User Scheduling for Heterogeneous Multiuser MIMO Systems: A Subspace Viewpoint

Xinping Yi and Edward K. S. Au, *Member, IEEE*

***Abstract*—In downlink multiuser multiple-input multiple-output systems, users are practically heterogeneous. However, many existing user scheduling algorithms are designed with an implicit assumption that users are homogeneous. In this paper, we revisit the problem by exploring the characteristics of heterogeneous users from a subspace viewpoint. With an objective of minimizing interference non-orthogonality among users, three new angular-based user scheduling criteria are proposed. While the first criterion is heuristically determined by identifying the incapability of largest principal angle to characterize subspace correlation and hence the interference non-orthogonality between users, the second and third ones are derived by using, respectively, the sum rate capacity bounds with block diagonalization and the change in capacity by adding a new user into an existing user subset. Aiming at capturing fairness among heterogeneous users while maintaining multiuser diversity gain, two new hybrid user scheduling algorithms are proposed whose computational complexities are only linearly proportional to the number of users. We show by simulations the effectiveness of our proposed user scheduling criteria and algorithms with respect to those commonly used in homogeneous environment.**

***Index Terms*—Multiuser, MIMO, Principal Angles, Subspace, User Scheduling.**

I. Introduction

For Multiuser Multiple-Input Multiple-Output (MU-MIMO) systems, a great number of low-complexity linear precoding algorithms have been proposed to improve sum rate capacity with reasonable computational complexity. Block Diagonalization (BD) is one of the popular choices due to its capability of approaching the capacity and its ease in practical implementation [1]. In an overloaded system that supports a very large number of users, user scheduling is necessary as base station (BS) cannot usually serve such a large number of users simultaneously because of the following two key reasons. First, there are far more users to be supported than the number of transmit antennas available at BS, which violates the dimensionality constraint of BD [1]. Second, interference non-orthogonality among users always exists [2]. In other words, instantaneous channels among users are non-orthogonal to one another, which result in mutual inter-user interference.

There are two common types of scheduling algorithms: user selection and user grouping. For the former whose objective is to select a subset of users for scheduling, it is natural to find an optimal subset by using exhaustive search but it is very computationally demanding even for moderate number of users. In this context, a large number of sub-optimal yet simplified algorithms has been proposed whose fundamental idea is to maximize system performance according to various user selection criteria [2]–[4]. On the other hand, user grouping algorithms take fairness into account and schedule all users to be served over consecutive scheduling units. In particular, all users are divided into a number of groups by certain criteria [5], [6], whose aim is to maximize system performance while minimizing spatial correlation among users per group.

For the above-mentioned user scheduling algorithms, an effective performance metric is required in selecting either an optimal subset of users or an optimal scheduling arrangement. There is a large body of literature focused on uncorrelated downlink MU-MIMO systems with homogeneous users (e.g., [7]–[9]). For systems with heterogeneous users, the task of designing an efficient user scheduling metric becomes more challenging because there are more system parameters to consider. Though there are some recent works that consider scheduling strategies for users with different received SNRs [10] and different number of receive antennas [11], there is a lack of works addressing the combined problem based on the scheduling criteria. Naturally, an interesting question arouses in mind is, *whether those heuristical scheduling criteria/metrics employed in homogeneous MIMO broadcast channels are still applicable in heterogeneous environment*. In this paper, we try to answer this question by studying users' channel characteristics in a subspace approach and designing effective user scheduling metrics from a geometric viewpoint. The main contributions are summarized as follows.

- We take into account the characteristics of principal angles between channels of heterogeneous users and propose three angular-based scheduling criteria that achieve larger sum rare capacity than the existing ones.
- We propose two hybrid user scheduling algorithms that takes into account some key features of user grouping and selection algorithms, i.e., to capture fairness among users and to maximize the system performance in a greedy manner. No brute-force search is required and the computational complexities are only linearly, rather than exponentially, proportional to the total number of users.

