

Cluster-Based Massive Access for Massive MIMO Systems

Shiyu Liang^{1,2,3}, Wei Chen^{1,2,*}, Zhongwen Sun^{2,4}, Ao Chen^{2,5}, Bo Ai^{2,6}

¹ State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100091, China

² School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100091, China

³ Frontiers Science Center for Smart High-speed Railway System, Beijing Jiaotong University, Beijing 100091, China

⁴ Key Laboratory of Railway Industry of Broadband Mobile Information Communications, Beijing Jiaotong University, Beijing 100091, China

⁵ Beijing Engineering Research Center of High-speed Railway Broadband Mobile Communications, Beijing Jiaotong University, Beijing 100091, China

⁶ School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

* The corresponding author, email: weich@bjtu.edu.cn

Cite as: S. Liang, W. Chen, *et al.*, "Cluster-based massive access for massive mimo systems," *China Communications*, 2024, vol. 21, no. 1, pp. 24-33. **DOI:** 10.23919/JCC.fa.2023-0380.202401

Abstract: Massive machine type communication aims to support the connection of massive devices, which is still an important scenario in 6G. In this paper, a novel cluster-based massive access method is proposed for massive multiple input multiple output systems. By exploiting the angular domain characteristics, devices are separated into multiple clusters with a learned cluster-specific dictionary, which enhances the identification of active devices. For detected active devices whose data recovery fails, power domain nonorthogonal multiple access with successive interference cancellation is employed to recover their data via re-transmission. Simulation results show that the proposed scheme and algorithm achieve improved performance on active user detection and data recovery.

Keywords: compressive sensing; dictionary learning; multiuser detection; random access

I. INTRODUCTION

With the commercialization of 5G, research on 6G has gained significant attention. Compared to the performance indicator of 1 million connections per

square kilometer in 5G, 6G claims an ultra-high density performance indicator with a device connection density of over 100 per cubic meter, aiming to achieve nearly seamless global connectivity for air, sky, earth, and sea [1, 2]. Massive machine-type communication (mMTC) aims to support sporadic transmission of small data packets by machine-type devices (MTDs), and is a key part of realizing ubiquitous communication [3]. However, in the traditional grant-based random access (RA) scheme, MTDs need the grant from the base station (BS) before sending data, which includes several interactions and causes significant signaling overhead. Compared to the small data packets of MTDs, such signaling overhead is intolerable and causes a waste of resources. In addition, as many MTDs are battery-operated with limited power, multiple interactions in handshake reduce their lifetime.

In order to reduce the high control signaling overhead between the access phase and data transmission phase, grant-free RA (GFRA) schemes have attracted broad attention in recent years [4–6]. In GFRA, MTDs directly send the data along with the pilot instead of waiting for a response/grant from the BS. Due to the sporadic activity of MTDs in the mMTC system, compressive sensing-based multiuser detection (CS-MUD) has been proposed for detecting active users

Received: Jun. 19, 2023

Revised: Sep. 13, 2023

Editor: Lei Wang

and recovering their data [7–9]. Specifically, CS-MUD consists of two phases. In the first phase, the BS performs active user detection (AUD) and channel estimation based on the pilot sent by the user. In the second phase, the BS utilizes the estimated channel to recover the data. Many CS-MUD algorithms have been proposed to improve the performance of AUD and data recovery [10–14]. An enhanced block coordinate descend (BCD) algorithm is proposed in [10], which incorporates a novel candidate set pruning mechanism into the original BCD framework and leads to remarkable CS-MUD performance improvement. The use of sinusoidal spreading sequences is explored in [11]. Compared with the random or Zadoff-Chu spreading sequences, using sinusoidal spreading sequences allows some non-iterative algorithms to be utilized in CS-MUD. Besides, various kinds of prior information have been applied to enhance CS-MUD. For example, the activity pattern of adjacent time slots can be exploited in the CS-MUD algorithm design. Dynamic CS-MUD (DCS-MUD) in [12] takes the information of the former time slot as prior information. Then, the bidirectional DCS-MUD (BD-DCS-MUD) proposed in [13] further exploits the prior information of both former and later time slots. The information of modulation and cyclic redundancy check is exploited in CS-MUD to improve the accuracy of pilot detection and data recovery [14].

Furthermore, it has been proved that increasing the number of antennas at the BS enhances CS-MUD [15–18]. The existing CS-MUD methods for massive multiple input multiple output (mMIMO) systems mostly treat the CS-MUD as a row-sparse recovery problem [14, 15, 19], which leads to improved recovery performance in comparison to the conventional element-sparse recovery problem in the single antenna case. Owing to the limited scattering in the environment, the channel of MTDs is also sparse in the angular domain. In specific, each MTD only has a limited number of paths, which leads to the power spectrum density spanning a small range in the angular domain. By exploiting the angular domain channel characteristics, a novel structural based model for massive random access in mMIMO systems is proposed to enhance the performance of AUD and channel estimation in CS-MUD [16]. However, if two MTDs with the same pilot sequence are close to each other, it is still unable to differentiate them in the angular domain and this kind

of collision is difficult to resolve.

In this paper, we propose a new massive access method that exploits the angular domain channel characteristics in mMIMO systems to enhance the performance of CS-MUD. Different from the approach in [16], we propose to separate MTDs into several clusters. MTDs within the same cluster are difficult to differentiate in the channel spatial or angular domain, and are assigned with different pilots to avoid collisions. MTDs in distinct clusters can be differentiated in the channel spatial or angular domain, so pilots are reused across different clusters. To accurately differentiate MTDs from different clusters, we propose to learn a cluster-specific dictionary. Under the learned dictionary, the spatial channel has more prominent block sparse characteristics. When an MTD is active, it directly sends the pre-assigned pilot and data. The BS performs AUD, channel estimation and data recovery using the learned dictionary and the proposed corresponding algorithm. For detected active MTDs whose data recovery fails, as the BS already knows their activity, we exploit power domain nonorthogonal multiple access (PD-NOMA) with successive interference cancellation (SIC) to recover their data via retransmission. Numerical results demonstrate the superiority of the proposed method against conventional methods.

