \right\|_2^2 + \lambda \sum_{k=1}^K \|\mathbf{x}_k\|_2^2, \quad (6)$$

where $\lambda > 0$ is the penalty parameter for sparsity constraint, also called the tuning factor, which is usually selected by trial-and-error. More importantly, the choice of λ should be meticulous, since it is critical for obtaining the optimal solution towards the above problem. Specifically, given a large value of λ , the solution of (6) could be sufficiently sparse. On the other hand, if the value of λ approaches to zero, the sparsity of the solution tends to be K . It is noted that (6) is a convex optimization problem and the optimal objective is to identify non-zero rows in \mathbf{X} . In the following, a series of the BCD based algorithms are applied, which can address this group LASSO problem efficiently in a globally optimal way.

In the BCD framework, (6) can be decoupled into a sequence of K small subproblems. In particular, each group of variables \mathbf{x}_k in certain subproblem will be optimized one by one while keeping other group of variables \mathbf{x}_l ($l \in \{1, 2, \dots, K\} \setminus \{k\}$) fixed. To be specific, the k -th subproblem degenerating from (6) can be expressed as

$$\min_{\mathbf{x}_k} \frac{1}{2} \left\| \mathbf{R}_k - \mathbf{g}_k \mathbf{x}_k^T \right\|_2^2 + \lambda \|\mathbf{x}_k\|_2^2, \quad (7)$$

where $\mathbf{R}_k = \mathbf{Y} - \sum_{l=1, l \neq k}^K \mathbf{g}_l \mathbf{x}_l^T$. Here, we define a convex function as follows

$$f(\mathbf{x}_k) \triangleq \frac{1}{2} \left\| \mathbf{R}_k - \mathbf{g}_k \mathbf{x}_k^T \right\|_2^2 + \lambda \|\mathbf{x}_k\|_2^2. \quad (8)$$

According to (8), we can obtain the optimal solution of \mathbf{x}_k in a closed form by setting the first-order derivation of $f(\mathbf{x}_k)$ to zero. Hence, the optimal solution \mathbf{x}_k^* can be obtained by

$$(\mathbf{x}_k^*)^T = \frac{\mathbf{g}_k^H \mathbf{R}_k}{\mathbf{g}_k^H \mathbf{g}_k + \lambda_0}, \quad (9)$$

where $\lambda_0 = 2\lambda$. Updating the optimal solution following the (8) and (9), the results will reach convergence after sufficient iterations.

The main procedure of the original BCD algorithm is stated as **Algorithm 1**. As can be seen, the data updating process happens in Step 2. In the sequel, the energy-based threshold is utilized to identify the non-zero positions in Step 3. Particularly, V_{th} is relative to the signal-to-noise ratio, which can be selected empirically [11]. After total T iterations, once the signal power of user k exceeds the threshold, i.e., $\|\hat{\mathbf{x}}_k\|_2^2 > V_{th}$, user k will be regarded as active user and the index k will be stored in the final candidate set Γ . Finally, for a more balanced SER performance and computational complexity trade-off, LS method is exploited to recover the transmitted signal with the aid of the estimated candidate set in Step 4, rather than directly using the results from the previous iteration process. The reason is that the interference introduced by inactive users have adverse influences on estimating the transmitted signals. Conversely, as for the LS method, we only use the estimated support to reconstruct the signals for a more reliable detection results.

Algorithm 1 The Original BCD Algorithm.

Input:

- The received signal: \mathbf{Y} ;
- Equivalent channel matrix: \mathbf{G} ;
- The maximum number of iterations: T ;
- The equivalent regularization parameter: λ_0 ;
- The threshold parameter: V_{th} ;

Output:

Reconstructed sparse signals: $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1^T, \hat{\mathbf{x}}_2^T, \dots, \hat{\mathbf{x}}_K^T]$.

• Step 1 (Initialization)

- 1: (Iteration initialization): the iterative index $t = 1$, the support set $\Gamma = \emptyset$, the initial transmitted signal $\hat{\mathbf{x}}_k = \mathbf{0}, \forall k = 1, 2, \dots, K$.

• Step 2 (Iteration)

for $t = 1, 2, \dots, T$ **do**
 for $k = 1, 2, \dots, K$ **do**

- 2: Compute $\mathbf{R}_k^{(t)}$.
- 3: Update $\hat{\mathbf{x}}_{k,(t)}^T$ using (9).

end for

end for

• Step 3 (Obtain The Estimated Candidate Set)

for $k = 1, 2, \dots, K$ **do**

if $\|\hat{\mathbf{x}}_k\|_2^2 > V_{th}$ **then**

- 4: $\Gamma \leftarrow \Gamma \cup \{k\}$.

end if

end for

• Step 4 (Obtain Data Symbols)

- 5: $\hat{\mathbf{X}}[\Gamma] = (\mathbf{G}[\Gamma])^\dagger \mathbf{Y}$, $\hat{\mathbf{X}}[\{1, 2, \dots, K\} \setminus \Gamma] = \mathbf{0}$.
-

B. The Proposed EBCD Algorithm

Note that the original BCD method is a global optimization algorithm, where the variables from \mathbf{x}_1^T to \mathbf{x}_K^T are updated sequentially at each iteration. However, since the majority of elements in \mathbf{x}_k^T are zeros due to sporadic transmissions, the updating process of those zero positions can be regarded as computational wastage. Moreover, updating the signals of inactive users indeed is to reconstruct the noise. In this case, signals from inactive users are bound to deteriorate the

detection performance since they are used to estimate other users' signals at each iteration when calculating \mathbf{R}_k . Motivated by this, we develop a novel EBCD algorithm, which can identify the majority of inactive users in advance, thanks to the design of the candidate set pruning mechanism. To be specific, in our proposed EBCD algorithm, a majority of inactive users can obtain their own optimal solutions (i.e., $\mathbf{0}$) in advance instead of running the iterative algorithm until convergence, which is beneficial for the identification of active users due to the interference mitigation. Furthermore, since a great number of inactive users can be discarded halfway through the iteration process, our proposed EBCD algorithm also enjoys a considerable computational complexity reduction.

For describing our proposed EBCD algorithm explicitly, we firstly define a parameter $\Omega^{(t)}$ to indicate the candidate set at the t -th iteration and $\Omega_i^{(t)}$ denotes the i -th element in $\Omega^{(t)}$. Correspondingly, $K_a^{(t)}$ is defined to measure the scale of $\Omega^{(t)}$. In particular, the candidate set Ω in the original BCD method consists of all the potential users in the system, i.e. $\Omega = \{1, 2, \dots, K\}$. In other words, if $K_a^{(t)}$ always equals to K at each iteration, our proposed EBCD algorithm boils down to the original one. Therefore, at each iteration, only $K_a^{(t)}$ elements in $\Omega^{(t)}$ need to be updated according to (9).