Notation: Matrices and vectors are represented as uppercase and lowercase letters, respectively. Transpose and conjugate transpose of a matrix are denoted as $(\cdot)^T$ and $(\cdot)^H$, respectively. Further, we reserve $diag\{\cdot\}$ for an diagonal matrix, while $\det(\cdot)$, $rank(\cdot)$, $tr(\cdot)$, $\Lambda(\cdot)$, $\lambda_i(\cdot)$, and $||\cdot||_F$ represent the determinant, rank, trace, diagonal part, the i-th singular

X. Yi is with the Mobile Communications Dept., EURECOM, 06560 Sophia Antipolis, France (email: yix@eurecom.fr) and E. Au is with Huawei Technologies, Shenzhen, China (email: edward.au@huawei.com).

value, and Frobenius norm of a matrix.

II. System Model

Consider a downlink MU-MIMO system with M_T transmit antennas at BS and K heterogeneous users[1] that are equipped with M_{R_k} receive antennas at the k-th user. We consider overloaded scenarios (i.e., $M_T \ll \sum_{k=1}^{K} M_{R_k}$) in which the BS cannot serve all users simultaneously. For this reason, user scheduling is necessary to serve either a subset of users at one time or all users once over an entire scheduling period.

Consider a subset of users $\mathcal{T}$ that has been scheduled for transmission. Denote $\mathbf{x}_k$ as the transmit signal of user-k in the group (i.e., $k \in \mathcal{T}$). Its receive signal $\mathbf{y}_k$ is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{F}_k \mathbf{x}_k + \sum_{l \in \mathcal{T}, l \neq k} \mathbf{H}_k \mathbf{F}_l \mathbf{x}_l + \mathbf{n}_k,$$

where $\mathbf{H}_k = \sqrt{\rho_k} \bar{\mathbf{H}}_k \in \mathcal{C}^{M_{R_k} \times M_T}$ is the channel matrix between the BS and user-k with ρ_k being the average received power of the k-th user and $\bar{\mathbf{H}}_k$ is an arbitrary matrix that depends on the channel model employed. Further, $\mathbf{F}_k$ is a precoding matrix of the k-th user. For practical consideration, we adopt BD [1] as the linear precoder. $\mathbf{n}_k$ is the additive white Gaussian noise at the receive antennas of user-k and is assumed to be zero-mean independent and identically distributed (i.i.d.) Gaussian vector with variance σ_n^2.

Assume all channels are time-invariant during the scheduling period and the BS has perfect channel state information for all users. In order to track the influence of received SNR, we follow [10] and consider it as a function of the distance between the BS and a user, path loss exponent, and average transmit power per antenna.

III. Scheduling Criteria for Heterogeneous Users

A. Review on Metric Choice for Homogeneous Users

Considering channel matrices as the subspace spanned by their column vectors, mutual interference across users can be represented as the correlation of the corresponding subspaces. Some recent works have been attempted to measure the level of subspace correlation in either angular or subspace domains and utilize them as user scheduling metrics. In the following, we will review the three common ones.

1) Largest Principal Angle: To facilitate our subsequent discussion, we first review the definition of principal angle.

Definition 1 (Principal Angle [12], [13]): For any two nonzero subspaces $\mathcal{U}_k, \mathcal{V}_j \subseteq \mathcal{C}^n$, the principal angles between $\mathcal{U}_{k,1} = \mathcal{U}_k$ and $\mathcal{V}_{j,1} = \mathcal{V}_j$ are recursively defined to be the numbers $0 \leq \theta_{k,j,i} \leq \pi/2$ such that

$$\begin{aligned} \cos\theta_{k,j,i} &= \max_{\{\mathbf{u}_k \in \mathcal{U}_{k,i}, \mathbf{v}_j \in \mathcal{V}_{j,i}, \|\mathbf{u}_k\|_2 = \|\mathbf{v}_j\|_2 = 1\}} \mathbf{v}_j^H \mathbf{u}_k \\ &= \mathbf{v}_{j,i}^H \mathbf{u}_{k,i}, \quad i = 1, \cdots, p, \end{aligned}$$