The rest of the paper is organized as follows: Section II describes the system model and the proposed cluster-based CS access scheme. Section III provides the receiver algorithm designed to conduct cluster-based dictionary learning, AUD, channel estimation, and PD-NOMA. Numerical results are presented in Section IV, followed by conclusions in Section V.

II. SYSTEM MODEL AND THE PROPOSED CLUSTER-BASED CS ACCESS SCHEME

We consider a typical mMTC scenario containing a BS equipped with M antennas and N stationary single-antenna MTDs. N_{act} devices are active in a coherence time, i.e., have messages to transmit, and perfect synchronization is assumed in this paper. The spatial MIMO channel described in [20] is considered, which is affected by the geometry characteristics of the cell scattering. For the active MTD k , the spatial MIMO

channel is given as

$$\mathbf{h}_k = \sum_{i=1}^{N_c} \sum_{l=1}^{N_s} \beta_{i,j,k} \boldsymbol{\alpha}(\theta_{i,j,k}), \quad (1)$$

where N_c and N_s denote the number of scattering clusters and the number of propagation subpaths, respectively. Scattering clusters include N_u primary clusters and N_f far scattering clusters (FSCs), and thus $N_c = N_u + N_f$. $\beta_{i,j,k}$ is the gain of the l -th subpath of the i -th scattering channel cluster of the k -th MTD. $\boldsymbol{\alpha}(\theta_{i,j,k}) = \left[1, e^{j2\pi \frac{d}{\lambda} \sin(\theta_{i,l,k})}, \dots, e^{j2\pi \frac{d}{\lambda} (M-1) \sin(\theta_{i,l,k})} \right]^T$ is the array response vector, where $\theta_{i,l,k}$ is the angle of arrival (AOA) of the l -th path of the MTD k , and the antenna spacing d is set as half-wavelength, i.e., $d = \frac{\lambda}{2}$. The limited scattering clusters restricts the AOAs, and thus the channel becomes sparse in the angular domain.

To support more devices with limited resources, we propose a cluster-based alternating access scheme, which is illustrated in Figure 1. Based on the angular information of MTDs, we separate stationary MTDs into $2C$ clusters, where C is the number of cluster allowed to access in one frame, so half of the clusters are served in one frame. To avoid the influence caused by angle spread between adjacent clusters, we allow non-adjacent clusters to access in one frame. In specific, the BS emits beams of different angles in different frames, and each MTD's cluster index is determined by the beam index, where it has the strongest receive signal strength. Besides, all clusters share the same pilot pool, and MTDs within the same cluster are pre-assigned with different pilots. Actually, MTDs can randomly select pilots to send, but intra-cluster collision needs to be resolved. Here, we consider the scenario of pre-assigned pilots for simplicity. As pilots are reused in different clusters, the number of needed pilots in the proposed scheme is $\frac{1}{2C}$ of that in traditional non-cluster based scheme, which can reduce inter-pilot interference and improve the performance. Although collision may happen between non-adjacent clusters, collided MTDs could be differentiated by their AOAs, i.e., the sparse structure of channel representation in the angular domain. Without loss of generality, we consider the AUD and data recovery problem for one frame in the sequel.

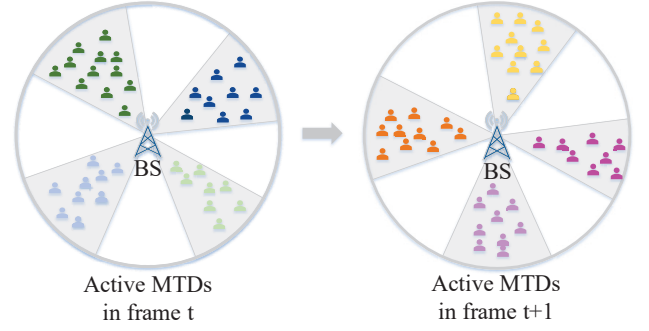


Figure 1. Cluster-based alternating access scheme.

Let $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{N_p}] \in \mathbb{R}^{L_p \times N_p}$ denote the pilot pool, where L_p and N_p are the length and the number of pilots, respectively. Elements in the pilot can be generated randomly following $\mathcal{CN}(0, \frac{1}{L_p})$. Corresponding to each pilot, there is a unique spreading sequence for data transmission, which is denoted as $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_p}] \in \mathbb{R}^{L_s \times N_p}$, whose elements can be generated randomly following $\mathcal{CN}(0, \frac{1}{L_s})$. Each active MTD directly transmits L_p pilot symbols and multiple data symbols to the BS. The BS first detects transmitted pilots and estimates the channel, then performs data recovery with the estimated channel.

Define $\mathcal{K}$ as the index set of active MTDs. $\mathbf{p}_k$ and $\mathbf{h}_k$ denote the pilot and channel of the MTD $k \in \mathcal{K}$, respectively. Then, the received pilot $\mathbf{Y}_p \in \mathbb{R}^{L_p \times M}$ at the BS is

$$\begin{aligned} \mathbf{Y}_p &= \sum_{k \in \mathcal{K}} \mathbf{p}_k \mathbf{h}_k^T + \mathbf{N}_p \\ &= \mathbf{P} \mathbf{H} + \mathbf{N}_p, \end{aligned} \quad (2)$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{N_p}]^T \in \mathbb{C}^{N_p \times M}$ and $\mathbf{N}_p \in \mathbb{C}^{L_p \times M}$ denote the channel matrix and the white Gaussian noise matrix, respectively. $\mathbf{h}_p$ is the superposition channel of MTDs transmitting the p -th pilot, and we denote $\mathcal{P}_p$ as the set of active MTDs that transmit the p -th pilot. $\mathbf{H}$ is a row-sparse matrix with non-zeros rows indicating transmitted pilots.

