IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 20, 2017, accepted July 10, 2017, date of publication July 20, 2017, date of current version August 8, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2729878

Overlapping User Grouping in IoT Oriented Massive MIMO Systems

RUN TIAN[1,2], YUAN LIANG[2], (Student Member, IEEE), XUEZHI TAN[1], (Member, IEEE), AND TONGTONG LI[2], (Senior Member, IEEE)

[1]Harbin Institute of Technology, Harbin 150001, China
[2]Michigan State University, East Lansing, MI 48824, USA

Corresponding author: Run Tian (tianrun@msu.edu)

ABSTRACT This paper considers network capacity and user coverage improvement in Internet of Things (IoT)-oriented massive MIMO systems. In the literature, user grouping approaches have been used in massive MIMO to improve the network capacity, where users are generally divided into non-overlapping groups, and those users with less favorable channel conditions are dropped for capacity optimization. As a result, users may suffer from unpredicted interruptions and delays, even long time disconnection from the network. Moreover, non-overlapping user grouping also leads to unnecessary resource waste. As an effort to overcome these limitations, in this paper, we introduce the concept of overlapping user grouping by exploiting the favorable propagation property in massive MIMO. More specifically, we propose two new user grouping approaches. First, we present a greedy search-based user grouping method by allowing overlapping among the selected subgroups. Second, we introduce a new channel similarity measure, and develop a low complexity overlapping user grouping approach based on the spectral clustering algorithm in machine learning. Both the theoretical and numerical results demonstrate that: overlapping user grouping can achieve much higher network capacity, and can ensure that at any given time, each IoT device will be served in at least one subgroup.

INDEX TERMS Massive multiple-input multiple-output (MIMO), Internet of Things (IoT), overlapping user grouping.

I. INTRODUCTION

In Internet of Things (IoT) oriented massive multiple input and multiple output (MIMO) systems, where one base station (BS) with an enormous number of antennas serves a large group of single antenna devices simultaneously, channel correlation among the users or devices has significant impact on system capacity [1], [2]. More specifically, when applying the beamforming technology, systems with improper user selection may allocate users with high correlation into the same beamforming subgroup, resulting in severe capacity loss.

To maximize system capacity, a number of user selection and user grouping methods have been proposed in literature. The basic idea there is to select the subgroup of users that can achieve the highest system aggregate capacity. For example, in [3], Dimic *et al.* considered the case of more users than transmit antennas, and proposed a suboptimal greedy user selection algorithm in multiuser MIMO systems. The algorithm iteratively selects the users that have the greatest contributions to the aggregate capacity until no more increase can be achieved. Although the computational complexity is much lower than the exhaustive search-based approach, it only achieves a fraction of the optimal capacity. In [4], Wang *et al.* proposed a generalized greedy user selection scheme based on sequential water-filling (SWF). This SWF algorithm introduces the recursive LQ decomposition to simplify the calculation of Moore-Penrose matrix inverse, and achieves a higher sum rate than the conventional greedy user selection method. In [5], Huang *et al.* proposed a greedy user selection algorithm with swap (GUSS). Relying on the ‘‘delete’’ and ‘‘swap’’ operations, GUSS may escape from potential *local optimum* and achieves a better performance at the cost of increased number of iterations and higher computational complexity. In [6], Nam *et al.* proposed a two-stage precoding strategy based on user grouping. In order to reduce the overhead of channel estimation in MIMO systems, it partitions users into pre-beamforming subgroups, then simplify the channel estimation and beamforming in the grouping stage. Furthermore, in [7], Shen *et al.* proposed a branch-and-bound iterative user grouping algorithm named SIEVE. By adjusting the sieve size K dynamically, in each iteration, the SIEVE algorithm can achieve a trade-off between search complexity and achievable capacity.

As can be seen, existing user selection approaches have been focused on capacity maximization, but sort of overlooked the user coverage problem: for each round, users with favorable channel conditions will be selected, but those with less favorable channel conditions will be dropped. This turns out to be a major limitation with existing approaches, as users may suffer from unpredicted interruptions and delays. For fixed IoT devices with quasi-static channels, the situation can be even worse. Some devices may not be served for a long time since they might be dropped in every round.

Moreover, influenced by the traditional "collision-free" user division concept in wireless communications, existing user grouping allows no overlapping among user groups. However, we should note that a key property of the radio channels in massive MIMO is *favorable propagation* [8], which is defined as the mutual orthogonality among vector-valued channels corresponding to each terminal. If we insist that the user groups is non-overlapping, then we actually cannot take the full advantage of the favorable propagation property.

Motivated by these observations, in this paper, we introduce the concept of "overlapping user grouping", aiming to increase the system capacity and at the same time ensure full user coverage. More specifically, we present two new user grouping approaches. First, we propose a greedy search based user grouping method by allowing overlapping among the selected subgroups. Second, we introduce a new channel similarity measure, and develop a low complexity overlapping user grouping approach based on the spectral clustering algorithm in machine learning. Both the theoretical and numerical results demonstrate that: with respect to non-overlapping schemes, overlapping user grouping can achieve much higher network capacity, and can ensure that at any given time, each IoT device will be served in at least one subgroup.

The rest of the paper is organized as follows. In Section II, the system model is introduced and the problem is formulated. In section III, the two proposed overlapping user grouping approaches are presented. Capacity analysis and complexity evaluation of the proposed approaches are carried out in Section IV. Numerical simulation results are presented in Section V and we conclude in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

Without loss of generality, we consider the downlink of a centralized massive MIMO system, formed by an M-antenna base station and K single antenna IoT devices. The transmit antennas can have different geometric structures, e.g., being placed along a line to form a uniform linear array (ULA), or along a circle to form a uniform circular array (UCA). Let $\mathbf{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ denote the actual channel matrix, where $(\cdot)^{\mathrm{T}}$ is the transposition of a matrix or vector, and $\boldsymbol{h}_k$ the complex channel vector between the base station and user k. We assume that the channels are quasi-static and flat-fading, such that the matrix $\mathbf{H}$ can be taken as invariant for a few time slots. Thus, the received signals of the K users can be expressed as

$$\boldsymbol{y} = \mathbf{HB}\boldsymbol{d} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{C}^{K\times 1}$ denotes the received data for all the K users in a single time slot, $\mathbf{B} \in \mathbb{C}^{M\times K}$ the precoding matrix at the base station, $\boldsymbol{d} \in \mathbb{C}^{K\times 1}$ the data vector for all the K users, and $\boldsymbol{n} \sim \mathcal{CN}(0, \mathbf{I}_K)$ the additive white Gaussian noise vector of zero mean and unit variance. Throughout this paper, we use bold upper case and lower case letters to denote matrices and vectors, respectively; and use normal letters to represent scalars.

B. CHANNEL CORRELATION MODEL

Consider a set of spatially correlated Rayleigh channels with non-line-of-sight (NLOS) propagation. According to the Kronecker correlation model [9], the channel matrix can be expressed as

$$\mathbf{H} = \mathbf{R}_{RX}^{\frac{1}{2}}\mathbf{H}_{iid}\mathbf{R}_{TX}^{\frac{1}{2}}, \tag{2}$$

where $\mathbf{H}_{iid} \in \mathbb{C}^{K\times M}$ is an uncorrelated Rayleigh channel matrix, whose elements are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance. $\mathbf{R}_{RX} \in \mathbb{C}^{K\times K}$ and $\mathbf{R}_{TX} \in \mathbb{C}^{M\times M}$ denote the *spatial correlation matrix* at the receiver and the transmitter, respectively.