where $p = \min\{dim(\mathcal{U}_k), dim(\mathcal{V}_j)\}$, $\mathbf{u}_{k,i}$ and $\mathbf{v}_{j,i}$ are the vectors that construct the i-th principal angle $\theta_{k,j,i}$, $\|\mathbf{u}_{k,i}\|_2 = \|\mathbf{v}_{j,i}\|_2 = 1, \mathcal{U}_{k,i} = \mathcal{U}_{k,i-1} \cap \mathbf{u}_{k,i-1}^{\perp}$ and $\mathcal{V}_{j,i} = \mathcal{V}_{j,i-1} \cap \mathbf{v}_{j,i-1}^{\perp}$. Further, $\theta_{k,j,min} = \theta_{k,j,1} \leq \cdots \leq \theta_{k,j,p} \leq \theta_{k,j,max}$ and $\theta_{k,j,p} = \theta_{k,j,max}$ when $dim(\mathcal{U}_k) = dim(\mathcal{V}_j)$. □

From the definition, the cosine of the principal angle is the inner product of two vectors, and the minimal principal angle represents the largest inner product of any vectors in the two subspaces. If there exists intersection between two subspaces, the minimal principal angle would be zero.

In [6], a user scheduling algorithm is proposed that aims at minimizing spatial correlation across users by arranging users with low inter-user spatial correlation into one group. In order to determine the impact of spatial correlation of the k-th user due to the remaining users, the largest principal angle between the orthogonal basis of the row spaces of user-k's transmission channel $\mathbf{H}_k$ and its interference channel $\tilde{\mathbf{H}}_k = \left[\mathbf{H}_1^T, \cdots, \mathbf{H}_{k-1}^T, \mathbf{H}_{k+1}^T, \cdots, \mathbf{H}_K^T\right]^T$ is utilized and the resulting largest principal angle of each user is used as a scheduling metric. In Section IV, we will discuss why this algorithm motivates us to propose two reduced-complexity greedy-based hybrid user scheduling algorithms.

2) Subspace Collinearity: Subspace collinearity is a criterion that reflects the similarity of two matrix subspaces and it can be used for characterizing users' spatial separability [14]. In general, given two matrices $\mathbf{M}_A$ and $\mathbf{M}_B$, their collinearity can be represented as [15]

$$col(\mathbf{M}_A, \mathbf{M}_B) = \frac{abs\left(tr(\mathbf{M}_A \mathbf{M}_B^H)\right)}{\|\mathbf{M}_A\|_F \|\mathbf{M}_B\|_F}.$$

It is clear that the smaller the collinearity is, the less similarity of the two matrix subspaces.

3) Chordal Distance: Chordal distance is commonly used in limited feedback systems for codebook design but it has also been recently considered as a user scheduling criterion [16]. As referred to [17], the chordal distance between two subspaces $\mathcal{U}_k$ and $\mathcal{V}_j$ with dimensions p and q is defined in terms of the user's principal angles as follows

$$d_c(\mathcal{U}_k, \mathcal{V}_j) = \frac{1}{\sqrt{2}} \|\mathbf{P}_{\mathcal{U}_k} - \mathbf{P}_{\mathcal{V}_j}\|_F = \sqrt{\sum_{i=1}^{\min\{p,q\}} \sin^2\theta_{k,j,i}},$$

where $\mathbf{P}_{\mathcal{U}_k}$ and $\mathbf{P}_{\mathcal{V}_j}$ are the projection matrices of $\mathcal{U}_k$ and $\mathcal{V}_j$, respectively.

B. Proposed User Scheduling Criteria

In downlink MU-MIMO channel with heterogeneous users, principal angles between two subspaces of dimensions[2] p and q possess the following specific characteristics, namely

$$\begin{aligned} 0 &= \underbrace{\theta_{k,j,1} = \cdots = \theta_{k,j,m}}_{Part\ I} < \underbrace{\theta_{k,j,m+1} \leq \cdots \leq \theta_{k,j,n}}_{Part\ II} \\ &< \underbrace{\theta_{k,j,n+1} = \cdots = \theta_{k,j,p}}_{Part\ III} = \frac{\pi}{2}, \end{aligned} \tag{1}$$

where $\theta_{k,j,1}$ and $\theta_{k,j,p}$ are the minimum and maximum principal angles, respectively. These three parts represent different physical meanings on subspace correlation (that models the mutual interference among users). For Part I that consists of

[1]Unlike homogeneous counterpart, heterogeneous users are of different antenna configurations and/or experience different channel environments.