With respect to the candidate set pruning, the signal power is employed as the criteria, since a higher value of $\|\hat{\mathbf{x}}_{\Omega_i^{(t)}}\|_2^2$ signifies that the user with the corresponding index $\Omega_i^{(t)}$ is more likely to be active. According to this principle, $\psi^{(t)}$ users holding the lower signal power are deemed as inactive users and will be discarded from the current candidate set $\Omega^{(t)}$. Therefore, the scale of the new candidate set $\Omega^{(t+1)}$ is updated by

$$K_a^{(t+1)} = K_a^{(t)} - \psi^{(t)}. \quad (10)$$

Then, we can obtain a new candidate set $\Omega^{(t+1)}$ by pruning the old candidate set $\Omega^{(t)}$ as follows:

$$\Omega^{(t+1)} = \max \left(\|\hat{\mathbf{x}}_{\Omega_i^{(t)}}\|_2^2, K_a^{(t+1)} \right), \Omega_i^{(t)} \in \Omega^{(t)}. \quad (11)$$

The key idea of (11) is to select the largest $K_a^{(t+1)}$ elements from $\Omega^{(t)}$ to constitute the new candidate set $\Omega^{(t+1)}$. Meanwhile, the remaining elements from $\Omega^{(t)}/\Omega^{(t+1)}$ are set to zeros, i.e. $\hat{\mathbf{x}}_{\Omega^{(t)}/\Omega^{(t+1)}} = \mathbf{0}$. Actually, zeros are the optimal solutions for the inactive users. Hence, (10) and (11) can not only prune the candidate set for the next iteration, but also provide the optimal solutions for $\psi^{(t)}$ discarded users. That is, in the subsequent iterations, the amount of computations will be decreased while the calculation results of \mathbf{R}_k will also be more accurate for estimating the active users' signals. It is important to note that after T_b iterations, the final candidate set keeps invariant and only contains \overline{K}_a users, which can be calculated by

$$\overline{K}_a = K - \sum_{t=1}^{T_b} \psi^{(t)}. \quad (12)$$

To avoid discarding the non-zero candidates, we set $\overline{K}_a \geq 0.2K$, considering the sporadic transmission in the practical scenarios³.

It is noted that the procedure of our proposed EBCD algorithm will be exhibited together with our proposed CR-EBCD algorithm in order to avoid too much repetition. In the

sequel, it will be explained detailedly.

Algorithm 2 Proposed CR-EBCD Algorithm.

Input:

The received signal: \mathbf{Y} ;
 Equivalent channel matrix: \mathbf{G} ;
 The maximum number of iterations: T ;
 The equivalent regularization parameter: λ_0 ;
 The threshold parameter: V_{th} ;
 The parameters for pruning the candidate set:
 $\Psi = \{\psi^{(1)}, \psi^{(2)}, \dots, \psi^{(T_b)}\}$;

Output:

Reconstructed sparse signals: $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1^T, \hat{\mathbf{x}}_2^T, \dots, \hat{\mathbf{x}}_k^T]$.

• Step 1 (Initialization)

1: (Iteration initialization): The iterative index $t = 1$, the support set $\Gamma = \emptyset$, the initial candidate set $\Omega^{(1)} = \{1, 2, \dots, K\}$, the initial scale of the candidate set $K_a^{(1)} = K$, the transmitted signals $\hat{\mathbf{x}}_{k,(1)}^T = \mathbf{0}, \forall k = 1, 2, \dots, K$.

• Step 2 (Iteration)

for $t = 1, 2, \dots, T$ **do**
 for $i = 1, \dots, K_a^{(t)}$ **do**
 if $i = 1$ **then**
 if $t = 1$ **then**
 $\mathbf{W}_{\Omega_i^{(t)}}^{(t)} = \mathbf{0}$;
 else if $2 \leq t \leq T_b$ **then**
 $\mathbf{W}_{\Omega_i^{(t)}}^{(t)} \leftarrow$ using (14);
 else
 $\mathbf{W}_{\Omega_i^{(t)}}^{(t)} \leftarrow$ using (16);
 end if
 else
 $\mathbf{W}_{\Omega_i^{(t)}}^{(t)} \leftarrow$ using (19);
 end if
 Compute $\mathbf{R}_{\Omega_i^{(t)}}^{(t)}$.
 Update $\hat{\mathbf{x}}_{\Omega_i^{(t)}}^T$ using (9).
 end for
 while $t \leq T_b$ **do**
 Obtain $\Omega^{(t+1)}$ using (10) and (11).
 $\hat{\mathbf{x}}_{\Omega^{(t)}/\Omega^{(t+1)}} = \mathbf{0}$.
 end while
 end for
 • Step 3 (Obtain The Estimated Candidate Set)
 10: Use the method in Algorithm 1.
 • Step 4 (Obtain Data Symbols)
 11: Use the method in Algorithm 1.

C. The Proposed CR-EBCD Algorithm

In the above two BCD based algorithms, the process of calculating $\mathbf{R}_k^{(t)}$ at each iteration dominates the total computational complexity, since it needs to perform a large-scale matrix multiplication. However, at the arbitrary t -th iteration, a large amount of calculations for obtaining $\mathbf{R}_k^{(t)}$ are futile. For instance, when solving the third subproblem for acquiring $\mathbf{R}_3^{(t)}$, the result of $\mathbf{g}_1 \hat{\mathbf{x}}_{1,(t)}^T$ is requisite. In fact, $\mathbf{g}_1 \hat{\mathbf{x}}_{1,(t)}^T$ has

³Under normal circumstances, the number of active users does not exceed 10% of the total number of potential users even in the busy hours [37], [38].

TABLE I
A LIST OF IMPORTANT VARIABLES

Variables	Descriptions
$\Omega^{(t)}$	the candidate set at the t -th iteration
$\Omega_n^{(t)}$	the n -th user in $\Omega^{(t)}$
T_b	the number of iterations in pruning the candidate set
$K_a^{(t)}$	the size of $\Omega^{(t)}$
\bar{K}_a	the size of the final constant candidate set after T_b iterations
$\mathbf{W}_n^{(t)}$	the auxiliary matrix of user n at the t -th iteration
$\hat{\mathbf{x}}_{n,(t)}^T$	the signals of user n at the t -th iteration
\mathbf{g}_n	the equivalent channel coefficients of user n at the t -th iteration

been calculated when obtaining $\mathbf{R}_2^{(t)}$. Thus, the process of calculating $\mathbf{g}_1 \hat{\mathbf{x}}_{1,(t)}^T$ can be regarded as computational wastage. Against this background, the CR-EBCD algorithm is developed to avoid such unnecessary computational overhead.