In practical downlink transmissions, base station is usually free of local scattering, which may result in high correlation among the transmit antennas. Let θ be the azimuth angle of the user location, S the distance between the base station and the far field scatterer ring with the radius r, and Δ the angle spread of the transmit signal, which can be approximated as $\Delta \approx \arctan(r/S)$. According to the one-ring MIMO channel model shown in Fig. 1, which is also adopted by [10] and [11], the spatial correlation coefficient between transmit antennas $1 \leq p, q \leq M$, can generally be

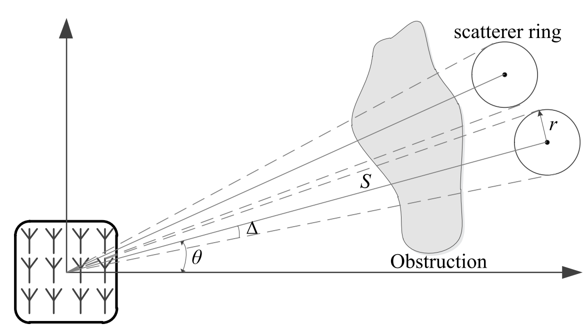

FIGURE 1. One-ring MIMO channel model with shadow fading.

modeled as:

$$[\mathbf{R}_{TX}]_{p,q} = \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} e^{j\mathbf{g}^{\mathrm{T}}(\alpha+\theta)(\mathbf{u}_p-\mathbf{u}_q)} d\alpha. \tag{3}$$

Here $\mathbf{g}(\alpha+\theta) = -\frac{2\pi}{\lambda}(cos(\alpha+\theta), sin(\alpha+\theta))^{\mathrm{T}}$ is the vector for a planar wave impinging the antennas with the Angle of Arrival (AoA) $\alpha+\theta$ and the carrier wavelength λ. $\mathbf{u}_p$ and $\mathbf{u}_q$ are the position of the transmit antennas p and q, respectively. It can be verified that the spatial correlation matrix $\mathbf{R}_{TX}$ is a normal matrix with eigen decomposition:

$$[\mathbf{R}_{TX}]_{p,q} = \mathbf{U}_{TX}\mathbf{\Sigma}_{TX}\mathbf{U}_{TX}^{\mathrm{H}}, \tag{4}$$

in which $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose of a matrix, $\mathbf{U}_{TX}$ is a unitary matrix composed of the eigenvectors of $\mathbf{R}_{TX}$, and $\mathbf{\Sigma}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{R}_{TX}$.

Furthermore, in IoT networks, devices or sensors located in houses or buildings usually experience *shadow fading*, that is, the signal power fluctuation due to obstacles on the transmit path. It has been shown that the channels affected by the same shadowing could be significant correlated when the receivers are geographically close to each other, and this may have a strong impact on system performance. In a random shadowing environment obeying the NeSH model [12], the correlation of channels between users i, j, where $1 \le i, j \le K$, is modeled as:

$$[\mathbf{R}_{RX}]_{i,j} = \frac{\sigma_s^2}{d_{cor}} e^{-\frac{|d_{i,j}|}{d_{cor}}}, \tag{5}$$

where $|d_{i,j}|$ represents the distance between the two users, σ_s is the standard deviation of shadow fading, and d_{cor} is defined as the *correlation distance* corresponding to the distance at which the correlation drops to 0.5. It also can be verified that at the receiver, the spatial correlation matrix $\mathbf{R}_{RX}$ is also a normal matrix. It then follows from (2) that the actual channel can be represented as

$$\mathbf{H} = \mathbf{U}_{RX}\mathbf{\Sigma}_{RX}^{\frac{1}{2}}\mathbf{H}_{iid}\mathbf{\Sigma}_{TX}^{\frac{1}{2}}\mathbf{U}_{TX}. \tag{6}$$

C. FORMULATION OF THE OVERLAPPING USER GROUPING PROBLEM

1) MULTIUSER INTERFERENCE CANCELLATION IN MASSIVE MIMO

To eliminate multi-user interference, zero-forcing beamforming (ZFBF) is often used to achieve channel orthogonalization. Let $\boldsymbol{x} = \mathbf{B}\boldsymbol{d}$ denote the transmit signal, which is subject to the transmit power constraint P_T with $\mathbb{E}[\|\boldsymbol{x}\|^2] \le P_T$. Zero-forcing beamforming eliminates the multi-user interference by choosing the proper beamforming weight matrix $\mathbf{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K]^{\mathrm{H}}$, to satisfy

$$\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{b}_j = \begin{cases} \gamma_i, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases} \tag{7}$$

That is, the beamforming vector for each user should be orthogonal to the subspace spanned by the channel vectors of all the other concurrent users in the same beamforming subgroup, i.e., $\mathbf{V}_i = span\{\boldsymbol{h}_j | j \in Group, j \neq i\}$. The projection matrix of the orthogonal subspace $\mathbf{V}_i^{\perp}$ is denoted by

$$\mathbf{P}_i = \left(\mathbf{I}_M - \mathbf{V}_i^{\mathrm{H}}(\mathbf{V}_i\mathbf{V}_i^{\mathrm{H}})^{-1}\mathbf{V}_i\right), \tag{8}$$

where $\mathbf{P}_i \in \mathbb{C}^{M\times M}$ and $\mathbf{V}_i^{\mathrm{H}}\mathbf{P}_i = \mathbf{0}$. This implies that $\mathbf{B}$ is the Moore-Penrose pseudo-inverse of the channel matrix [13]

$$\mathbf{B} = \mathbf{H}^{\dagger} = \mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}. \tag{9}$$

The optimal power allocation which achieves the maximum sum rate can be derived using the water-filling scheme

$$\beta_k = \max\left(\left(\frac{1}{\mu} - \frac{1}{\lambda_k}\right), 0\right), \tag{10}$$

where β_k is the power allocation factor, $\lambda_k = \|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{b}_k\|$ is the equivalent channel gain after beamforming and μ is called the water level, satisfying

$$\sum_{k\in Group}\left(\frac{1}{\mu} - \frac{1}{\lambda_k}\right) = P_T. \tag{11}$$

Since we have $\mathbf{B} = \mathbf{H}^{\dagger}$, the multiuser interference within the beamforming subgroup can be eliminated by projecting the intended user's signal into the null space of all the other concurrent users.

2) EXPLOITING THE FAVORABLE PROPAGATION IN MASSIVE MIMO

As we know, the projection operation in zero-forcing beamforming may reduce the channel gain and result in capacity decreasing [14]. The loss of the channel gain after projection is illustrated in Fig. 2.

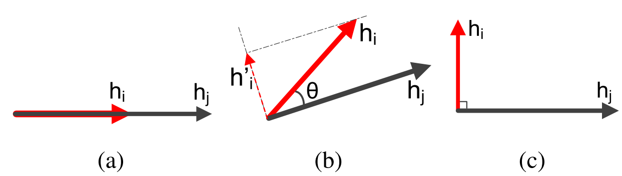

FIGURE 2. Zero-forcing beamforming in two-dimensional channel vector spaces. (a) Overlap channels. (b) General case. (c) Orthogonal channels.