[2]Without loss of generality, we assume $p \leq q$.

m zero principal angles, they represent m overlapped and fully-correlated basis of the two subspaces. As for the second part that is composed of $n-m$ principal angles whose values lie between 0 and $\pi/2$, the corresponding basis of the two subspaces are non-overlapped but non-orthogonal with one another, which result in partial subspace correlation. Regarding the third part that contains $q-n$ principal angles of $\pi/2$, it means there are $q-n$ orthogonal principal angles.

In order to understand the characteristics of (1), we consider an example in which there are two heterogeneous users k and j with $p = M_{R_k}$ and $q = M_{R_j}$ receive antennas, respectively, and their instantaneous channel matrices are non-orthogonal to one another because of, e.g., their close proximity. Denote $\mathcal{U}_k$ and $\mathcal{V}_j$ as the two subspaces of dimensions p and q that are spanned by the columns of their channel matrices. Assuming that the overlapped dimension of the channel subspaces is m, we have $dim(\mathcal{U}_k \cap \mathcal{V}_j) = m$ and hence m zero principal angles (c.f., Part I in (1)). For the non-overlapped counterpart of dimension $p-m$, there exists $p-n$ mutually orthogonal components that correspond to $p-n$ largest principal angle of $\pi/2$ (c.f., Part III), while the remaining $n-m$ ones are non-orthogonal with one another and they represent those principal angles with values between 0 and $\pi/2$ (c.f., Part II).

As referred to the characteristics above, subspace correlation is reduced if the principal angles are as large as possible. For user scheduling algorithms that aim at minimizing subspace correlation and hence interference non-orthogonality among users, it means that the user selection/grouping criteria should be designed in such a way that users with a smaller dimension of Part I and a larger dimension of Part III can be served simultaneously. This important observation leads us to understand that single-dimensional information, e.g., the largest/smallest principal angles, is sometimes not enough to characterize the correlation between two subspaces. For example, if $\mathcal{U}_k$ and $\mathcal{V}_j$ are subspaces of unequal dimensions that have a nontrivial intersection, then $\theta_{k,j,min} = 0$ and $\theta_{k,j,max} = \pi/2$, but neither of them might convey the desired information on whether these subspaces are highly correlated. Similarly, though subspace collinearity reflects the similarity of two subspaces, it is an indirect measure because of its heuristic reflection of the orthogonality of the channel matrices. Likewise, chordal distance requires the compared subspaces $\mathcal{U}_k$ and $\mathcal{V}_k$ to be of the same dimension. If the subspaces are of different dimensions, the lower dimension is usually adopted, which may result in an inaccurate measure for systems with heterogeneous users. Therefore, it would be interesting to consider not only the largest and smallest principal angles, but also those "intermediate angles". In the following, we propose three user scheduling criteria that take into account all principal angles as given in (1).

1) Geometrical Angle: Let $\mathcal{U}_k = span\{\mathbf{u}_{k,1}, \cdots, \mathbf{u}_{k,p}\}$ and $\mathcal{V}_j = span\{\mathbf{v}_{j,1}, \cdots, \mathbf{v}_{j,q}\}$ be two subspaces with $1 \leq p \leq q$. Geometrical angle, i.e., the angle $\psi_{k,j} = \measuredangle(\mathcal{U}_k, \mathcal{V}_j)$ between the two subspaces, is defined as [18]

$$\cos^2 \psi_{k,j} = \prod_{i=1}^{p} \cos^2 \theta_{k,j,i},$$

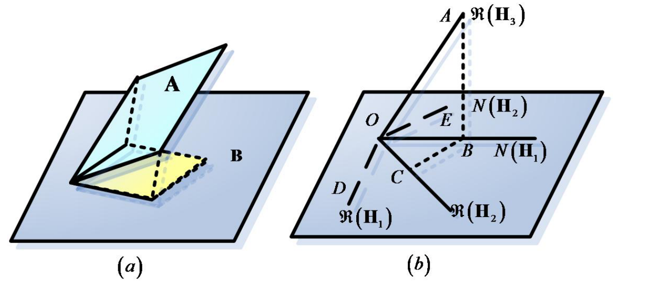

Fig. 1. Geometrical illustration of (a) geometrical angle; and (b) the loss in sum rate capacity due to a new incoming user.