Denote $\mathbf{d}_k \in \mathbb{C}^{L_d}$ and $\mathbf{s}_k \in \mathbb{C}^{L_s}$ as the transmitted symbols and the spreading sequence of the k -th MTD, respectively. Then the received data of the m -th antenna at the BS is given as

$$\mathbf{Y}_d^m = \sum_{k \in \mathcal{K}} h_k^m \mathbf{s}_k \mathbf{d}_k^T + \mathbf{N}_d^m, \quad (3)$$

where h_k^m and $\mathbf{N}_d^m \in \mathbb{C}^{L_s \times L_d}$ are the channel coefficient of MTD k at the m -th antenna and the additive white Gaussian noise, respectively. With some simple manipulations, the received data $\mathbf{Y}_d = [\mathbf{Y}_d^1, \mathbf{Y}_d^2, \dots, \mathbf{Y}_d^M] \in \mathbb{C}^{L_s \times ML_d}$ at all antennas can be written as

$$\mathbf{Y}_d = \mathbf{S}\mathbf{X} + \mathbf{N}_d, \quad (4)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_p}]^T \in \mathbb{C}^{N_p \times ML_d}$, $\mathbf{x}_p = [\mathbf{x}_p^1, \dots, \mathbf{x}_p^M] \in \mathbb{C}^{ML_d}$, $\mathbf{x}_p^m = \sum_{k \in \mathcal{P}_p} h_k^m \mathbf{d}_k \in \mathbb{C}^{L_d}$.

III. RECEIVER ALGORITHM DESIGN

The mMIMO channel matrix $\mathbf{H}$ in Eq. (2) is a row-sparse matrix owing to the sporadic activity, and can be estimated by sparse recover algorithms, e.g., OMP-MMV [21]. The index of non-zero rows in the recovered channel matrix $\hat{\mathbf{H}} \in \mathbb{C}^{N_p \times M}$ indicates the set of detected active pilots $\hat{\mathcal{M}}$. As the data matrix $\mathbf{X}$ in Eq. (4) and the channel matrix $\mathbf{H}$ in Eq. (2) have the same support, the recovered data $\hat{\mathbf{X}}$ can be obtained via the least square (LS) method

$$\hat{\mathbf{X}}(\mathcal{M}, :) = (\mathbf{S}(:, \mathcal{M})^T \mathbf{S}(:, \mathcal{M}))^{-1} \mathbf{S}(:, \mathcal{M})^T \mathbf{Y}_d. \quad (5)$$

Given the estimated channel, the transmitted data $\mathbf{d}_p$ corresponding to the p -th pilot can be recovered by

$$\hat{\mathbf{d}}_p = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{x}}_p^i / \hat{h}_p^i. \quad (6)$$

If the recovered data passes the cyclic redundancy check (CRC), it suggests the AUD and channel estimation are successful for the MTD transmitting the p -th pilot. If MTDs in different clusters select the same pilot, the recovered channel is the superposition of all these MTD's channel. In this case, the recovered data is not likely to pass the CRC.

To differentiate active MTDs of different clusters who transmitting the same pilot, we propose a dictionary learning assisted AUD method. Dictionary learning is a highly effective approach for extracting crucial features of the signals from a training dataset, particularly when the signals have sparse representations. Previous researches have proved its effectiveness when applied to classification tasks [22–24]. Concretely, we could learn a cluster-specific dictionary, where the channel of MTDs in different clusters

could be represented by different sub-dictionaries, and its representation possesses more prominent block sparse characteristics. To improve the accuracy of classification, it is essential to minimize similarities between the atoms of different sub-dictionaries, which makes it easier to distinguish MTDs from different clusters.

3.1 Dictionary Learning Assisted AUD

We consider to represent the active MTDs' channel with a learned cluster-specific dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C] \in \mathbb{C}^{M \times G}$, where C denotes the number of clusters, G denotes the number of atoms in the dictionary, and $\mathbf{D}_i \in \mathbb{C}^{M \times g}$ ($i = 1, 2, \dots, C$) denotes the i -th sub-dictionary corresponding to the i -th cluster, with $g = G/C$ denoting the number of dictionary atoms in the sub-dictionary. Let $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2, \dots, \tilde{\mathbf{H}}_C] \in \mathbb{C}^{M \times CN_s}$ be the training data set, where $\tilde{\mathbf{H}}_i \in \mathbb{C}^{M \times N_s}$ ($i = 1, 2, \dots, C$) contains N_s different channels of MTDs in the i -th cluster. Mathematically, we can express $\tilde{\mathbf{H}}_i$ as $\tilde{\mathbf{H}}_i = \mathbf{D}_1 \tilde{\mathbf{X}}_i^1 + \mathbf{D}_2 \tilde{\mathbf{X}}_i^2 + \dots + \mathbf{D}_C \tilde{\mathbf{X}}_i^C$, where $\tilde{\mathbf{X}}_i^j \in \mathbb{C}^{g \times N_s}$ is the coefficient matrix of the i -th cluster corresponding to the j -th sub-dictionary. To differentiate the MTDs in different clusters, we would like $\mathbf{D}_i$ to capture the features of channels for the i -th cluster, so that $\tilde{\mathbf{X}}_i^i$ contains non-zero elements. A similar approach of learning sub-dictionaries for image classification is investigated in [24]. Now, we can formulate the overall dictionary learning problem as follows