Undoubtedly, the worst-case of the two channel vectors is depicted in Fig. 2a, in which $\boldsymbol{h}_j$ overshadows $\boldsymbol{h}_i$ in the user channel space, which makes it almost impossible to decode the received signals. Fig. 2b shows a general case that the angle between two channel vectors takes a certain value. The channel gain of h_i experiences a severe loss after being projected into the the null space of channel h_j (see h_i'). The most favorable case is shown in Fig. 2c where channels are orthogonal to each other, which is known as the *favorable propagation* property in massive MIMO [8]. In this case, the projected channel vector is equal to the original one, yielding the highest channel gain.

More specifically, favorable propagation is defined as the mutual orthogonality among vector-valued channels corresponding to each terminal [8]. Consider a massive MIMO system where the base station transmits data simultaneously and independently to K single antenna terminals. The base station has M transmit antennas and all K terminals share the same time-frequency resources. Let $\boldsymbol{h}_k \in \mathbb{C}^{M\times 1}$ denote the channel vector between the base station and the kth terminal. Under both independent Rayleigh fading case (i.e., when there is no line-of-sight path) and uniform random LOS case, it can be proved that

$$\frac{1}{M}\|\boldsymbol{h}_k\|^2 = 1$$
$$\frac{1}{M}\boldsymbol{h}_k^H\boldsymbol{h}_j \to 0, \quad M\to\infty,\ k\neq j.$$

3) CAPACITY OPTIMIZATION BASED ON OVERLAPPING USER GROUPING

The aggregate capacity of the system is largely determined by the channel correlation among the users in the same subgroup, and can be optimized with proper user grouping method. Motivated by the approximate orthogonality in massive MIMO systems, in this research, we propose to exploit favorable propagation to enable users to share the time-frequency resources more efficiently, so as to achieve higher capacity and full user coverage.

Let $\mathbb{S} = \{k|\ k = 1, 2, \ldots, K\}$ denote the whole user set and suppose that we have G subgroups. For $i = 1, 2, \ldots, G$, let $\mathbb{S}_i$ denote the i-th subgroup, the sum capacity of each subgroup can be obtained as

$$R_i(\mathbf{H},\mathbf{B}) = \sum_{k\in\mathbb{S}_i}\log_2(1+\beta_k\|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{b}_k\|). \tag{12}$$

We define the optimal user grouping strategy as the one that maximizes the sum rate. Note that the optimal user grouping problem is determined by channel matrix $\mathbf{H}$ and power constraint P_T. Let $\boldsymbol{S}^*(\mathbf{H}, P_T) \triangleq \{\mathbb{S}_1, \ldots, \mathbb{S}_G\}$ denote the optimal user grouping, and $\boldsymbol{\beta}^*(\mathbf{H}, P_T) \triangleq \{\beta_1, \ldots, \beta_K\}$ the power allocation factor that yields the maximum sum rate under the grouping strategy $\boldsymbol{S}^*(\mathbf{H}, P_T)$. Then the optimal user grouping problem can be formulated as

$$\left\{\boldsymbol{S}^*(\mathbf{H}, P_T), \boldsymbol{\beta}^*(\mathbf{H}, P_T)\right\} = \arg\max \sum_{i=1}^{G} R_i$$
$$\text{subject to} \bigcup_{i=1}^{G}\mathbb{S}_i = \mathbb{S},$$
$$\sum_{k=1}^{K}\beta_k \le P_T \tag{13}$$

As can be seen, here we allow overlapping among subgroups when solving the optimization problem. In the following, we propose to solve the overlapping user grouping problem using the greedy search based approach, and machine learning based spectral clustering.

III. THE PROPOSED OVERLAPPING USER GROUPING APPROACHES

In this section, we introduce two overlapping user grouping approaches: overlapping user grouping based on the greedy algorithm (OUG-Greedy), and overlapping user grouping based on spectral clustering (OUG-SC) in machine learning.

A. OVERLAPPING USER GROUPING BASED ON THE GREEDY ALGORITHM

We first introduce a traditional greedy user selection method called zero-forcing with selection (ZFS) which was proposed in [3], then improve the method by performing ZFS iteratively while allowing overlapping among the selected subgroups.

1) ZERO-FORCING WITH SELECTION

The optimal user grouping can be achieved through exhaustive search or combinatorial search algorithms. However, the combinatorial nature of these problems make it computationally infeasible due to the large number of users and antennas in massive MIMO systems [15]. Therefore, many suboptimal schemes have been proposed for user selection and one typical method is the ZFS algorithm. ZFS is a greedy algorithm which adopts the forward selection strategy. That is, in each iteration of ZFS, the algorithm selects one user that can yield the largest increase to the sum capacity until the number of selected users reaches the number of antennas, or no more improvement can be achieved on the sum capacity. To simplify the presentation of the proposed OUG-Greedy algorithm, we rewrite the ZFS algorithm in a recursive way, which is summarized in Table 1, where the function $C_{ZFBF}(\mathbb{X})$ calculates the sum capacity of user set $\mathbb{X}$ using zero-forcing beam forming. For initialization, we set input $\mathbb{S} = \emptyset$ and $\mathbb{C} = \{1, 2, \ldots, K\}$.

TABLE 1. Zero-forcing with selection algorithm [3].

Algorithm	$ZFS(\mathbb{S}_i, \mathbb{C}, \mathbf{H}, \beta_i)$
Input:	$\mathbb{S}_i$: The set of users that have been selected in the current subgroup. $\mathbb{C}$: The set of remaining users that have not been selected. $\mathbf{H}$: The channel matrix of all the users. $\boldsymbol{\beta}_i$: The power constraint for this group.
Output:	$\mathbb{S}_i^*$: The set of users that are finally selected. $\mathbb{C}^*$: The set of remaining users after current iteration.
Step 1:	**If** $\mathbb{C} == \emptyset$ **or** $\lvert\mathbb{S}_i\rvert = M$ /* M: The number of antennas at base station */ **Return** $\mathbb{S}_i$ **Else** $k_{opt} \leftarrow \arg\max_{k\in\mathbb{C}} C_{ZFBF}(\mathbb{S}_i \bigcup k)$, where $C_{ZFBF}(\mathbb{S}_i) = \max \sum_{k\in\mathbb{S}_i}\log_2\left(1+\frac{\beta_k}{[(\mathbf{H}_{\mathbb{S}_i}^{\mathrm{H}}\mathbf{H}_{\mathbb{S}_i})^{-1}]_{k,k}}\right)$ s.t. $\sum_{k\in\mathbb{S}_i}\beta_k = \beta_i$ /* β_k is power allocation factor for the k-th user in set $\mathbb{S}_i$ */
Step 2:	**If** $C_{ZFBF}(\mathbb{S}_i\bigcup\{k_{opt}\}) > C_{ZFBF}(\mathbb{S}_i)$ **Return** $\mathbb{S}_i^* = \mathbb{S}_i\bigcup\{k_{opt}\}$, $\mathbb{C}^* = \mathbb{C}\backslash k_{opt}$ **Else** **Return** $\mathbb{S}_i^*$, $\mathbb{C}^*$

Based on the ZFS, in [4], Wang *et al.* proposed the sequential water-filling (SWF) algorithm, which is a generalized and simplified version of ZFS joint with water-filling power allocation. The computational complexity of SWF is reduced to $\mathcal{O}(KM^3)$ per subgroup.