where $\mathbf{u}_{k,i}$, $\mathbf{v}_{j,i}$ and $\theta_{k,j,i}$ are defined in Definition 1. Given two users k and j with channel matrices $\mathbf{H}_k$ and $\mathbf{H}_j$, without loss of generality, we assume $p = rank(\mathbf{H}_k) \leq rank(\mathbf{H}_j) = q$. Geometrical angle can be alternatively defined as [18]

$$\cos^2 \psi_{k,j} = \frac{\det(\mathbf{M}_{k,j}\mathbf{M}_{k,j}^H)}{\det(\mathbf{H}_k\mathbf{H}_k^H)}, \quad (2)$$

where $\mathbf{M}_{k,j} = \mathbf{H}_k\mathbf{H}_j^H$ is a cross-correlation matrix that is represented in inner product form.

In general, the value of $\cos^2 \psi_{k,j}$ represents the ratio between the volume of the parallelepiped spanned by the projection of the basis vectors of the lower dimension subspace on the higher dimension subspace and the volume of the parallelepiped spanned by the basis vectors of the lower dimension subspace. Fig. 1(a) shows a two-dimensional example. By definition, geometrical angle refers to the ratio of the area of the projection of plane-A onto plane-B to that of plane-A itself. It is clear that $\cos^2 \psi_{k,j}$ is larger (or the angle $\psi_{k,j}$ is smaller) when the planes are closer, and vice versa.

In our case, the level of subspace correlation between two users is characterized by the degree of overlapping between the corresponding channel subspaces. In particular, if the correlation is severe, the corresponding channel subspaces get closer to each other and hence the geometrical angle is smaller. In other words, there are more principal angles of zeros and less principal angles of $\pi/2$ in (1). Aiming at minimizing subspace correlation and hence interference non-orthogonality among the scheduled users, an effective user scheduling criterion should select for users with larger geometrical angles $\psi_{k,j}$ or equivalently, a smaller value of $\cos^2 \psi_{k,j}$, i.e.,

$$\mathcal{M}(\mathbf{H}_k, \mathbf{H}_j) = \arg \min_{k,j \in \mathcal{C}} \cos^2 \psi_{k,j}, \quad (3)$$

where $\mathcal{C}$ is the candidate user pool. The metric $\mathcal{M}(\mathbf{H}_k, \mathbf{H}_j)$ in (3) is introduced to denote the correlation of $\mathbf{H}_k$ and $\mathbf{H}_j$, and is used as a scheduling criterion for user grouping.

2) Grouping-Oriented Criterion: Recall from our discussion in (1) that users with a smaller dimension of correlated basis (i.e., Part I) and a larger dimension of orthogonal basis (i.e., Part III) are preferred to be served together. Therefore, it is important for a user grouping algorithm to take into account subspace correlation among users so as to minimize interference orthogonality on each group while maximizing the sum rate capacity. In the following, we re-express the sum rate capacity with BD in terms of principal angles and derive a scheduling metric that satisfies these two objectives.

Theorem 1 (Sum rate capacity bounds of BD): For a K-heterogeneous user downlink MU-MIMO system, the sum rate capacity with BD is bounded as

$$\sum_{k=1}^{K}\log_2\left(1+\frac{\rho_k\lambda_{k,min}^2}{\sigma_n^2}\sum_{i=1}^{M_{R_k}}\sin^2\theta_{k,\tilde{k},i}\right)\leq C_{sum}$$
$$\leq\sum_{k=1}^{K}M_{R_k}\log_2\left(1+\frac{\rho_k\lambda_{k,max}^2}{M_{R_k}\sigma_n^2}\sum_{i=1}^{M_{R_k}}\sin^2\theta_{k,\tilde{k},i}\right),\quad(4)$$

where $\theta_{k,\tilde{k},i}$ is the i-th principal angle between the range space of the k-th user's channel $\bar{\mathbf{H}}_k$ and its interference channel $\tilde{\mathbf{H}}_k$. Moreover, $\lambda_{k,min}$ and $\lambda_{k,max}$ are the minimum and maximum singular values of $\bar{\mathbf{H}}_k$. □

Proof 1: Please refer to Appendix A for details.