For simplicity, we introduce an auxiliary matrix $\mathbf{W}_{\Omega_i^{(t)}} \triangleq \sum_{l \in \Omega^{(t)}, l \neq \Omega_i^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T$. In addition, $\mathbf{W}_{\Omega_i^{(t)}}^{(t)}$ can be further categorized into three different cases and represented in the following expression:

$$\mathbf{W}_{\Omega_i^{(t)}}^{(t)} = \begin{cases} \sum_{l \in \Omega_n^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, & i = 1, \\ \sum_{l \in \Omega_m^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T + \sum_{l \in \Omega_n^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, & 2 \leq i \leq K_a^{(t)} - 1, \\ \sum_{l \in \Omega_m^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T, & i = K_a^{(t)}, \end{cases} \quad (13)$$

where $1 \leq m \leq i-1$ and $i+1 \leq n \leq K_a^{(t)}$. $\hat{\mathbf{x}}_{l,(t-1)}^T$ and $\hat{\mathbf{x}}_{l,(t)}^T$ denote the perviously estimated signal at the $(t-1)$ -th iteration and the newly estimated signal at the current t -th iteration, respectively. Then, according to the above three different cases, we will analyse the relationship between $\mathbf{W}_{\Omega_i^{(t-1)}}^{(t-1)}$ and $\mathbf{W}_{\Omega_i^{(t)}}^{(t)}$ in order to achieve a tremendous computational complexity reduction. For ease of understanding, some important variables have been summarized in **Table I**.

Firstly, in the case when $i = 1$, three different situations need to be considered, $t = 1$, $2 \leq t \leq T_b$ and $T_b + 1 \leq t \leq T$, respectively. First of all, when $t = 1$, $\mathbf{W}_{\Omega_1^{(t)}}^{(t)} = \mathbf{0}$ because

of the initialization of $\left\{ \hat{\mathbf{x}}_{k,(1)}^T \right\}_{k=1}^K$. Hence, no computational cost incurs under this circumstance. Subsequently, when $2 \leq t \leq T_b$, since the candidate set $\Omega^{(t)}$ changes at each iteration due to the pruning mechanism, the relationship between $\mathbf{W}_{\Omega_1^{(t)}}^{(t)}$ and the previously estimated signals also varies dynamically. In this case, $\mathbf{W}_{\Omega_1^{(t)}}^{(t)}$ can only be acquired by

$$\mathbf{W}_{\Omega_1^{(t)}}^{(t)} = \sum_{l \in \Omega_n^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, \quad (14)$$

where $i+1 \leq n \leq K_a^{(t)}$. Finally, when $T_b + 1 \leq t \leq T$, the candidate set $\Omega^{(t)}$ remains unchanged. In this case, we replace $\Omega^{(t)}$ by Ω for notational simplicity. Based on (13), we can establish the following relationship between the $(t-1)$ -th iteration and the t -th iteration:

$$\mathbf{W}_{\Omega_1}^{(t)} = \sum_{l \in \Omega_n} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, \quad (15-1)$$

$$\mathbf{W}_{\Omega_{\bar{K}_a}}^{(t-1)} = \sum_{l \in \Omega_m} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T. \quad (15-2)$$

Combining (15-1) and (15-2), $\mathbf{W}_{\Omega_1}^{(t)}$ can be further ex-

pressed as

$$\mathbf{W}_{\Omega_1}^{(t)} = \mathbf{W}_{\Omega_{\bar{K}_a}}^{(t-1)} + \mathbf{g}_{\Omega_{\bar{K}_a}} \hat{\mathbf{x}}_{\Omega_{\bar{K}_a},(t-1)}^T - \mathbf{g}_{\Omega_1} \hat{\mathbf{x}}_{\Omega_1,(t-1)}^T. \quad (16)$$

It is obvious that $\mathbf{W}_{\Omega_{\bar{K}_a}}^{(t-1)}$ and $\mathbf{g}_{\Omega_1} \hat{\mathbf{x}}_{\Omega_1,(t-1)}^T$ can be reused since both of them have been computed at the $(t-1)$ -th iteration. Therefore, only the calculation of $\mathbf{g}_{\Omega_{\bar{K}_a}} \hat{\mathbf{x}}_{\Omega_{\bar{K}_a},(t-1)}^T$ is required. To sum up, $\mathbf{W}_{\Omega_1}^{(t)}$ can be summarized as

$$\mathbf{W}_{\Omega_1}^{(t)} = \begin{cases} \mathbf{0}, & t = 1, \\ \sum_{l \in \Omega_n^{(t)}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, & 2 \leq t \leq T_b, \\ \mathbf{W}_{\Omega_{\bar{K}_a}}^{(t-1)} + \mathbf{g}_{\Omega_{\bar{K}_a}} \hat{\mathbf{x}}_{\Omega_{\bar{K}_a},(t-1)}^T - \mathbf{g}_{\Omega_1} \hat{\mathbf{x}}_{\Omega_1,(t-1)}^T, & T_b + 1 \leq t \leq T. \end{cases} \quad (17)$$

Next, we discuss the case when $2 \leq i \leq K_a^{(t)} - 1$. Due to the constant of $\Omega^{(t)}$ at the arbitrary t -th iteration, we still use Ω to replace $\Omega^{(t)}$ for simplicity. Here, $\mathbf{W}_{\Omega_{i-1}}^{(t)}$ and $\mathbf{W}_{\Omega_i}^{(t)}$ can be expressed as

$$\mathbf{W}_{\Omega_{i-1}}^{(t)} = \sum_{l \in \Omega_{\tilde{m}}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T + \sum_{l \in \Omega_{\tilde{n}}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, \quad (18-1)$$

$$i \neq 2, 1 \leq \tilde{m} \leq i-2, i \leq \tilde{n} \leq K_a^{(t)}.$$

$$\mathbf{W}_{\Omega_i}^{(t)} = \sum_{l \in \Omega_m} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T + \sum_{l \in \Omega_n} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t-1)}^T, \quad (18-2)$$

$$1 \leq m \leq i-1, i+1 \leq n \leq K_a^{(t)}.$$

It is worth noting that if $i = 2$, $\mathbf{W}_{\Omega_1}^{(t)}$ should be computed by (17), rather than (18-1). According to (18), the relationship between the $(i-1)$ -th candidate and the i -th candidate in Ω can be formulated as

$$\mathbf{W}_{\Omega_i}^{(t)} = \mathbf{W}_{\Omega_{i-1}}^{(t)} + \mathbf{g}_{\Omega_{i-1}} \hat{\mathbf{x}}_{\Omega_{i-1},(t)}^T - \mathbf{g}_{\Omega_i} \hat{\mathbf{x}}_{\Omega_i,(t-1)}^T. \quad (19)$$

Following the same analysis as (16), only computing $\mathbf{g}_{\Omega_{i-1}} \hat{\mathbf{x}}_{\Omega_{i-1},(t)}^T$ incurs computational cost for obtaining $\mathbf{W}_{\Omega_i}^{(t)}$, since $\mathbf{W}_{\Omega_{i-1}}^{(t)}$ and $\mathbf{g}_{\Omega_i} \hat{\mathbf{x}}_{\Omega_i,(t-1)}^T$, as previous computing results, can be reutilized directly.