2) THE PROPOSED ALGORITHM

In our proposed algorithm, we resolve the following two issues that were not addressed in the existing user grouping methods:

- To ensure at any given time, each IoT device is served by at least one subgroup.
- To further increase the sum rate by allowing overlapping among the subgroups.

TABLE 2. Overlapping user grouping based on the greedy algorithm.

Algorithm 1 $OUG\text{-}Greedy(\mathbb{C}_i, \mathbf{H}, P_T)$	
Input:	$\mathbb{C}_i$: The set of candidate users to be grouped in the i-th iteration . $\mathbf{H}$: The channel matrix of all the users. P_T: The total power constraint.
Output:	$\boldsymbol{S}_o = \{\mathbb{S}_{1,o}, \ldots, \mathbb{S}_{G,o}\}$: The sets of users of all the overlapping beamforming subgroups. $\boldsymbol{\beta}_o = \{\beta_{1,o}, \ldots, \beta_{G,o}\}$: The power allocation factor for overlapping beamforming subgroups.
Initialization1:	$i \leftarrow 1$ /* i: Iteration index */
Loop1:	**While** $\mathbb{C}_i \neq \emptyset$ $\mathbb{S}_i \leftarrow ZFS(\emptyset, \mathbb{C}_i, \mathbf{H}, \beta_i)$ /* Perform ZFS algorithm for group i */ /* $\boldsymbol{\beta}_i$ can be obtained by the water-filling scheme, see (10) */ $\mathbb{C}_{i+1} \leftarrow \mathbb{C}_i \backslash \mathbb{S}_i$
Initialization2:	$i \leftarrow 1$
Loop2:	**If** $i > 1$ **And** $i \leq G$ $\mathbb{S}_{i,o} \leftarrow ZFS(\mathbb{S}_i, \bigcup_{j=1}^{i-1} \mathbb{S}_j, \mathbf{H}, \beta_i)$ $i \leftarrow i + 1$
Return:	$\boldsymbol{S}_o = \{\mathbb{S}_{1,o}, \ldots, \mathbb{S}_{G,o}\}$, $\boldsymbol{\beta}_o = \{\beta_{1,o}, \ldots, \beta_{G,o}\}$

The proposed OUG-Greedy algorithm is an iterative greedy search algorithm and summarized in Table 2. Let the whole user set be denoted as $\mathbb{S}$ and the set of users which have not been assigned into any subgroup in the i-th iteration, be denoted as $\mathbb{C}_i$. Start with the first iteration $i = 1$ and $\mathbb{C}_1 = \mathbb{S}$, which means no user has been grouped, we perform the typical ZFS algorithm [3] in each iteration until all the users have been grouped. More specifically, in the i-th iteration, the proposed OUG-Greedy algorithm selects the bunch of users out from $\mathbb{C}_i$ to form the subgroup $\mathbb{S}_i$, which yields the maximum sum rate with the water-filling power allocation, i.e.

$$\mathbb{S}_i = \arg\max_{\mathcal{A} \subset \mathbb{C}_i} \left\{ \max_{\beta_k} \sum_{k \in \mathcal{A}} \log_2(1 + \beta_k \|\boldsymbol{h}_k \boldsymbol{b}_k\|) \right\}. \tag{14}$$

where β_k denotes the power allocation factor for user k. For the next iteration, we first remove the users in group $\mathbb{S}_i$ from $\mathbb{C}_i$, i.e.

$$\mathbb{C}_{i+1} = \mathbb{C}_i \backslash \mathbb{S}_i. \tag{15}$$

We then move on to the $(i + 1)$-th iteration to select the next subgroup $\mathbb{S}_{i+1}$ from the remaining user set $\mathbb{C}_{i+1}$ using ZFS, and continue the process until all the users have been assigned to a subgroup.

After the user selection for each subgroup using ZFS, we extend the searching space of subgroup $\mathbb{S}_i$, where $i = 2, \ldots, G$, to the users that have been assigned to the previous subgroups. That is, for the i-th subgroup $\mathbb{S}_i$ obtained from the previous steps, we reset the searching space as

$$\mathbb{C}_{i,o} = \bigcup_{j=1}^{i-1} \mathbb{S}_j \tag{16}$$

to perform the *overlapping user selection*. Such that, *whenever possible, users with favorable channel conditions* (that is, the channel vectors of these users are approximately orthogonal to all the other users), *can be reselected and assigned to multiple beamforming subgroups simultaneously*.

Let $\mathbb{S}_{i,o}$ denote the newly obtained subgroup i with overlapping. As we will show in Section IV, for $i = 1, \ldots, G$, $\mathbb{S}_{i,o}$ is always larger than or equal to the corresponding subgroup obtained with non-overlapping user group. It then follows that, the overall sum rate achieved by all the users in $\mathbb{S}_{i,o}$ is greater than or equal to that in $\mathbb{S}_i$.

As can be seen, our approach breaks the collision-free barrier in existing user grouping algorithms and increases the overall capacity by exploiting favorable propagation property and allowing overlapping among the subgroups.

B. OVERLAPPING USER GROUPING BASED ON SPECTRAL CLUSTERING

To explore the trade-off between performance and complexity, instead of using the forward selection strategy as in the greedy based approaches, we propose an alternative way of overlapping user grouping based on spectral clustering in machine learning.

1) METHOD DESCRIPTION

In zero-forcing beamforming, multi-user interference is canceled by projecting the channel vector of each user onto the null space of the channel vectors of all the other concurrent users in the same subgroup. As discussed in Section II-C, if the angle between the channel vectors of two concurrent users is too small, i.e., the two channel vectors are approximately in the same direction in Euclidean space, these two users may cause significant interference to each other. Thus, instead of optimizing the sum capacity directly, we propose to optimize the difference between the directions of channel vectors of concurrent users. More specifically, the steps of the proposed OUG-SC algorithm are listed as follow:

- Cluster the users such that the channel vectors of the users from the same cluster lie in the same direction in Euclidean space, while the channel vectors of users from different clusters lie in relatively different directions.
- Generate beamforming subgroups by selecting one user from each cluster. This implies the users which are assigned into the same subgroup hold relatively

different directions and guarantees the multi-user interference within each subgroup is maintained at a lower level. Note that the cluster number should be smaller than or equal to the number of antennas M.

2) ALGORITHM DESIGN

Spectral clustering refers to a class of clustering methods that approximate the problem of partitioning nodes in a weighted graph as eigenvalue problems. The weighted graph represents a similarity matrix between the objects associated with the nodes in the graph.