It is clear from Theorem 1 that the sum rate capacity of the k-th user is increased monotonically with its $\sum_{i=1}^{M_{R_k}}\sin^2\theta_{k,\tilde{k},i}$ and the performance metric can be written as

$$\mathcal{M}(\mathbf{H}_k,\tilde{\mathbf{H}}_k)\quad=\quad\arg\max_{k\in\mathcal{C}}\sum_{i=1}^{M_{R_k}}\sin^2\theta_{k,\tilde{k},i}.$$

This metric reflects the correlation of channel matrices between user-k and the other users that intend to group together. In order to take into account those zero principal angles corresponding to the overlapped subspaces (c.f., Part I in (1)), however, we prefer to use the cosine function because the sine of zero principal angle is equal to zero. Hence, our grouping-oriented scheduling criterion is embodied as

$$\mathcal{M}(\mathbf{H}_k,\tilde{\mathbf{H}}_k)\quad=\quad\arg\min_{k\in\mathcal{C}}\sum_{i=1}^{M_{R_k}}\cos^2\theta_{k,\tilde{k},i}.\qquad(5)$$

Although the bound (5) is not tight enough, it is sufficient to determine the relationship between one user and the other group members as will be shown by simulations in Section V.

3) Selection-Oriented Criterion: While user grouping algorithm serves all users once over an entire scheduling period, user selection algorithm serves only a subset of users at one time. For a selected user subset, an addition of a new user would induce a change in sum rate capacity ΔC that can be separated into two components [19], namely the incremental gain in sum rate capacity C_{gain} and the incremental capacity degradation C_{loss} due to the interference of this new user on the existing users. If the gain surpasses the loss, the incoming user would exert a positive influence on the sum rate capacity, and vice versa. In view of this, one promising user selection criterion is to evaluate the impact of each user from the candidate user pool on the change in sum rate capacity of the user subset, followed by enrolling a user into the subset if it brings the largest and positive value of ΔC.

Theorem 2 (Change in capacity due to a new user [20]): For a K-heterogeneous user downlink MU-MIMO system, the change in sum rate capacity when a new user is added in a selected user subset $\mathcal{T}$ is

$$\Delta C\quad=\quad C_{gain}-C_{loss},\qquad(6)$$

where C_{gain} quantifies the gain in sum rate capacity due to a new incoming user (user k) and it is approximated as

$$C_{gain}\quad\approx\quad\log_2(\frac{\rho_k}{\sigma_n^2}\det\left(\bar{\mathbf{H}}_k\bar{\mathbf{H}}_k^H\right)\sin^2\psi_{k,s}),\qquad(7)$$

with $\psi_{k,s}=\measuredangle(\bar{\mathbf{H}}_k,\mathbf{H}_s)$ being the geometrical angle between the range spaces of the channel matrix of the new user ($\bar{\mathbf{H}}_k$) and the aggregated channel matrix of the existing users in the selected user subset ($\mathbf{H}_s$), while C_{loss} denotes the loss in sum rate capacity resulting from the interference of this new user to existing users and it is given by

$$C_{loss}\approx\sum_{j\in\mathcal{T}}\log_2(\frac{\rho_j}{\sigma_n^2}\det\left(\bar{\mathbf{H}}_j\bar{\mathbf{H}}_j^H\right)\sin^2\psi_{j,s\backslash j}\sin^2\psi_{k,s\backslash j}),\quad(8)$$

with $\psi_{j,s\backslash j}=\measuredangle(\bar{\mathbf{H}}_j,\mathbf{H}_s\backslash\mathbf{H}_j)$ being the geometrical angle between the range spaces of the channel matrix of the j-th user in the subset ($\bar{\mathbf{H}}_j$) and the aggregated channel matrix of the other selected users $\mathbf{H}_s\backslash\mathbf{H}_j$, and similar definition holds for $\psi_{k,s\backslash j}=\measuredangle(\bar{\mathbf{H}}_k,\mathbf{H}_s\backslash\mathbf{H}_j)$. □

Proof 2: Please refer to Appendix B for details.