$$\begin{aligned} \min_{\{\mathbf{D}_i\}, \{\tilde{\mathbf{X}}_i\}} & \sum_{i=1}^C \|\tilde{\mathbf{H}}_i - \mathbf{D}_i \tilde{\mathbf{X}}_i^i\|_F^2 + \lambda_1 \|\mathbf{D}_i^T \mathbf{D}_i - \mathbf{I}\|_F^2 \\ & + \lambda_2 \|\mathbf{D}_i^T \mathbf{D}_{-i}\|_F^2 \\ \text{s.t.} & \quad \|\mathbf{D}_i(:, k)\|_2^2 = 1, \forall i, k, \end{aligned} \quad (7)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\mathbf{D}_{-i} = [\mathbf{D}_1, \dots, \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, \dots, \mathbf{D}_C] \in \mathbb{C}^{M \times (C-1)g}$ denotes the submatrix by removing $\mathbf{D}_i$ from $\mathbf{D}$. $\mathbf{I} \in \mathbb{R}^{p \times p}$ is an identity matrix. The term $\|\mathbf{D}_i^T \mathbf{D}_i - \mathbf{I}\|_F^2$ and $\|\mathbf{D}_i^T \mathbf{D}_{-i}\|_F^2$ are proposed to improve the irrelevance of atoms within one sub-dictionary and reduce the similarity between different sub-dictionaries, respectively. Similar to [24], $\lambda_1 = \frac{\eta_1 N_s}{g^2} > 0$ and $\lambda_2 = \frac{\eta_2 N_s}{(C-1)g^2} > 0$ are the regularization coefficients for the self-incoherent term and cross-incoherent term, re-

spectively. N_s , C and g denote the number of training samples in one cluster, the number of clusters allowed to access in a frame, and the number of atoms in a sub-dictionary. $\eta_1 > 0$ and $\eta_2 = 2C\eta_1 > 0$ are adjustable parameters. We constrain dictionary atoms with normalized energy to avoid trivial solutions.

Since the objective function is not convex, we alternatively update the sparse coefficient of each cluster and the corresponding sub-dictionary with other variables fixed. First, we fixed $\mathbf{D}_i$ to update the sparse coefficient $\tilde{\mathbf{X}}_i^i$. Owing to the limited scattering, spatial channel in the MIMO system can be represented sparsely under the discrete fourier transform (DFT) [16]. Thus, we initialize the dictionary $\mathbf{D}$ as the redundant DFT matrix obtained by interpolation. Then, the optimization problem for updating sparse coefficients becomes

$$\min_{\tilde{\mathbf{X}}_i^i} \|\tilde{\mathbf{H}}_i - \mathbf{D}_i \tilde{\mathbf{X}}_i^i\|_F^2, \quad (8)$$

which is a least squares (LS) problem. $\tilde{\mathbf{X}}_i^i$ can be updated with the closed-form expression

$$\tilde{\mathbf{X}}_i^i = (\mathbf{D}_i^T \mathbf{D}_i)^{-1} (\mathbf{D}_i^T \tilde{\mathbf{H}}_i). \quad (9)$$

Given the sparse coefficient $\tilde{\mathbf{X}}_i^i$, the optimization problem with respect to the i -th sub-dictionary becomes

$$\begin{aligned} \min_{\mathbf{D}_i} & \|\tilde{\mathbf{H}}_i - \mathbf{D}_i \tilde{\mathbf{X}}_i^i\|_F^2 + \lambda_1 \|\mathbf{D}_i^T \mathbf{D}_i - \mathbf{I}\|_F^2 \\ & + \lambda_2 \|\mathbf{D}_i^T \mathbf{D}_{-i}\|_F^2 \\ \text{s.t.} & \quad \|\mathbf{D}_i(:, k)\| = 1, \forall k, \end{aligned} \quad (10)$$

which can be solved using the gradient descent algorithm. We choose the step size following the Armijo rule.

By iteratively updating the dictionary and sparse coefficients until the reconstruction error no longer decreases, we obtain the learned cluster-specific dictionary. The learned dictionary is then used to identify the clusters of active MTDs. Firstly, according to the detected active pilot set $\hat{\mathcal{M}}$ and the estimated channel matrix $\hat{\mathbf{H}}$, the recovered data symbols is given in Eq. (6). With the assistant of data recovery, the user

detection problem can be formulated as

$$\begin{aligned} \min_{\hat{\mathbf{X}}} & \|\hat{\mathbf{H}}(\hat{\mathcal{M}}, :)^T - \mathbf{D} \hat{\mathbf{X}}\|_F^2 \\ \text{s.t.} & \quad \|\hat{\mathbf{x}}_p\|_{0,d} \leq \hat{s}_b, \quad p \in \hat{\mathcal{M}}. \end{aligned} \quad (11)$$

$\hat{\mathbf{H}}(\hat{\mathcal{M}}, :) \in \mathbb{C}^{\text{card}(\hat{\mathcal{M}}) \times M}$ and $\mathbf{D} \in \mathbb{C}^{M \times G}$ denote the recovered channel of active pilots and the learned cluster-dictionary, respectively. $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_{\text{card}(\hat{\mathcal{M}})}] \in \mathbb{C}^{G \times \text{card}(\hat{\mathcal{M}})}$ denotes the sparse representation under the learned dictionary, where G is the number of atoms in the learned dictionary and $\text{card}(\hat{\mathcal{M}})$ is the Cardinality of set $\hat{\mathcal{M}}$. $\|\hat{\mathbf{x}}_p\|_{0,d} (p \in \hat{\mathcal{M}})$ is the ℓ_0 -norm with block size d and counts the number of non-zero blocks. $\hat{s}_b$ denotes the block sparsity level. If the pilot is sent by only one MTD and the recovered data passes the CRC, $\hat{s}_b$ is set to 1. If the pilot is sent by multiple MTDs from different clusters and the recovered data does not pass the CRC, $\hat{s}_b$ is set to 2 (As the probability of three or more MTDs from different clusters transmitting the same pilot is small, we mainly focus on the condition of two active MTDs collide). Considering the computational complexity, block-OMP (BOMP) [25] is exploited to find the block-sparse representation of the channel under the learned cluster-specific dictionary. The non-zero blocks in the sparse representation indicate the clusters of the MTDs. Therefore, the proposed cluster-specific dictionary enables AUD in the case of same pilot selected by multiple MTDs, where existing methods fail. The proposed cluster-specific dictionary-learning BOMP (CSDL-BOMP) algorithm for AUD is given in Algorithm 1.