To perform the proposed spectral clustering based user grouping, the first step is to define the similarity measure, more specifically, the similarity matrix between the channel directions. Then to identify and cluster users automatically based on their channel direction similarities in Euclidean space, which is a typical scenario of spectral clustering [16], [17]. Let $w_{i,j}$ be the similarity measure between the i-th and j-th channel vectors. We choose the generalized Fubini-Study distance [18]

$$w_{i,j} = d_{FS}(\boldsymbol{h}_i, \boldsymbol{h}_j) = \arccos \frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{h}_j\boldsymbol{h}_j^{\mathrm{H}}\boldsymbol{h}_i|}{\|\boldsymbol{h}_i\|\|\boldsymbol{h}_j\|} \tag{17}$$

as the similarity measure. This Fubini-Study distance first calculates the projection of auto-correlation between different channel vectors as the similarity measure of the channel directions, then performs normalization and amplifies the difference with the inverse cosine function. Note that the inverse cosine operation makes it more sensitive to the small angle deference between two channel vectors, in which case these two users may suffer from significant interference.

With the similarity measure defined in (17), we will cluster channel vectors into C clusters, denoted by $\mathbb{A}_1, \mathbb{A}_2, \ldots, \mathbb{A}_C$, note that $C \leq M$. Recall that the whole user set $\mathbb{S} = \{k | k = 1, 2, \ldots, K\}$, the optimization problem of the spectral clustering algorithm is defined as

$$\begin{aligned} \{\mathbb{A}_1, \mathbb{A}_2, \ldots, \mathbb{A}_C\} = \arg\min \ & \frac{1}{2}\sum_{c=1}^{C} \frac{\sum_{i\in\mathbb{A}_c, j\notin\mathbb{A}_c} w_{i,j}}{|\mathbb{A}_c|} \\ \text{subject to } & \mathbb{A}_c \subset \mathbb{S} \quad \text{for } 1 \leq c \leq C \\ & \mathbb{A}_i \cap \mathbb{A}_j = \emptyset \quad \text{for any } i \neq j \\ & \bigcup_{c=1}^{C} \mathbb{A}_c = \mathbb{S}. \end{aligned} \tag{18}$$

Define the degree of the i-th channel vector as

$$d_i = \sum_{j=1}^{K} w_{i,j}.$$

The *degree matrix* $\mathbf{D}$ is defined as the diagonal matrix with the degrees $d_1, d_2, \ldots, d_K$ on the diagonal, and the weighted *adjacency matrix* is defined as $\mathbf{W} = [w_{i,j}]_{i,j=1,2,\ldots,K}$. Then the unnormalized graph *Laplacian matrix* can be obtained as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}.$$

Due to the symmetric structure of the Laplacian matrix $\mathbf{L}$, it is positive semi-definite, and for an arbitrary vector $\boldsymbol{f} = [f_1, f_2, \ldots, f_K], \boldsymbol{f} \in \mathbb{C}^{K\times 1}$, matrix $\mathbf{L}$ satisfies

$$\boldsymbol{f}^{\mathrm{T}}\mathbf{L}\boldsymbol{f} = \frac{1}{2}\sum_{i,j=1}^{K} w_{i,j}(f_i - f_j)^2. \tag{19}$$

As we can see, Eqn. (19) holds the same structure as the objective function in optimization problem (18) when $(f_i - f_j)^2 = \frac{1}{|\mathbb{A}_c|}$.

With the arguments above, the optimization problem (18) is converted to a *RatioCut* minimization problem

$$\begin{aligned} \mathbf{F}^* = \arg \min_{\mathbb{A}_1,\mathbb{A}_2,\ldots,\mathbb{A}_C} & \operatorname{Tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F}) \\ \text{subject to } & \mathbf{F} = [f_{i,j}]_{i=1,\ldots,K,j=1,\ldots,C}, \\ \text{where } f_{i,j} = & \begin{cases} 1/\sqrt{|\mathbb{A}_j|} & \text{if } i \in \mathbb{A}_j \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{20}$$

Note that the columns in $\mathbf{F}$ are orthonormal to each other, which means $\mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{I}$. The problem is then relaxed to

$$\begin{aligned} \mathbf{F}^* = \arg \min_{\mathbf{F}\in\mathbb{R}^{K\times C}} & \operatorname{Tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F}) \\ \text{subject to } & \mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{I}, \end{aligned} \tag{21}$$

whose solution $\mathbf{F}^*$ can be obtained by selecting the first C eigenvectors (in the ascending order of eigenvalues) of $\mathbf{L}$ as its columns. In order to satisfy the constraints on $\mathbf{F}$ in the original optimization problem, *k-means* clustering [19] is then applied on the row vectors of $\mathbf{F}^*$, where each row corresponds to one user. After the clustering, we build the beamforming subgroup by picking only one user in each cluster. User k in cluster $\mathbb{A}_c$, for $c = 1, 2, \ldots, C$, is randomly selected with the probability p_k to form each beamforming subgroup. p_k can be determined by the channel state information, requirements of QoS or fairness measures, which should satisfy the constraint $\sum_{p_k\in\mathbb{A}_c} p_k = 1$. Power allocation for each user can also be performed using the water-filling scheme as discussed in Section III-A. A detailed description of OUG-SC is listed in Table 3.

Note that after spectral clustering, the number of elements in each cluster may not be the same. Those users that have fewer neighbors within their clusters are the ones with more favorable channels. That is, they would have more chances to be selected and assigned into more than one beamforming subgroups while the multiuser interference in these subgroups can be kept at a rather low level. As we can see, this approach is intrinsically an overlapping user grouping algorithm.

IV. CAPACITY AND COMPLEXITY ANALYSIS FOR OVERLAPPING USER GROUPING

In this section, we provide a detailed capacity and complexity analysis to illustrate the effectiveness of the proposed overlapping user grouping approaches.

TABLE 3. Overlapping user grouping based on spectral clustering.

Algorithm 2	$OUG\text{-}SC(\mathbb{S}, \mathbf{H})$
Input:	$\mathbb{S}$: The set of user indices to be grouped. $\mathbf{H}$: The channel matrix of all the users..
Output:	$\boldsymbol{S} = \{\mathbb{S}_1, \ldots, \mathbb{S}_G\}$: The sets of user indices i in each beamforming subgroup.
Initialization:	$\mathbf{W} = (w_{i,j})_{i,j=1,\ldots,K}$ $w_{i,j} = w_{j,i} = \arccos \frac{\lvert \boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{h}_j \boldsymbol{h}_j^{\mathrm{H}} \boldsymbol{h}_i \rvert}{\lVert \boldsymbol{h}_i \rVert \lVert \boldsymbol{h}_j \rVert}$ /* Generate the similarity matrix $\mathbf{W}$ using the generalized Fubini-Study distance */
Laplacian Matrix:	$d_i = \sum_{j=1}^{K} w_{i,j}, \quad \mathbf{D} = diag\{d_i\}$ $\mathbf{L} = \mathbf{D} - \mathbf{W}$ $\mathbf{L}_{rw} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W}$ $\mathbf{L}_{sym} = \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$ /* $\mathbf{L}_{rw}$ and $\mathbf{L}_{sym}$ are normalized Laplacians called *random walk* and *symmetric* matrix, respectively. */
Eigen-Decomposition:	$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_C \leftarrow \mathbf{L}_{sym\setminus rw}\boldsymbol{u} = \lambda \boldsymbol{u}$ $\mathbf{F} = (\boldsymbol{f}_1^T, \ldots, \boldsymbol{f}_C^T)^T = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_C) \in \mathbb{C}^{K \times C}$ /*$\mathbf{F}$ is formed by the eigenvectors of $\mathbf{L}$ as columns */
K-means:	$\mathbb{A}_1, \mathbb{A}_2, \ldots, \mathbb{A}_C \leftarrow$ K-means$(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_C)$ /* *K-means* clustering in respect of $\mathbf{F}$ */
Return:	$S_i \xleftarrow{p_k} \{k_1, \ldots, k_C \mid k_1 \in \mathbb{A}_1, \ldots, k_C \in \mathbb{A}_C\}$, for $i = 1, 2, \ldots, G$ /* Pick one user in each cluster to form subgroup */

A. CAPACITY ANALYSIS

In this subsection, we will show that the proposed overlapping user grouping approaches can achieve better performance than the non-overlapping user grouping approaches in terms of the sum capacity. The proof is broken down into two steps:

- First, we prove that the greedy user selection achieves higher sum capacity than that of random user picking.
- Second, we prove that the overlapping among multiple subgroups will further improve the sum capacity.