C_{loss} can be geometrically interpreted by the following example. Suppose there are two users (user-1 and user-2) in an existing user subset and a new incoming user (user-3) whose effective channel lies in the intersection of the null spaces of $\mathbf{H}_1$ and $\mathbf{H}_2$. As referred to (8), the capacity loss due to user-3 is mainly due to two components: one is the projection of user-3's channel onto the null space of $\mathbf{H}_1$ followed by the range space of $\mathbf{H}_2$, and the other one is the projection of user-3's channel onto the null space of $\mathbf{H}_2$ and then the range space of $\mathbf{H}_1$. As shown in Fig. 1(b), the first component is equivalent to recursively project OA onto OB and then OC. Similarly, the second component is geometrically equivalent to a recursive projection of OA onto OE followed by OD.

From Theorem 2, it is clear that the performance metric is to maximize the incremental improvement in (6), i.e.,

$$\mathcal{M}(\mathbf{H}_k,\mathbf{H}_s)$$
$$=\arg\max_{k\in\mathcal{C}}\frac{\sigma_n^{2(|\mathcal{T}|-1)}\rho_k\det\left(\bar{\mathbf{H}}_k\bar{\mathbf{H}}_k^H\right)\sin^2\psi_{k,s}}{\prod_{j\in\mathcal{T}}\rho_j\det\left(\bar{\mathbf{H}}_j\bar{\mathbf{H}}_j^H\right)\sin^2\psi_{j,s\backslash j}\sin^2\psi_{k,s\backslash j}},\qquad(9)$$

which measures the influence introduced by the incoming user ($\mathbf{H}_k$) on the already selected users ($\mathbf{H}_s$), and it is used as a scheduling criterion for user selection or hybrid user scheduling. Alternatively, a simplified user selection criterion is to solely consider C_{gain}, namely

$$\mathcal{M}(\mathbf{H}_k,\mathbf{H}_s)\quad=\quad\arg\max_{k\in\mathcal{C}}\rho_k\det\left(\bar{\mathbf{H}}_k\bar{\mathbf{H}}_k^H\right)\sin^2\psi_{k,s},$$
$$=\quad\arg\max_{k\in\mathcal{C}}\rho_k\det\left(\bar{\mathbf{H}}_k\mathbf{H}_s^{\perp H}\mathbf{H}_s^{\perp}\bar{\mathbf{H}}_k^H\right)(10)$$

Geometrically, this simplified criterion (10) refers to the volume of the parallelepiped spanned by the projection of the basis vectors of the range space of $\bar{\mathbf{H}}_k$ onto the null space of $\mathbf{H}_s$. When compared (10) with an alternative expression of geometrical angle to (3) in terms of sine function, their difference lies on the volume of channel matrix, i.e., $\rho_k\det\left(\bar{\mathbf{H}}_k\bar{\mathbf{H}}_k^H\right)$. Nevertheless, the impact of the channel volume on $\mathcal{C}_k$ vanishes when the number of users is asymptotically large because a best user with $\rho_k \det\left(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^H\right) \to 1$ can always be found.

To sum up, our proposed criteria can be applied in various user scheduling algorithms. In particular, the geometrical angle (3) and grouping-oriented criterion (5) are more suitable for user grouping algorithms because both criteria emphasize on the integrated effect of all involved users. On the other hand, the selection-oriented criteria (9) and (10) focus more on the impact of a newly-recruited user on the sum rate capacity and therefore, they are more appropriate for user selection algorithms or our proposed hybrid user scheduling algorithms.