3.2 PD-NOMA Based Data Recovery

For MTDs selecting the same pilot, the BS is unable to correctly recover their data due to their superimposed channel. As these conflicted MTDs could be identified by the proposed CSDL-BOMP, the BS broadcasts information to inform the conflicted MTDs to retransmit their pilot on orthogonal resources for channel estimation. Then, conflicted MTDs resend their data on non-orthogonal resources via PD-NOMA, which can improve the spectrum efficiency of data transmission. The complete process of massive access and data transmission is shown in Figure 2.

The basic idea of PD-NOMA is to differentiate users

Algorithm 1. Cluster-specific dictionary-learning BOMP algorithm for AUD.

Input: The received signal $\mathbf{Y}_p$ and $\mathbf{Y}_d$, pilot matrix $\mathbf{P}$, the spreading matrix $\mathbf{S}$, and the learned dictionary $\mathbf{D}$

Output: The detected active user set $\hat{\mathcal{K}}$

- 1: Compute the estimated active pilot set $\hat{\mathcal{M}}$ and the estimated channel $\hat{\mathbf{H}}$ by OMP-MMV
- 2: Compute $\hat{\mathbf{X}}$ with Eq. (5)
- 3: **for** $p \in \hat{\mathcal{M}}$ **do**
- 4: Compute the recovered data symbol $\hat{\mathbf{d}}_p$ with Eq. (6)
- 5: **if** $\hat{\mathbf{d}}_p$ satisfy CRC **then**
- 6: Applying BOMP to $\hat{\mathbf{h}}_p$ with one block. Identify which cluster the channel $\hat{\mathbf{h}}_p$ belongs to based on the index of non-zero block
- 7: **else**
- 8: Applying BOMP to $\hat{\mathbf{h}}_p$ with two blocks. Identify which two clusters of channels are included in the superposed channel based on the index of two non-zero blocks
- 9: **end if**
- 10: Obtain the estimated set of active MTDs that transmit the p -th pilot $\hat{\mathcal{P}}_p$. The detected active user set $\hat{\mathcal{K}} = \hat{\mathcal{K}} \cup \hat{\mathcal{P}}_p$
- 11: **end for**

through different power levels [26]. In uplink PD-NOMA, users in the cell are divided into several user groups, each usually containing 2 ~ 3 users. The BS allocates orthogonal resources to different user groups, while users in one group share the same resource. After receiving the superimposed user signal, the BS performs SIC to recover data. In this paper, before conflicting MTDs resend their data, the BS informs the detected active MTDs to resend orthogonal pilots to estimate their channel. After obtaining the channel state information (CSI), the BS pairs MTDs with the far near pairing method [27] and allocates different power to MTDs of the same pair. Then, MTDs assigned to the same group resend data according to the specified power.

Take a group of two MTDs as an example. The data symbols resent by MTD 1 and MTD 2 are denoted as $\mathbf{d}_1 \in \mathbb{C}^{L_d}$ and $\mathbf{d}_2 \in \mathbb{C}^{L_d}$, respectively. The transmit power of MTD 1 and MTD 2 are p_1 and p_2 . Then, the

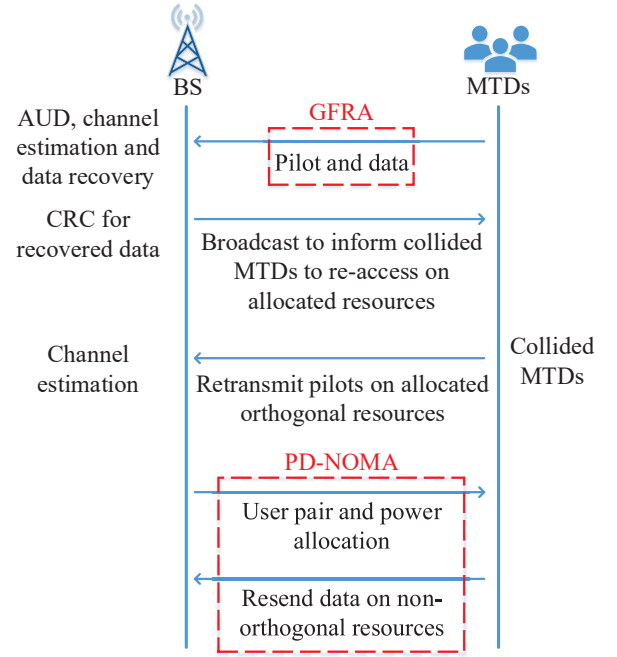


Figure 2. The complete process of massive access and data transmission.

received signal $\mathbf{Y} \in \mathbb{C}^{M \times L_d}$ at the BS is

$$\mathbf{Y} = \sqrt{p_1} \mathbf{h}_1 \mathbf{d}_1^T + \sqrt{p_2} \mathbf{h}_2 \mathbf{d}_2^T + \mathbf{N}, \quad (12)$$

where $\mathbf{h}_1 \in \mathbb{C}^M$ and $\mathbf{h}_2 \in \mathbb{C}^M$ are channel coefficient vectors of MTD 1 and MTD 2, respectively. Without loss of generality, we assume that $\mathbf{h}_1$ and $\mathbf{h}_2$ are accurate, i.e., channel estimation is perfect. $\mathbf{N}$ denotes the additive white Gaussian noise. After the BS receives the superimposed signal, SIC is applied for multi-user detection. Due to the fact that users with better channel conditions have stronger signal to noise ratios (SNRs), users with better channel conditions are usually decoded first. Here, we assume the channel condition of MTD 1 is better. Then, treating the interference from MTD 2 as noise, the BS will first decode the symbol of MTD 1 utilizing the LS method

$$\hat{\mathbf{d}}_1 = \left(\frac{1}{\sqrt{p_1}} (\mathbf{h}_1^T \mathbf{h}_1)^{-1} \mathbf{h}_1^T \mathbf{Y} \right)^T. \quad (13)$$