1) THE SUPERIORITY OF GREEDY USER SELECTION OVER RANDOM USER PICKING

We demonstrate the superiority of the greedy user selection by proving that when selecting a certain number l of users from the user set $\mathbb{S}$ of size K, the greedy user selection can achieve higher capacity than random user picking does. Let $\mathbf{H}_i$ denote the channel matrix of the current group $\mathbb{S}_i$, with the zero-forcing beamforming in (9), the achieved sum rate can be written as

$$\begin{aligned} R_i(\mathbf{H}_i, \boldsymbol{\beta}_i) &= \sum_{k \in \mathbb{S}_i} \log_2(1 + \beta_k \lVert \boldsymbol{h}_k \boldsymbol{b}_k \rVert) \\ &= \sum_{k \in \mathbb{S}_i} \log_2 \left(1 + \frac{\beta_k}{[(\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}})^{-1}]_{(k,k)}} \right), \end{aligned} \quad (22)$$

where $[\cdot]_{(k,k)}$ denotes the k-th diagonal element in the matrix. Since the optimal power allocation for random user picking is not necessarily the optimal one for the proposed grouping approach, it suffices to show the effectiveness by proving that $R_{grouping} \geq R_{random}$ under the optimal power allocation $\boldsymbol{\beta}_k$ for random user picking.

Following (22), for user k in $\mathbb{S}_i$, the effective channel gain is $1 \big/ [(\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}})^{-1}]_{(k,k)}$. Since the elements of $\mathbf{H}_i$ are i.i.d. complex Gaussian random variables with zero-mean and unit variance, it is known that $\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}}$ is central complex Wishart distributed [20], i.e. $\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}} \sim \mathcal{W}_{|\mathbb{S}_i|}^{C}(M, \mathbf{I}_M)$. Then $(\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}})^{-1}$ is complex inverse Wishart distributed with M degrees of freedom. According to [21], when the users in this subgroup are randomly selected, $1 \big/ [(\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}})^{-1}]_{(k,k)}$ is a chi-square distributed random variable with $2(M - |\mathbb{S}_i| + 1)$ degrees of freedom. If $l = |\mathbb{S}_i| = M$ users are to be selected, the expectation of the channel gain of user k is

$$E\left[\frac{1}{[(\mathbf{H}_i \mathbf{H}_i^{\mathrm{H}})^{-1}]_{(k,k)}} \right] = E\left[\chi^2(2(M - |\mathbb{S}_i| + 1)) \right] = 2. \quad (23)$$

So that for each group, the capacity of the random user picking algorithm can be written as

$$R_{random} = \sum_{k \in \mathbb{S}_i} \log_2 \left(1 + 2\beta_k \right). \quad (24)$$

In [4], Wang *et al.* prove that the greedy user selection in the proposed user grouping algorithm is a special case of the *asymptotically optimal G-greedy* (AOG) algorithm and show that the effective channel gain for user k has the lower bound

$$\gamma_{k_{(j)}} \geq \frac{(a_j + 2) b_j}{a_j (1 + a_j)^{2(M-j)} + 2} \lVert \boldsymbol{h}_k \rVert, \quad (25)$$

where j is the iteration number in AOG algorithm, a_j and b_j are positive constant parameters which satisfy the following conditions:

1) $\lim_{K \to +\infty} a_j = 0$ and $\lim_{K \to +\infty} a_j^2 \log K \geq c > 0$ for some positive c,
2) $\lim_{K \to +\infty} b_j = 1$ and $\lim_{K \to +\infty} K^{1-b_j} = +\infty$.

As the channel vectors of different user are independent and converge to complex Gaussian vectors with distribution $\mathcal{CN}(0, \boldsymbol{I}_m)$, we have

$$\lim_{K \to +\infty} \left[E[R_{grouping}] - \sum_{k \in \mathbb{S}_i} \log_2 \Big(1 + \beta_k \log K \Big) \right] = 0. \quad (26)$$

Combine (24) and (26), we can conclude that, *when K, the size of the whole user set, is large enough*, even without overlapping among subgroups, *the sum rate achieved by the proposed grouping approach is greater than that achieved by random user picking*.

2) THE IMPROVEMENT INTRODUCED BY OVERLAPPING USER GROUPING

We next show that the overlapping user grouping can further improve the sum capacity over the non-overlapping approaches. Let $\mathbb{S}_1, \ldots, \mathbb{S}_G$ be the subgroups obtained by the greedy based non-overlapping user grouping, and $\mathbb{S}_{1,o}, \ldots, \mathbb{S}_{G,o}$ the subgroups obtained using the overlapping user grouping as described in Section III-A, we now prove that

$$sum\_rate(\mathbf{H}, \mathbb{S}_{i,o}, \boldsymbol{\beta}_{i,o}) \geq sum\_rate(\mathbf{H}, \mathbb{S}_i, \beta_i). \quad (27)$$

where *sum_rate*$(\mathbf{H}, \mathbb{S}_{i,o}, \boldsymbol{\beta}_{i,o})$ means the overall capacity that can be achieved for all the users in subgroup $\mathbb{S}_{i,o}$ with power allocation $\boldsymbol{\beta}_{i,o}$, and *sum_rate*$(\mathbf{H}, \mathbb{S}_i, \beta_i)$ is defined similarly.

Recall that $\mathbb{S}_{i,o}$ is obtained from $\mathbb{S}_i$ by re-searching over the set $\mathbb{C}_{i,o} = \bigcup_{j=1}^{i-1} \mathbb{S}_j$, the search is carried out through an iterative procedure as before. In each step of overlapping searching, for subgroup $\mathbb{S}_i$ and any user $k \in \bigcup_{j=1}^{i-1} \mathbb{S}_j$, it will be added to $\mathbb{S}_i$ to obtain $\mathbb{S}_{i,o}$, only if the overall sum rate achieved by adding user k is larger than that without it. The iteration goes on until all the users in $\mathbb{C}_{i,o} = \bigcup_{j=1}^{i-1} \mathbb{S}_j$ are explored. It then follows that $\mathbb{S}_i \subset \mathbb{S}_{i,o}$ and *sum_rate*$(\mathbf{H}, \mathbb{S}_{i,o}, \boldsymbol{\beta}_{i,o}) \geq$ *sum_rate*$(\mathbf{H}, \mathbb{S}_i, \beta_i)$.