Finally, when $i = K_a^{(t)}$, we replace $K_a^{(t)}$ by K_a for notational simplicity. Then, the expression of $\mathbf{W}_{\Omega_{(K_a-1)}}^{(t)}$ and $\mathbf{W}_{\Omega_{K_a}}^{(t)}$ can be written as

$$\mathbf{W}_{\Omega_{(K_a-1)}}^{(t)} = \sum_{l \in \Omega_{\tilde{m}}} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T + \mathbf{g}_{\Omega_{K_a}} \hat{\mathbf{x}}_{\Omega_{K_a},(t-1)}^T, \quad (20-1)$$

$$1 \leq \tilde{m} \leq K_a - 2,$$

$$\mathbf{W}_{\Omega_{K_a}}^{(t)} = \sum_{l \in \Omega_m} \mathbf{g}_l \hat{\mathbf{x}}_{l,(t)}^T, \quad 1 \leq m \leq K_a - 1. \quad (20-2)$$

Combining (20-1) and (20-2), we obtain

$$\mathbf{W}_{\Omega_{K_a}}^{(t)} = \mathbf{W}_{\Omega_{(K_a-1)}}^{(t)} + \mathbf{g}_{\Omega_{K_a-1}} \hat{\mathbf{x}}_{\Omega_{K_a-1},(t)}^T - \mathbf{g}_{\Omega_{K_a}} \hat{\mathbf{x}}_{\Omega_{K_a},(t-1)}^T. \quad (21)$$

Similarly, to calculate $\mathbf{W}_{\Omega_{K_a}}^{(t)}$, we only need to compute $\mathbf{g}_{\Omega_{K_a-1}} \hat{\mathbf{x}}_{\Omega_{K_a-1},(t)}^T$ in (21). Additionally, it can be observed that (19) and (21) become identical when setting $i = K_a$ in (19). Therefore, these two situations can be combined into one.

The procedure of our proposed CR-EBCD algorithm is summarized in **Algorithm 2**. Following the above discussion, we can perform the data update with extremely low computational overhead. Furthermore, after iteration process, the method to obtain the estimated candidate set and transmitted data symbols is same as that of the original BCD algorithm (Step 3 and Step 4). It is worth noting that even though the majority of inactive user have been discarded in Step 2, the interference cannot be absolutely eliminated. Thus, LS method is still applied to estimate the data symbols for a more balanced MUD performance and computational complexity trade-off.

Compared with our proposed EBCD algorithm, the iteration process (Step 2) in **Algorithm 2** is refined, while other steps are the same. To be specific, if the lines 2-5 in **Algorithm 2** are removed, our proposed CR-EBCD algorithm will degenerate as the proposed EBCD algorithm. In essence, by constantly reusing the previous calculation results, our proposed CR-EBCD algorithm is able to avoid substantial repeating computations, contributing to dramatic computational cost saving without sacrificing any detection performance.

IV. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section, the computational complexity of our proposed BCD based algorithms for uplink grant-free NOMA systems is discussed. It is noted that the total number of complex multiplications is evaluated as a criterion to measure the computational complexity [11].

A. The Original BCD Algorithm

Firstly, for the original BCD algorithm, the complexity of calculating \mathbf{R}_k (line 2) is $NJ(K-1)$. Then, the complexity of estimating $\hat{\mathbf{x}}_k^T$ (line 3) is $NJ+N$. Thus, the total computational cost for completing the iteration process is $TK(KNJ+N)$. After that, the complexity for obtaining the estimated support which involves the operations of l_2 -norm and support merge is $2K$. In Step 4, the LS operation has a complexity of $J(2Ns^2+s^3)$, where s represents the estimated user sparsity level. Consequently, the entire computational complexity of the original BCD algorithm can be approximated as follows:

$$C_{\text{Original BCD}} = TK(KNJ+N) + 2K + J(2Ns^2+s^3). \quad (22)$$

B. Proposed EBCD Algorithm

With respect to our proposed EBCD algorithm, the computational complexity of updating signals at the t -th iteration is $(K_a^{(t)}NJ+N)$. When the number of iterations t is less than T_b , the operation of pruning the candidate set requires $2K_a^{(t)}$ at each iteration. Furthermore, when the iteration number exceeds T_b , the complexity analysis can refer to that in **Algorithm 1** due to the invariant candidate set. Finally, after iteration process, the complexity of Step 3 is $2\bar{K}_a$. For Step 4, the computational cost is the same as that of the original one (Algorithm 1), i.e., $J(2Ns^2+s^3)$. Hence, the total computational

complexity of our proposed enhanced BCD algorithm can be expressed as

$$C_{\text{EBCD}} = \underbrace{\sum_{t=1}^{T_b} \left(K_a^{(t)} \left(K_a^{(t)} NJ + N \right) + 2K_a^{(t)} \right)}_{t \leq T_b} + \underbrace{(T - T_b) \bar{K}_a \left(\bar{K}_a NJ + N \right)}_{T_b + 1 \leq t \leq T} + 2\bar{K}_a + J(2Ns^2 + s^3). \quad (23)$$

C. Proposed CR-EBCD Algorithm

We discuss the computational complexity of our proposed CR-EBCD algorithm in three distinct cases, $t = 1$, $2 \leq t \leq T_b$ and $T_b + 1 \leq t \leq T$, respectively.

Firstly, we consider the situation when $t = 1$. In this case, when estimating the first user's signals, calculating $\mathbf{W}_{\Omega_1^{(1)}}^{(1)}$ incurs no computational cost due to the initialization. Thus, the complexity of obtaining $\hat{\mathbf{x}}_{1,(1)}^T$ only comes from (9), which is $(NJ+N)$. Except for the first user, we update all the remaining users' signals sequentially based on (19) and (9), requiring $(2NJ+N)(K-1)$ in total. Furthermore, $2K$ complex multiplications are needed for pruning the candidate set. Thus, when $t = 1$, the computational complexity can be represented as

$$C_{\text{CR-EBCD}}^{(1)} = NJ + N + (2NJ + N)(K - 1) + 2K = 2NJK + NK - NJ + 2K. \quad (24)$$

As for the second case with $2 \leq t \leq T_b$, when $i = 1$, (14) is employed to compute $\mathbf{W}_{\Omega_1^{(t)}}^{(t)}$, which requires $(K_a^{(t)} - 1)NJ$ complex multiplications. while updating $\hat{\mathbf{x}}_{1,(t)}^T$ using (9), the complexity for acquiring $\hat{\mathbf{x}}_{1,(t)}^T$ involves $K_a^{(t)}NJ + N$ multiplications. Similar to the case when $t = 1$, the complexity for estimating all the remaining users' signals is $(K_a^{(t)} - 1)(2NJ + N)$. Meanwhile, at each iteration, extra $2K_a^{(t)}$ multiplications are needed for pruning the candidate set. As a result, the total computational complexity from $t = 2$ to $t = T_b$ is given by

$$C_{\text{CR-EBCD}}^{(2)} = \sum_{t=2}^{T_b} \left(\left(K_a^{(t)} NJ + N + (K_a^{(t)} - 1)(2NJ + N) \right) + 2K_a^{(t)} \right) = \sum_{t=2}^{T_b} \left(3K_a^{(t)} NJ + K_a^{(t)} N - 2NJ + 2K_a^{(t)} \right). \quad (25)$$