After successfully decoding the symbol of MTD 1, the BS removes its interference from the received signal $\mathbf{Y}$, as the channel of MTD 1 is known to the BS. Then the signal of MTD 2 can be decoded through the LS

method

$$\hat{\mathbf{d}}_2 = \left(\frac{1}{\sqrt{p_2}} (\mathbf{h}_2^T \mathbf{h}_2)^{-1} \mathbf{h}_2^T (\mathbf{Y} - \sqrt{p_1} \mathbf{h}_1 \hat{\mathbf{d}}_1^T) \right)^T. \quad (14)$$

IV. SIMULATION RESULTS

In this section, we investigate the performance of the proposed cluster-based CS massive access scheme with CSDL-BOMP for AUD and PD-NOMA for data re-transmission. The accuracy of pilot detection is defined as

$$\mu_{\text{pilot}} = \frac{\text{card}(\hat{\mathcal{M}})}{\text{card}(\mathcal{M})}, \quad (15)$$

where $\mathcal{M}$ and $\hat{\mathcal{M}}$ denote the set of active pilot and detected active pilot, respectively. $\text{card}(\cdot)$ is the Cardinality of the set. The successful AUD ratio μ_{user} is defined as

$$\mu_{\text{user}} = \frac{\text{card}(\mathcal{K} \cap \hat{\mathcal{K}})}{\text{card}(\mathcal{K})}, \quad (16)$$

where $\mathcal{K}$ and $\hat{\mathcal{K}}$ denote the set of active MTDs and the set of detected active MTDs, respectively. Although the data recovery is conducted in both activity detection phase and data re-transmission phase, we use the user detection ratio to evaluate the performance in the first phase, while consider the data recovery ratio in the re-transmission to evaluate the performance of the second phase. The successful data recovery ratio μ_{data} is defined as

$$\mu_{\text{data}} = \frac{\text{card}(\mathcal{Q})}{N_{\text{re}}}, \quad (17)$$

with $\mathcal{Q}$ and N_{re} denoting the set of re-transmitted MTDs whose data is correctly recovered and the number of re-transmitted MTDs, respectively. For competitor methods in AUD, we choose the vector approximate message passing (VAMP) [15], sparse Bayesian learning (SBL) with angular domain enhancement (SBL-ADE) [16], and generalized MMV-AMP (GMMV-AMP) [19] for comparison. For resolving collisions, we consider the contention-based method re-access method for comparison against the use of PD-NOMA.

In the experiments, we consider $N = 2000$ stationary MTDs equipped with one antenna are randomly distributed in the cell, and $N_f = 21$ fixed FSCs are uniformly placed in the cell with the radius of 900 meters. Each channel response between the BS and

the MTD consists of AOA from 3 FSCs closest to the MTD and one AOA from a device location-related scattering. MTDs are separated into $2C = 16$ clusters according to their angular information. MTDs in $C = 8$ non-adjacent clusters are allowed to access in one frame. To learn the cluster-specific dictionary, we generate a dataset containing $N_s = 1000$ channel responses per cluster and set the number of atoms in a sub-dictionary g to 8. As for the regularization coefficients for the self-incoherent term and cross-incoherent term, we set η_2 and η_1 to 0.1 and 0.1/16, respectively. A common pilot pool with $N_p = 125$ and $L_p = 64$ is reused between clusters. The length of the spreading sequence is $L_d = 18$. The data is coded by using concatenated code, where the outer layer is a Reed-Solomon (RS) code with parameters (6,4) and the inner layer is a convolutional code with parameters (2,1,6). The modulation type is QPSK. Moreover, the numerical results are averaged over 300 trials.

First, we conducted an experiment to demonstrate the performance gain brought by cluster-based scheme. The comparison is non-cluster based scheme, i.e., not separating MTDs and assigning a unique pilot to all potential MTDs. Figure 3 shows the performance of active pilot detection under a varying numbers of active MTDs. The cluster based scheme outperforms the non-cluster based scheme, as it reduces the number of needed pilot, leading to reduced inter-pilot interference.

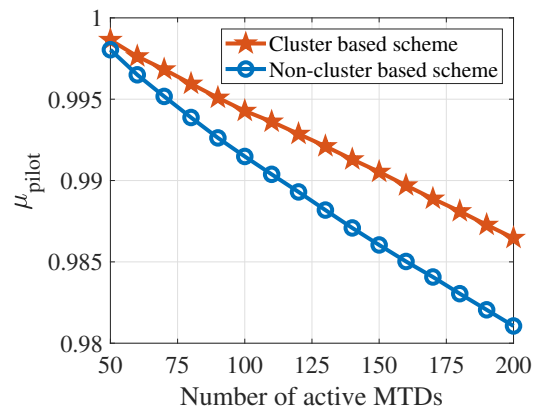


Figure 3. Active pilot detection performance with a varying number of active MTDs.

Figure 4 and Figure 5 show the performance of AUD under various SNRs and number of active users. The proposed method outperforms other competitors for

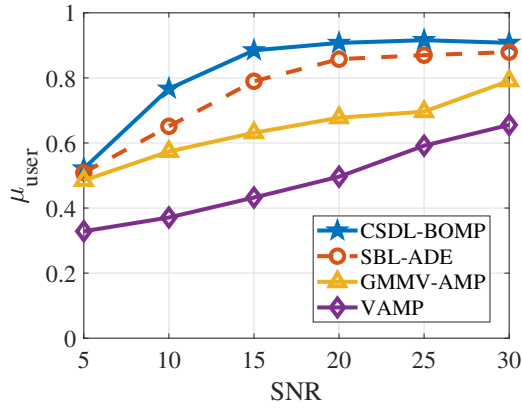


Figure 4. AUD performance under a varying SNRs.