More specifically, suppose we have grouped K_1 users and we are selecting users for the current subgroup $\mathbb{S}_i$ from the remaining user set $\mathbb{C}_i$ of size K_2. By using the greedy user selection algorithm, we sequentially select the users from $\mathbb{C}_i$ which can maximize the sum-capacity for the current group. The selection will stop if no capacity improvement can be achieved by adding any additional user into the current group. In [4], it has been proved that the user selection process will stop with a probability upper bounded by

$$Pr\{stop\} \leq e^{-(1-e^{-c})^{i-1} b_i^{M-i} (\log K)^{M-i} K^{1-b_i}}. \tag{28}$$

When we introduce the overlapping among subgroups by extending the searching space from $\mathbb{C}_i$ to $\mathbb{C}_{i,o}$, we actually increase K in (28), the size of searching space, from K_1 to $K_1 + K_2$, so that the upper bound of the stop searching probability will decrease. It means that $\mathbb{S}_{i,o}$ can select more users with higher probability and is always larger than or equal to the corresponding subgroup $\mathbb{S}_i$. Furthermore, by (26), for sufficiently large K, especially in IoT oriented massive MIMO systems, the achievable capacity of user grouping algorithm can be approximated by

$$E[R_{grouping}] \approx \sum_{k \in \mathbb{S}_g} \log_2 \left(1 + \beta_k \log K\right). \tag{29}$$

Clearly, based on (29), when extending the searching space, it always results in larger achievable capacity.

It can then be concluded that, by exploiting the orthogonality of users in massive MIMO and assigning them into multiple subgroups simultaneously, the overlapping among subgroups introduces new degrees of freedom while the multi-user interference is kept within a lower level. As a result, the overall capacity is further increased with overlapping user grouping.

B. COMPLEXITY ANALYSIS

Computational complexity is also a primary concern in IoT systems due to the very limited energy budget of IoT devices. In this subsection, we will analyze the computational complexity of the proposed overlapping user grouping approaches.

In each iteration of the OUG-Greedy algorithm, the system needs to select $C \leq M$ out of K users which yield the the largest sum rate. With the simplified algorithm proposed in [3], it evaluates the 2-norm of matrix $\boldsymbol{h}_K \mathbf{P}_{k,i} \boldsymbol{h}_K^{\mathrm{H}}$ for $(K - C + 1)$ times in a single iteration. Each evaluation of $\boldsymbol{h}_K \mathbf{P}_{k,i} \boldsymbol{h}_K^{\mathrm{H}}$ involves a vector-matrix multiplication with $1 \times M$ vector and $M \times M$ matrix. The complexity of this step is $\mathcal{O}(M^2)$. Repeating this over $\mathcal{O}(K)$ users in at most $\mathcal{O}(M)$ steps (since in each group, it can select at most M users, i.e. $C \leq M$), hence, the computational complexity for selecting one subgroup is limited by $\mathcal{O}(KM^3)$. To accommodate all the users, OUG-Greedy algorithm has to assign at least $\lceil K/M \rceil$ beamforming subgroups. Therefore, the computational complexity of OUG-Greedy is at least $\mathcal{O}(K^2M^2)$, which is of the same order as that of the SWF with only a linear increase.

For OUG-SC user grouping approach, the computational complexity roughly equals to that of the spectral clustering. In the step of the graph construction, the spectral clustering algorithm needs to calculate the similarity measure $w_{i,j}$ for each pair of users (actually, the computation could be reduced by half since the weighted adjacency matrix is symmetric), which has the complexity on the order of $\mathcal{O}(K^2)$. Then the algorithm moves on to the eigen-decomposition of *Laplacian matrix* with the complexity on the order of $\mathcal{O}(K^3)$ [22], followed by the *K-means* clustering based on the *Lloyd's algorithm*, which is of linear complexity $\mathcal{O}(K)$ [19]. From the analysis above, the computational complexity of OUG-SC algorithm is on the order of $\mathcal{O}(K^3 + K^2 + K) = \mathcal{O}(K^3)$, which is at least one order of magnitude lower than that of the greedy based user grouping methods. The computational complexities of these user grouping algorithms are summarized in Table 4.

TABLE 4. Comparison of computational complexity.

User Grouping Algorithm	Computational Complexity
Capacity Based ZFS	$O(KM^5)$
SWF	$O(KM^3)$
OUG-Greedy	$O(K^2M^2)$
OUG-SC	$O(K^3)$

V. SIMULATION RESULTS

In this section, the numerical results are presented. To evaluate the effectiveness of the proposed approaches, we compare the achieved sum rate of the proposed OUG-Greedy algorithm and OUG-SC algorithm with that of the Sequential Water-Filling (SWF) algorithm proposed in [4] and the Random User Picking (RUP) method discussed in [2]. For system set-up, we adopt the one-ring MIMO channel model [10] with the NeSH shadowing model [12], and the elements of the channel correlation matrix can be derived from equations (3) and (5). System parameters used in the simulation are listed in Table 5. The simulation of each user grouping algorithm is performed by averaging over 500 random channel realizations generated with the Monte-Carlo method.

Assuming that the base station in a massive MIMO system has 100 omni-directional antennas, serving a total number of

TABLE 5. System parameters in simulations.

parameters	value	parameters	value
θ	$[-180°, 180°]$	M	100
Δ	$[5°, 15°]$	K	300
d_{cor}	20 (m)	D	0.5

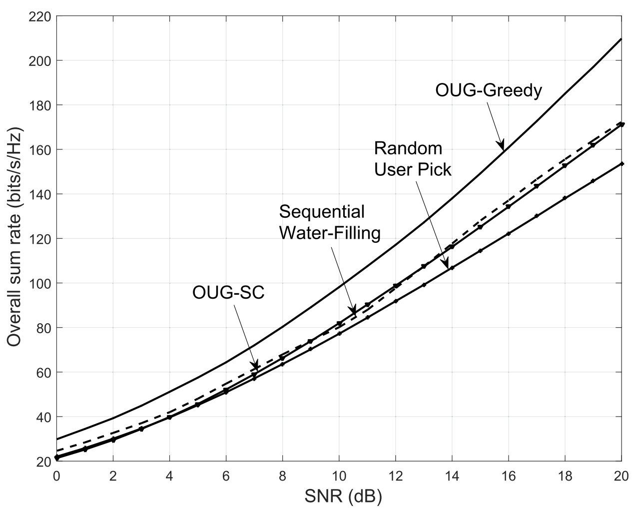

FIGURE 3. Sum rate comparison under uncorrelated MIMO channels.

300 IoT users, we generate the user families according to geometric modeling. First, we evaluate the proposed overlapping user grouping algorithms in the scenario of rich-scattering Rayleigh environment where users are mutually independent. Fig. 3 shows the sum capacity comparison of the proposed algorithms and the existing algorithms. It can be seen that the proposed OUG-Greedy outperforms the other three algorithms, and OUG-SC achieves approximately the same capacity as SWF [4] algorithm with a lower complexity. This is because that when all the channels are uncorrelated, the directions of channel vectors in Euclidean space obey the uniform distribution. In this case, the similarity measures for any two-user pairs are independent, and it may be difficult for the spectral clustering algorithm to partition users into clusters. It is also observed that the random user picking has a relatively high capacity since the angle between two channel vectors may be rather large due to the channel mutual independence. In such case, the capacity improvement of the proposed approaches is mainly achieved by the overlapping among subgroups.