For the last case, we consider $T_b + 1 \leq t \leq T$. Since the candidate set remains unchanged during this period, (16) and (19) can be exploited to calculate $\mathbf{W}_{\Omega_1^{(t)}}^{(t)}$ and $\mathbf{W}_{\Omega_i^{(t)}}^{(t)}$, $i \neq 1$, respectively. Furthermore, the procedure of pruning the candidate set is discarded. Hence, the computational complexity in this case only stems from lines 6-7 in **Algorithm 2**, which can be expressed by

$$C_{\text{CR-EBCD}}^{(3)} = (2NJ + N) \bar{K}_a (T - T_b). \quad (26)$$

As a consequence, the entire computational complexity of our proposed CR-EBCD algorithm can be computed according to (27).

To show the complexity comparison more explicitly, the computational complexities of three BCD based algorithms have been summarized into **Table II**.

$$\begin{aligned}
C_{\text{CR-EBCD}} &= C_{\text{CR-EBCD}}^{(1)} + C_{\text{CR-EBCD}}^{(2)} + C_{\text{CR-EBCD}}^{(3)} + 2\bar{K}_a + J(2Ns^2 + s^3) \\
&= \underbrace{(2NJK + NK - NJ + 2K)}_{t=1} + \underbrace{\sum_{t=2}^{T_b} (3K_a^{(t)}NJ + K_a^{(t)}N - 2NJ + 2K_a^{(t)})}_{2 \leq t \leq T_b} \\
&\quad + \underbrace{(T - T_b)(2NJ + N)\bar{K}_a}_{T_b+1 \leq t \leq T} + 2\bar{K}_a + J(2Ns^2 + s^3).
\end{aligned} \tag{27}$$

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON AMONG THREE BCD BASED ALGORITHMS

Algorithm	Number of complex multiplications
Original BCD	$TK(KNJ + N) + 2K + J(2Ns^2 + s^3)$
Proposed EBCD	$\sum_{t=1}^{T_b} (K_a^{(t)}(K_a^{(t)}NJ + N) + 2K_a^{(t)})$ $+ (T - T_b)\bar{K}_a(\bar{K}_a NJ + N) + 2\bar{K}_a + J(2Ns^2 + s^3)$
Proposed CR-EBCD	$\sum_{t=2}^{T_b} (3K_a^{(t)}NJ + K_a^{(t)}N - 2NJ + 2K_a^{(t)})$ $+ (T - T_b)\bar{K}_a(2NJ + N) + 2\bar{K}_a + J(2Ns^2 + s^3)$

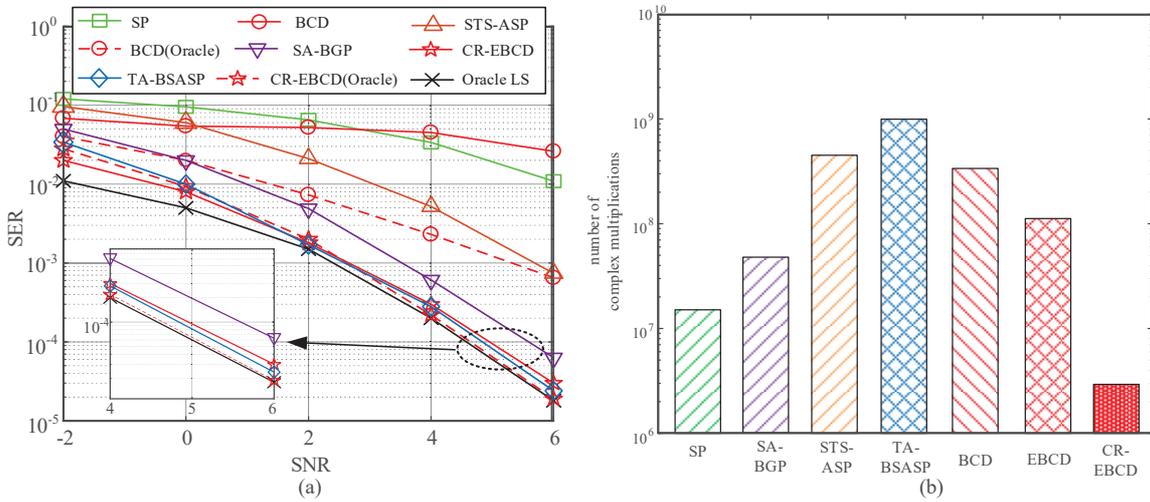


Fig. 2. (a) SER performance comparison of different algorithms against SNR; (b) Computational complexity comparison of different algorithms.

V. NUMERICAL SIMULATIONS

In this section, we evaluate the detection performances of our proposed BCD-based algorithms in uplink grant-free NOMA systems. Since our proposed EBCD algorithm and CR-EBCD algorithm achieve the same SER performance, we only show the performance of CR-EBCD algorithm in the following simulations. Moreover, to better validate the performance of our proposed BCD-based algorithms, oracle BCD algorithm and oracle CR-EBCD algorithm are adopted, both of which know the exact user sparsity before MUD.

For performance benchmarking, we consider the classical SP [17], the spatial-temporal structure enhanced adaptive SP (STS-ASP) [13], the TA-BSASP [25] and the sparsity adaptive block gradient pursuit (SA-BGP) [33] as baseline algorithms. In particular, it is assumed that the classical SP knows the exact user sparsity in advance while other existing algorithms can identify the active users adaptively. Additionally, as an upper bound of the MUD performance, the oracle LS method is employed, where the perfect knowledge of the active users is assumed at the BS.

A. Parameters Setting

In our simulations, the spreading sequences for each user are generated from random Gaussian sequences with the length $N = 100$, which also equals to the number of subcarriers. The total potential users is set as $K = 200$ and the actual user sparsity level varies from 18 to 20 randomly in each frame. Each frame consists of $J = 7$ continuous symbol durations according to the LTE-Advanced standard [39]. Furthermore, QPSK modulation is adopted. The channel coefficients in all simulations follows $h_{n,k} \sim \mathcal{CN}(0, 1)$.

For all BCD-based algorithms, the equivalent tuning factor λ_0 is set as 0.7 and the maximum number of iterations is equal to 12. In addition, we set the threshold V_{th} to 1.31, 1.05, 0.97, 0.86, and 0.73 when signal-to-noise ratio (SNR) equals to -2 dB, 0 dB, 2 dB, 4 dB and 6 dB, respectively.