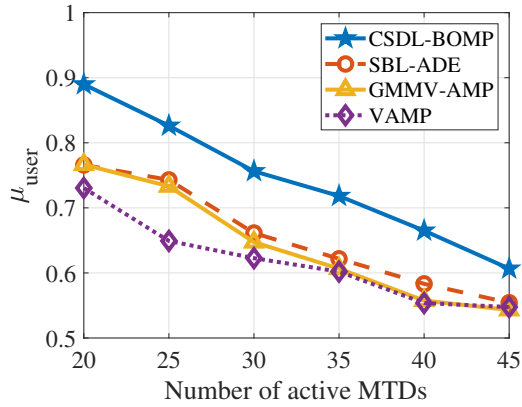


Figure 5. AUD performance with a varying number of active MTDs.

all cases. The proposed method exploits both the spatial channel features learned by dictionary learning and the non-orthogonal pilots to distinguish MTDs, which results in higher capability in AUD. It is worth noting that the dictionary learning is done in an off-line manner using some training dataset collected from the cell. The CSDL-BOMP algorithm only uses the learned cluster-specific dictionary for AUD, which has a relatively low computational complexity. In our simulation platform, we observe that the computing time of the proposed method is only one-third of the SBL-ADE.

Figure 6 shows the performance of data recovery under various SNRs. Consistent with the results of AUD, the proposed method also shows superior performance in data recovery. A part of the gain comes from the improved AUD accuracy, as active MTDs of different clusters can be identified even they transmit the same pilot, and these MTDs' data would be resolved

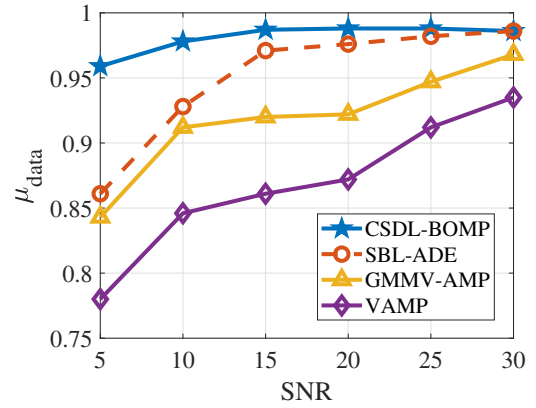


Figure 6. Data recovery performance under a varying SNRs in the data re-transmission phase.

by PD-NOMA during the re-transmission phase. For the SNR of 5 dB, the successful data recover ratio of the propose method reaches over 95%, while the other compared can only achieve about 78 – 86%.

Figure 7 shows the performance of data recovery under a varying number of active users. Under the experimental settings, the successful data recovery ratio keeps at a high level with the increasing number of active MTDs, while the performance of other methods degrades severely. This result is rationale. With more MTDs activated, the number of conflicted MTDs from different clusters, who select the same pilot, tends to increase. The proposed method can identify MTDs from different cluster and resolve collisions via re-transmission with PD-NOMA, while the other methods cannot differentiate these MTDs and the performance of contention based re-transmission degrades with the increasing number of involved MTDs.

ACKNOWLEDGEMENT

This research was supported by Natural Science Foundation of China (62122012, 62221001), the Beijing Natural Science Foundation (L202019, L211012), the Fundamental Research Funds for the Central Universities (2022JBQY004).

V. CONCLUSION

In this paper, we proposed a new massive access method for mMIMO systems, which separates MTDs into clusters and the same pilot pool could be reused among clusters. A cluster-specific dictionary can be

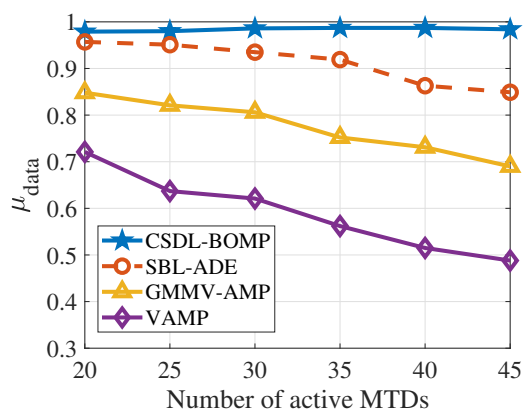


Figure 7. Data recovery performance with a varying number of active MTDs in the data re-transmission phase.

learned from the historical channel dataset of the cell, and used to differentiate active MTDs of different clusters. For detected active MTDs whose data recovery fails, we utilized PD-NOMA with SIC to recover the re-transmitted data. Numerical results reveal that the proposed method has superior performance in AUD and data recovery against the existing methods designed for mMIMO systems.

References

- [1] F. Guo, F. Yu, *et al.*, "Enabling massive iot toward 6g: A comprehensive survey," *IEEE Internet of Things Journal*, 2021, vol. 8, no. 15, pp. 11 891–11 915.
- [2] W. Chen, Y. Liu, *et al.*, "Signal processing and learning for next generation multiple access in 6g," 2023.
- [3] W. Chen, R. He, *et al.*, "Ai assisted phy in future wireless systems: Recent developments and challenges," *China Communications*, 2021, vol. 18, no. 5, pp. 285–297.
- [4] B. Yu and Y. Cai, "Age of information in grant-free random access with massive mimo," *IEEE Wireless Communications Letters*, 2021, vol. 10, no. 7, pp. 1429–1433.
- [5] Z. Rao, J. Jiao, *et al.*, "Grant-free code-domain random access for massive access in internet of things," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 4492–4497.
- [6] H. Jiang, D. Qu, *et al.*, "Enabling lpwan massive access: Grant-free random access with massive mimo," *IEEE Wireless Communications*, 2022, vol. 29, no. 4, pp. 72–77.
- [7] F. Monsees, M. Woltering, *et al.*, "Compressive sensing multi-user detection for multicarrier systems in sporadic machine type communication," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015, pp. 1–5.
- [8] J. Ahn, B. Shim, *et al.*, "Ep-based joint active user detection and channel estimation for massive machine-type communications," *IEEE Transactions on Communications*, 2019, vol. 67, no. 7, pp. 5178–5189.
- [9] B. K. Jeong, B. Shim, *et al.*, "Map-based active user and