Next, we consider the scenario with correlated shadow fading and evaluate the performance of the proposed approaches. It can be seen in Fig. 4 that both the proposed OUG-Greedy and OUG-SC algorithms achieve significant improvement on sum capacity over the existing RUP [2] and SWF [4] algorithms. This is because that the random user picking may assign users with high correlation into the same subgroup, in which case those users would generate strong interference to each other. In addition, the proposed user grouping

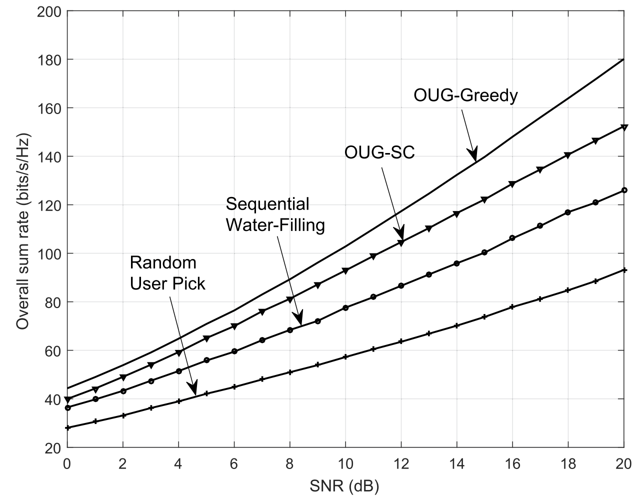

FIGURE 4. Sum rate comparison under correlated MIMO channels.

approaches allow overlapping between user subgroups, enable users with favorable propagation to be served in more than one beamforming subgroup without generating much interference to the concurrent users, which greatly enhances the sum capacity. We can also note that the OUG-Greedy algorithm outperforms the OUG-SC algorithm. This is because that the OUG-Greedy optimizes the sum capacity directly while the OUG-SC uses the indirect metric of the similarity measure. However, when the antenna number is large at the base station, the OUG-SC has a much lower computational complexity than the OUG-Greedy, and achieves a trade-off between complexity and performance.

VI. CONCLUSIONS

In this paper, we propose two new overlapping user grouping approaches by exploiting the favorable propagation property in IoT oriented massive MIMO systems. First, we propose a greedy search based user grouping method that allows overlapping between different beamforming subgroups. In this algorithm, system selects the users with more favorable channel conditions and assign them into multiple subgroups. As a result, OUG-Greedy algorithm improves the system capacity with the same order of computational complexity as the existing non-overlapping algorithms. To further reduce the complexity, we introduce a new channel similarity measure and develop a low complexity user grouping method based on the spectral clustering algorithm in machine learning. Both the theoretical and numerical results demonstrate that, with the proposed overlapping user grouping approaches, the IoT oriented massive MIMO system can achieve much higher network capacity, and ensures that at any given time, each IoT device will be served in at least one subgroup.

REFERENCES

[1] Z. Ni and D. Li, "Effect of fading correlation on capacity of distributed MIMO," in *Proc. IEEE 15th Int. Symp. Pers., Indoor Mobile Radio Commun.*, vol. 3. Sep. 2004, pp. 1637–1641.

[2] Q. Yang *et al.*, "BigStation: Enabling scalable real-time signal processingin large MU-MIMO systems," in *Proc. ACM SIGCOMM*, 2013, pp. 399–410.
[3] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.
[4] J. Wang, D. J. Love, and M. D. Zoltowski, "User selection with zero-forcing beamforming achieves the asymptotically optimal sum rate," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3713–3726, Aug. 2008.
[5] S. Huang, H. Yin, J. Wu, and V. C. M. Leung, "User selection for multiuser MIMO downlink with zero-forcing beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3084–3097, Sep. 2013.
[6] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
[7] W.-L. Shen, K. C.-J. Lin, M.-S. Chen, and K. Tan, "SIEVE: Scalable user grouping for large MU-MIMO systems," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2015, pp. 1975–1983.
[8] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2014, pp. 76–80.
[9] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, Jr., "Compressed channel feedback for correlated massive MIMO systems," *IEEE/KICS J. Commun. Netw.*, vol. 18, no. 1, pp. 95–104, Feb. 2016.
[10] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
[11] Y. Xu, G. Yue, N. Prasad, S. Rangarajan, and S. Mao, "User grouping and scheduling for large scale MIMO systems with two-stage precoding," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 5197–5202.
[12] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4024–4036, Aug. 2009.
[13] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.
[14] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
[15] S. Sigdel and W. A. Krzymien, "Efficient user selection and ordering algorithms for successive zero-forcing precoding for multiuser MIMO downlink," in *Proc. IEEE 69th Veh. Technol. Conf.*, Apr. 2009, pp. 1–6.
[16] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
[17] M. B. Blaschko and C. H. Lampert, "Correlational spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[18] J. Nam, Y.-J. Ko, and J. Ha, "User grouping of two-stage MU-MIMO precoding for clustered user geometry," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1458–1461, Aug. 2015.
[19] G. Hamerly, "Making *K*-means even faster," in *Proc. SIAM Int. Conf. Data Mining*, 2010, pp. 130–140.
[20] P. Graczyk, G. Letac, and H. Massam, "The complex wishart distribution and the symmetric group," *Ann. Statist.*, vol. 31, no. 1, pp. 287–309, 2003.
[21] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1740–1751, Feb. 1994.
[22] A. Banerjee, C. Krumpelman, J. Ghosh, S. Basu, and R. J. Mooney, "Model-based overlapping clustering," in *Proc. ACM 11th Int. Conf. Knowl. Discovery Data Mining*, 2005, pp. 532–537.

RUN TIAN received the B.S. and M.S. degrees in communication engineering from the Harbin Institute of Technology in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering. From 2015 to 2016, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Michigan State University, USA. His research interests include cognitive radio technology, interference alignment, and resource optimization in MIMO networks.

YUAN LIANG (S'16) received the B.S. degree in electrical engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2012, and the M.S. degree in electrical engineering from Southeast University, Nanjing, China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA. His research interests include architecture design, performance analysis, and resource allocation in heterogeneous wireless networks.

XUEZHI TAN (M'10) received the M.S. and Ph.D. degrees from the Harbin Institute of Technology (HIT), Harbin, China, in 1986 and 2005, respectively. He is currently a Professor with the School of Electrical and Information Engineering, HIT, where he is also the Vice Director of the Communication Research Center. His major research interests focus on data communication, broadband multimedia trunk communication, and cognitive radio networks. He is a Senior Member of the China Institute of Communications, a Senior Member of the China Institute of Electronics, the Executive Director of the Heilongjiang Province Institute of Electronics, the Vice President of the Software Association of Heilongjiang Province, and the Director of the China Radio Association.

TONGTONG LI (S'97–M'00–SM'08) received the Ph.D. degree in electrical engineering from Auburn University, Auburn, AL, USA, in 2000. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, Michigan State University. Her research interests include communication system design and networking, and statistical signal processing, with applications in computational neuroscience. She was a recipient of the National Science Foundation CAREER Award.