Finally, in Figs. 2, 5, 6 and 7, the parameter $\psi^{(t)}$ for pruning the candidate set in the proposed CR-EBCD algorithms is set as $\psi^{(1)} = \psi^{(2)} = \dots = \psi^{(T_b)} = 20$ and $T_b = 8$.

B. Simulation results

Fig. 2(a) depicts the SER performance comparison of different algorithms versus SNR. From Fig. 2(a), it is evident that our proposed CR-EBCD algorithm significantly outperforms the classical SP, due to the exploitation of the structured sparsity. In addition, our proposed CR-EBCD algorithm achieves better CS-MUD performance compared with STS-ASP and SA-BGP and almost the same performance as TA-BSASP, thanks to the property of global optimization and the well-designed pruning mechanism. To be specific, since the proposed CR-EBCD algorithm goes through all the potential users' signals in the iteration process, the missing-detection and false-alarm are more likely to be avoided. Meanwhile, since a majority of interference from the inactive users when computing \mathbf{R}_k is avoided with the aid of the pruning mechanism, the performance of our proposed CR-EBCD algorithm gets improved remarkably compared with that of the original BCD. It can be also observed that our proposed oracle CR-EBCD can approach the oracle LS performance. However, when it acquires the user sparsity level adaptively, it suffers from near 0.5 dB SER performance degradation at SNR of 6 dB. The reason is that the interference from inactive users cannot be completely removed due to lack of the accurate information about the actual number of active users. In this case, the stopping criterion sometimes loses its effectiveness, because the active users are hardly distinguishable from the inactive ones due to such interferences.

The comparison of the computational complexity among different schemes is shown in Fig. 2(b). It is apparent that all the BCD-based algorithms have lower computational cost than that of TA-BSASP and STS-ASP, since the computation of matrix inversion has been skipped when updating signals. Particularly, our proposed CR-EBCD algorithm saves more than 99.7% computational cost compared with that of TA-BSASP. Meanwhile, it enjoys significantly lower computational overhead compared with that of SA-BGP, whose matrix inversion operations are replaced by block gradient descend. The reason is that our proposed CR-EBCD algorithm requires neither sophisticated matrix inversions nor large-scale matrix multiplications. Furthermore, among all the BCD-based algorithms, our proposed EBCD algorithm saves about one-third of the complex multiplications compared with that of the original BCD, thanks to the operation of pruning the candidate set. More importantly, our proposed CR-EBCD algorithm enjoys about two orders of magnitude lower computational complexity than the proposed EBCD algorithm by removing redundant calculations. Therefore, our proposed CR-EBCD algorithm strikes a good balance between computational complexity and SER performance. Moreover, the ultra-low computational cost can lend itself for efficient implementation.

The influence of the final candidate set size \overline{K}_a in our proposed CR-EBCD algorithm is demonstrated in Fig. 3. Apparently, a better SER performance can be achieved while the value of \overline{K}_a becomes smaller. The reason is that when more inactive users are discarded from the candidate set and their corresponding signals are set to zeros, the calculation results of \mathbf{R}_k can be more accurate. In this case, when updating signals for active users, the impact of interferences from the inactive users can be largely alleviated. In contrast, severe interferences give rise to the threshold failure and further lead to the SER performance loss. Moreover, with respect to the

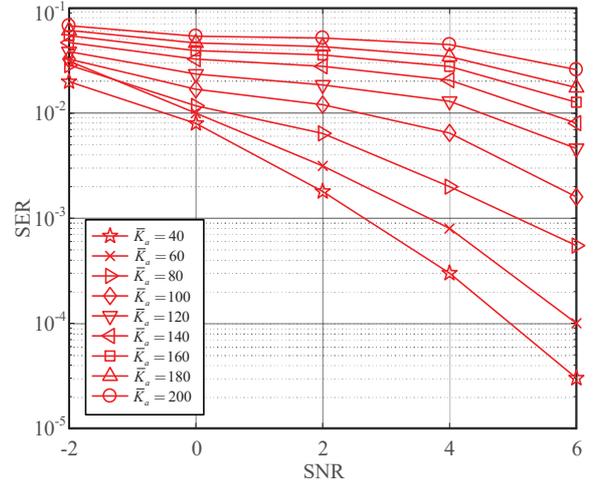


Fig. 3. SER performance comparison employing our proposed CR-EBCD algorithm against different \overline{K}_a .

actual deployment, the final candidate set \overline{K}_a is set around $0.2K$ considering the maximum number of potential active users in real communications [38]. By doing so, a majority of inter-user interference can be eliminated and the missing detection can also be avoided to a great extent. Besides, it is worth mentioning that when $\overline{K}_a = 200$, our proposed CR-EBCD algorithm equivalently degenerates to the original BCD algorithm.

Fig. 4 shows the SER performance comparison against different T_b in our proposed CR-EBCD algorithm. In this simulation, the final candidate set size \overline{K}_a is equal to 40 and we assume $\psi^{(1)} = \psi^{(2)} = \dots = \psi^{(T_b)}$. Correspondingly, when $T_b = 2, 4$ and 8 , $\psi^{(t)}$ for the above three cases is equal to $80, 40$ and 20 , respectively. From Fig. 4, it is clear that the SER performance of our proposed CR-EBCD algorithm gradually improves with decreasing T_b . This is because pruning the candidate set excessively in the beginning few iterations causes some actual active users discarded, resulting in the performance degradation. Meanwhile, if T_b is too small, more iterations are needed which bring about higher computational complexity. Hence, the selection of T_b should consider the trade-off between the speed of pruning candidate set and computational complexity. From Fig. 4, it can be seen that discarding $0.1K$ users from the candidate set in each iteration is appropriate in the practical applications.

Fig. 5 investigates the SER performance comparison of the considered algorithms against the number of consecutive time slots J in a certain frame, where $\text{SNR} = 4$ dB. As shown by the figure, the SER performances of the SP and oracle LS algorithm remain unchanged regardless of J , since both of them neglect the frame-wise joint sparsity structure. Conversely, as J increases, the original BCD, TA-BSASP and our proposed CR-EBCD algorithm can obtain SER performance gain, owing to the utilization of the structured sparsity. Moreover, our proposed CR-EBCD algorithm can achieve almost the same SER performance of the oracle LS method when $J \geq 7$.