- data detection for massive machine-type communications," *IEEE Transactions on Vehicular Technology*, 2018, vol. 67, no. 9, pp. 8481–8494.
- [10] P. Gao, Z. Liu, *et al.*, "Low-complexity block coordinate descend based multiuser detection for uplink grant-free noma," *IEEE Transactions on Vehicular Technology*, 2022, vol. 71, no. 9, pp. 9532–9543.
- [11] S. M. Hasan, K. Mahata, *et al.*, "Uplink grant-free noma with sinusoidal spreading sequences," *IEEE Transactions on Communications*, 2021, vol. 69, no. 6, pp. 3757–3770.
- [12] B. Wang, L. Dai, *et al.*, "Dynamic compressive sensing-based multi-user detection for uplink grant-free noma," *IEEE Communications Letters*, 2016, vol. 20, no. 11, pp. 2320–2323.
- [13] Y. Gao, J. Zheng, *et al.*, "Bidirectional side information aided compressed sensing multiuser detection for uplink gf-noma," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2148–2153.
- [14] Y. Bai, W. Chen, *et al.*, "Contention-based nonorthogonal massive access with massive mimo," *China Communications*, 2020, vol. 17, no. 11, pp. 79–90.
- [15] L. Liu and W. Yu, "Massive connectivity with massive mimo—part i: Device activity detection and channel estimation," *IEEE Transactions on Signal Processing*, 2018, vol. 66, no. 11, pp. 2933–2946.
- [16] W. Chen, H. Xiao, *et al.*, "Joint activity detection and channel estimation in massive mimo systems with angular domain enhancement," *IEEE Transactions on Wireless Communications*, 2022, vol. 21, no. 5, pp. 2999–3011.
- [17] L. Wu, Z. Wang, *et al.*, "Temporal correlation enhanced sparse activity detection in mimo enabled grant-free noma," *IEEE Transactions on Vehicular Technology*, 2022, vol. 71, no. 3, pp. 2887–2899.
- [18] Z. Gao, H. Xiu, *et al.*, "Multi-panel extra-large scale mimo based joint activity detection and channel estimation for near-field massive iot access," *China Communications*, 2023, vol. 20, no. 5, pp. 232–243.
- [19] M. Ke, Z. Gao, *et al.*, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive mimo," *IEEE Transactions on Signal Processing*, 2020, vol. 68, pp. 764–779.
- [20] J. Salo, G. Del Galdo, *et al.*, "Matlab implementation of the 3gpp spatial channel model (3gpp tr 25.996)," *on-line*, 2005.
- [21] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, 2006, vol. 54, no. 12, pp. 4634–4643.
- [22] L. Yang, J. Dong, *et al.*, "Simplified structured analysis dictionary learning for image classification," in *2022 7th International Conference on Computational Intelligence and Applications (ICCIA)*, 2022, pp. 156–160.
- [23] T. Liu, Y. Gu, *et al.*, "Separable coupled dictionary learning for large-scene precise classification of multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, 2022, vol. 60, pp. 1–14.
- [24] S. Gao, I. W. Tsang, *et al.*, "Learning category-specific dictionary and shared dictionary for fine-grained image categorization," *IEEE Transactions on Image Processing*, 2014, vol. 23, no. 2, pp. 623–634.
- [25] P. Sermwuthisarn, S. Auethavekiat, *et al.*, "A fast image

recovery using compressive sensing technique with block based orthogonal matching pursuit,” in *2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2009, pp. 212–215.

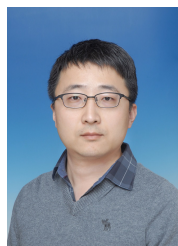
- [26] O. Maraqa, A. S. Rajasekaran, *et al.*, “A survey of rate-optimal power domain noma with enabling technologies of future wireless networks,” *IEEE Communications Surveys & Tutorials*, 2020, vol. 22, no. 4, pp. 2192–2235.
- [27] S. Chen, K. Peng, *et al.*, “A suboptimal scheme for uplink noma in 5g systems,” in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2015, pp. 1429–1434.

Biographies



liang@bjtu.edu.cn.

Shiyu Liang is pursuing the M.Eng. degree at Beijing Jiaotong University, and has received the B.Eng. degree from Beijing Jiaotong University in 2021. Her current research interests include AI-assisted communications and massive machine-type communications. Email: shiyu-



sparse signal processing. Email: weich@bjtu.edu.cn.

Wei Chen is a professor at Beijing Jiaotong University. He received both the B.Eng. degree and the M.Eng. degree from Beijing University of Posts and Telecommunications, and the Ph.D. degree in Computer Science from the University of Cambridge. His current research interests include AI-assisted communications, machine type communications and



Zhongwen Sun is pursuing the M.Eng. degree at Beijing Jiaotong University, and has received the B.Eng. degree from Beijing Jiaotong University in 2022. His current research interests include compressive sensing based on AI and critical machine-type communications. Email: 22120121@bjtu.edu.cn.



Ao Chen is pursuing the Ph.D. degree at Beijing Jiaotong University, and has received the B.Eng degree from Beijing Jiaotong University in 2023. His research interests include AI-assisted communications, deep learning, and massive machine-type communications. Email: aochen@bjtu.edu.cn.



Bo Ai is a professor at Beijing Jiaotong University. He received his Master and Ph.D. degree from Xidian University in 2002 and 2004 in China, respectively. His current interests are broadband mobile communication systems, specialized mobile communication, and artificial intelligence. Email: boai@bjtu.edu.cn.