Fig. 6 illustrates the impact of the user sparsity level on the SER performance of the considered algorithms when $\text{SNR} = 4$ dB. Simulation results indicate that as the user sparsity level rises, all the considered algorithms suffer from the SER performance degradation. Additionally, our proposed CR-EBCD algorithm outperforms the TA-BSASP algorithm

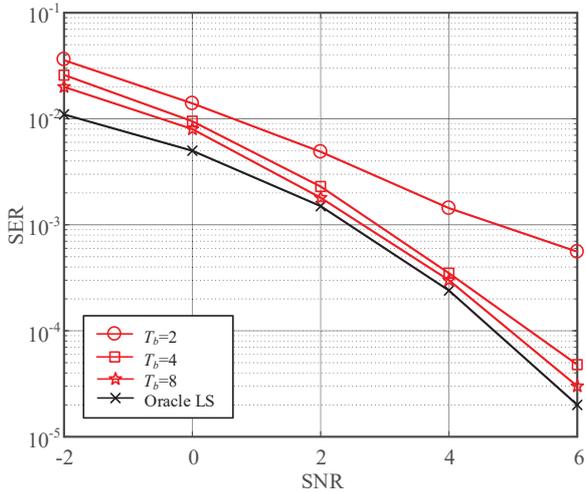


Fig. 4. SER performance comparison employing our proposed CR-EBCD algorithm against different T_b when the final candidate set size \bar{K}_a is equal to 40.

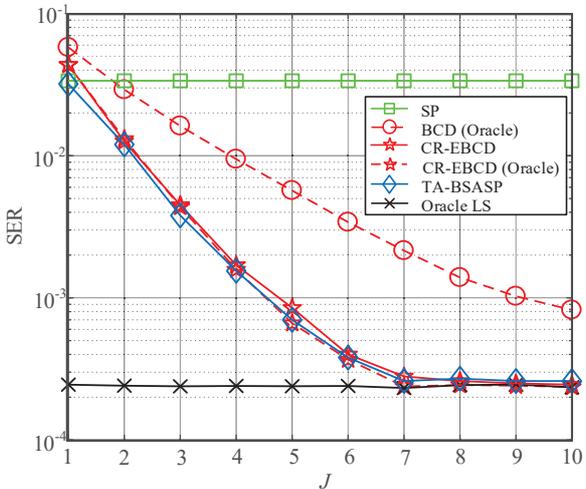


Fig. 5. SER performance comparison against the number of time slots in a whole frame when SNR = 4 dB.

when the user sparsity level is in a high region. Note that the computational cost of our proposed CR-EBCD algorithm remains the same regardless of the user sparsity, while the computational complexity of the TA-BSASP algorithm increases dramatically with the increasing actual user sparsity. Consequently, our proposed CR-EBCD algorithm is not only more tolerable to the increasing user sparsity, but also enjoys an extremely lower computational overhead compared with the TA-BSASP algorithm.

In the above simulations, similar to some existing works [24]-[29], it is assumed that the channel information is perfectly known at the receiver side and keeps static during a data frame. However, in the practical scenarios, small channel fluctuations might occur during an entire frame, which leads to channel estimation errors. Under this circumstance, we consider a more practical scenario, where the channel for each active user varies slowly and the initial channel information $\mathbf{h}_k^{(1)}$ obtained by channel estimation will be used for the whole frame [12]. Thus, channel information used in the detection is no longer perfect. In this scenario, for active user k , the real channel gain in the j -th time slot can be expressed as

$$\tilde{\mathbf{h}}_k^{(j)} = (1 - \beta) \mathbf{h}_k^{(j-1)} + \beta \Delta \mathbf{h}_k^{(j)}, \quad (28)$$

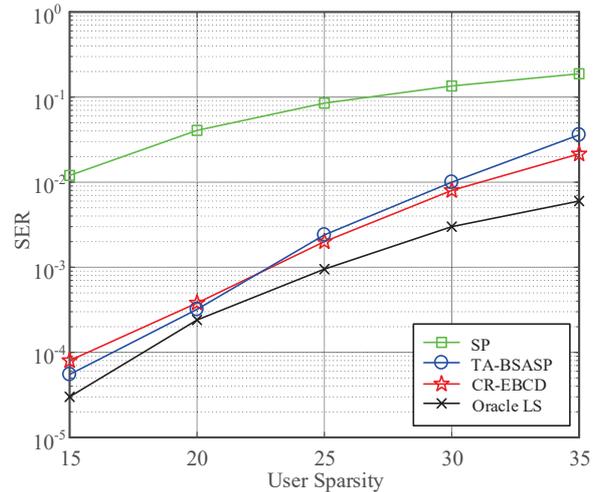


Fig. 6. SER performance comparison against the active user sparsity when SNR = 4 dB.

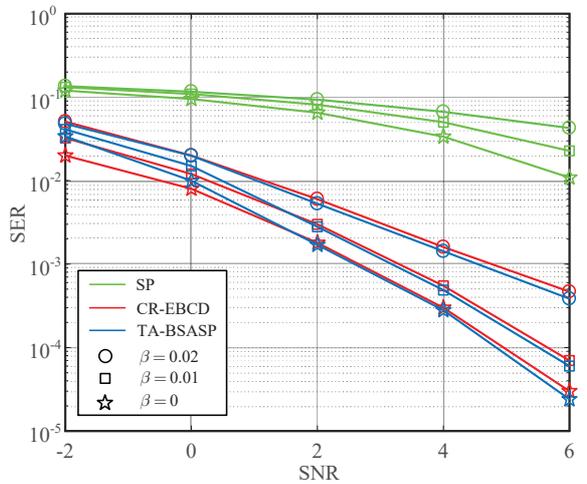


Fig. 7. SER performance comparison considering time-varying channel.

where $\beta \ll 1$ indicates the channel variation degree. Moreover, $\Delta \mathbf{h}_k^{(j)}$ represents the actual channel fluctuation of user k in the j -th time slot and each element $\Delta \mathbf{h}_{n,k}^{(j)}$ follows $\mathcal{CN}(0, 1)$. From Fig. 7, we can observe that all the considered algorithms experience SER performance loss in the presence of channel estimation errors ($\beta = 0$ indicates time-invariant channel and perfect channel estimation). In particular, the faster the channel varies, the more the SER performance degrades. Despite of this, when $\beta = 0.02$, the SER performance of our proposed CR-EBCD algorithm can still reach 4.6×10^{-3} when SNR = 6 dB.

VI. CONCLUSION

In this paper, we have proposed two modified BCD algorithms to carry out MUD in the uplink grant-free NOMA systems. Firstly, compared with the original BCD algorithm, a novel candidate set pruning mechanism has been applied to the proposed EBCD algorithm, which significantly alleviates the interference from the inactive users. Moreover, derived from the proposed EBCD algorithm, our proposed CR-EBCD algorithm achieves substantial computational complexity reduction without sacrificing any SER performance, attributing to the elimination of all the redundant calculations in the iteration process. Then, we have carried out the computational

complexity analysis of two proposed algorithms. Finally, simulation results have demonstrated that our proposed EBCD and CR-EBCD algorithms can achieve near oracle LS performance without the priori knowledge of user sparsity level. Additionally, our proposed CR-EBCD algorithm also enjoys a two orders of magnitude complexity reduction compared with the proposed EBCD algorithm, making it a feasible candidate for the practical implementation.

As a future work, it is interesting to study the case of asynchronous uplink grant-free NOMA systems. Alternatively, another promising future work is to design a multi-user detector, which enables the BS to execute the multi-user detection and synchronization jointly. In addition, the integration of grant-free NOMA systems and deep neural network (DNN) can be applied to achieve end-to-end communications for further performance enhancement [40]-[42].

REFERENCES

- [1] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanović, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59-65, Sep. 2016.
- [2] C. Bockelmann *et al.*, "Towards massive connectivity support for scalable mMTC communications in 5G networks," *IEEE Access*, vol. 6, pp. 28969-28992, 2018.
- [3] 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 Tuts.*, vol. 17, no. 4, pp. 2347-2376, 4th Quart. 2015.
- [4] H. Shariatmadari *et al.*, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10-17, Sep. 2015.
- [5] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications" *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120-128, Feb. 2017.
- [6] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen and L. Hanzo., "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys and Tuts.*, vol. 20, no. 3, pp. 2294-2323, 3rd Quart., 2018.
- [7] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. A. Dahir and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615-636, Mar. 2021.
- [8] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181-2195, Dec. 2017.
- [9] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanović, and E. D. Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88-99, Sep. 2018.
- [10] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: a survey," *IEEE Commun. Surveys and Tuts.*, vol. 22, no. 3, pp. 1805-1838, 3rd Quart., 2020.
- [11] Y. Du *et al.*, "Efficient multi-user detection for uplink grant-free NOMA: prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812-2828, Dec. 2017.
- [12] Y. Cui, W. Xu, Y. Wang, J. Lin and L. Lu, "Side-information aided compressed multi-user detection for up-link grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7720-7731, Aug. 2020.
- [13] L. Wu, P. Sun, Z. Wang and Y. Yang, "Joint user activity identification and channel estimation for grant-free NOMA: a spatial-temporal structure enhanced approach," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12339-12349, Aug. 2021.
- [14] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
- [15] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.
- [16] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
- [17] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230-2249, May. 2009.
- [18] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: useful tips and tricks," *IEEE Commun. Surveys and Tuts.*, vol. 19, no. 3, pp. 1527-1550, 3rd Quart., 2017.
- [19] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169-6183, Dec. 2015.
- [20] D. Cohen and Y. C. Eldar, "Sub-Nyquist sampling for power spectrum sensing in cognitive radios: a unified approach," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3897-3910, Aug. 2014.
- [21] H. Zhu and S. Brueck, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 454-465, Feb. 2011.
- [22] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972-974, Jul. 2012.
- [23] C. Bockelmann, H. F. Schepker, and A. Dekorsy, "Compressive sensing based multi-user detection for machine-to-machine communication," *Eur. Trans. Telecommun.*, vol. 24, no. 4, pp. 389-400, Jul. 2013.
- [24] A. T. Abebe and C. G. Kang, "Iterative order recursive least square estimation for exploiting frame-wise sparsity in compressive sensing-based MTC," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1018-1021, May 2016.
- [25] Y. Du, C. Cheng, B. Dong, Z. Chen, X. Wang, J. Fang, and S. Li, "Block-sparsity-based multiuser detection for uplink grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7894-7909, Dec. 2018.
- [26] S. Li, L. Xiao and T. Jiang, "An efficient matching pursuit based compressed sensing detector for uplink grant-free NOMA," *IEEE Trans. Vehi. Tech.*, vol. 70, no. 2, pp. 2012-2017, Feb. 2021.
- [27] C. Wei, H. Liu, Z. Zhang, J. Dang, and L. Wu, "Approximate message passing based joint user activity and data detection for NOMA," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 640-643, Mar. 2017.
- [28] B. Jeong, B. Shim and K. Lee, "MAP-based active user and data detection for massive machine-type communications," *IEEE Trans. Vehi. Tech.*, vol. 67, no. 9, pp. 8481-8494, Sep. 2018.
- [29] Y. Mei, Z. Gao, D. Mi, P. Xiao, M. Alouini, "Compressive sensing based grant-free random access for massive MTC," *Proc. IEEE Int. Conf. on UK-China Emerging Technologies (UCET)*, Glasgow, United Kingdom, Aug. 2020, pp. 1-4.
- [30] Q. He, T. Q. S. Quek, Z. Chen, Q. Zhang and S. Li, "Compressive channel estimation and multi-user detection in C-RAN with low-complexity methods," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3931-3944, Jun. 2018.
- [31] Y. Li, M. Xia and Y. Wu, "Activity detection for massive connectivity under frequency offsets via first-order algorithms," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1988-2002, Mar. 2019.
- [32] F. Jiang, G. Zheng, Y. Hu, Y. Wang and Y. Xu, "Low-complexity multi-user detection based on gradient information for uplink grant-free NOMA," *IEEE Access*, vol. 8, pp. 137438-137448, Jul. 2020.
- [33] P. Gao, Z. Liu, P. Xiao, C. Foh and J. Zhang, "Low-complexity channel estimation and multi-user detection for uplink grant-free NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 263-267, Feb. 2022.
- [34] N. Y. Yu, "Binary Golay spreading sequences and Reed-Muller codes for uplink grant-free NOMA," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 276-290, Oct. 2020.
- [35] Y. Mei, Z. Gao, Y. Wu, W. Chen, J. Zhang, D. W. K. Ng, and M. Di Renzo, "Compressive sensing based joint activity and data detection for grant-free massive IoT access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1851-1869, Mar. 2022.
- [36] A. Mousavi, M. Rezaee, and R. Ayanzadeh, "A survey on compressive sensing: classical results and recent advancements," Aug. 2019, *arXiv: 1908.01014*. [Online]
- [37] J. P. Hong, W. Choi, and B. Rao, "Sparsity controlled random multiple access with compressed sensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 998-1010, Feb. 2015.
- [38] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 2320-2323, Nov. 2016.
- [39] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 12)*, document 3GPP TS-36.211, Jan. 2016.
- [40] N. Ye, X. Li, H. Yu, A. Wang, W. Liu, and X. Hou, "Deep learning aided grant-free NOMA toward reliable low-latency access in tactile Internet of Things," *IEEE Trans. Indust. Inf.*, vol. 15, no. 5, pp. 2995-3005, May. 2019.
- [41] H. Yu, Z. Fei, Z. Zheng, and N. Ye, "Finite-alphabet signature design for grant-free NOMA: a quantized deep learning approach," *IEEE Trans. Vehi. Tech.*, vol. 69, no. 10, pp. 10975-10987, Jul. 2020.

- [42] N. Ye, J. An, and J. Yu, "Deep-learning-enhanced NOMA transceiver design for massive MTC: challenges, state of the art, and future directions," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 66-73, Aug. 2021.