IV. Proposed Scheduling Algorithms for Heterogeneous Users

A. Conventional User Grouping Algorithm [6]

Consider a downlink MU-MIMO system with $K = 2L$ homogeneous users. Assume that the channel remains unchanged during the entire scheduling period of L timeslots such that these K users are divided into L groups of size $G = 2$. The objective of [6] is to design a user scheduling algorithm that minimizes spatial correlation between two users per group while maximizing multiuser diversity by using the largest principal angle as the user grouping criteria. In particular, given $C_2^{2L} C_2^{2L-2} \cdots C_2^2 / L!$ possible arrangements, a max-min operation is performed in which the smallest largest principal angle for each arrangement is first identified, followed by selecting an arrangement with the largest value among all these $C_2^{2L} C_2^{2L-2} \cdots C_2^2 / L!$ smallest angles as the best one. However, there are two main drawbacks in applying this algorithm to systems with heterogeneous users.

1) *Reduced Average Sum Rate Capacity.* Since homogeneous users are equipped with the same number of receive antennas, the group size can be heuristically set as a constant $G = M_T / M_R$, where $M_{R_k} = M_R$ for all k. For heterogeneous users, however, it is not wise to determine the group size in advance because each user may have different number of receive antennas. For example, if we set $G = \lfloor M_T / \max M_{R_k} \rfloor$, it is apparent that either the total degree of freedom per group cannot be fully utilized or a larger number of groups is required[3], which results in a lower average sum rate capacity per group. On the other hand, if we set the group size according to the minimum number of receive antennas, the total number of receive antennas in a group will be larger than that of transmit antennas, which violates the dimensionality constraint of BD.
2) *Huge Computational Complexity.* Roughly speaking, the algorithm involves as many as $C_2^{2L} C_2^{2L-2} \cdots C_2^2 / L!$ possible arrangements. Since there are L groups for each arrangement, more than $\mathcal{O}(L^2)$ comparisons are required per arrangement.

[3]Take a $\{1,1,1,2,3,4\} \times 6$ MU-MIMO system as an example. There are 3 groups that separately consist of $G = 1$ user with single receive antenna, and the remaining degree of freedom (i.e., $M_T - \min M_{R_k} = 5$ dimensions) cannot be fully utilized. Since it requires $N_G = 6$ groups in serving all users, it is equivalent to TDMA in which only one user is served at a time. Due to the lack of spatial multiplexing among users, the sum rate capacity per group is significantly reduced.

Because of these concerns, we develop two hybrid user scheduling algorithms that takes into account some key features of user grouping and user selection algorithms, i.e., capture fairness among users and maximize the system performance in a greedy manner. These two algorithms, which aim at minimizing group size and maximizing degree of freedom, are outlined in Tables I and II and summarized as follows.

B. Algorithm 1: Group Number Minimization

In contrast to the conventional algorithm that considers a constant group size, we alternatively consider variable group size and minimize the number of groups N_G required by setting $N_G = \left\lceil \sum_{k=1}^{K} M_{R_k} / M_T \right\rceil$. Each group is allowed to have different number of group members as long as its total number of receive antennas is smaller than or equal to M_T, i.e., the total degree of freedom available for interference-free transmission with BD. In this case, we can ensure that the algorithm provides the same fairness as the conventional algorithm but requires a fewer number of groups[4].

For each group $\mathcal{T}^{(g)}$, where $g = 1, \cdots, N_G$, better users have higher priority in getting the resources. In particular, we select the best N_G users with the largest Frobenius norm from the candidate user pool $\mathcal{C} = \{1, \cdots, K\}$ and assign them to be the first user of each group (c.f., lines 4–8 of Table I). For the remaining $K - N_G$ users, they will be assigned to one of the N_G groups by certain criterion that aims at minimizing subspace correlation and interference non-orthogonality per group while maximizing the sum rate capacity. Taking into account the performance-and-complexity tradeoff, we consider the simplified selection-oriented criterion (10). It is important mentioning that the idea of our approach is inspired by the idea of greedy selection but with two main differences:

1) While typical user selection algorithms aims at choosing the "best" user for a group/user subset, our algorithm alternatively help users select the best group.
2) No user is allowed to be assigned into more than one group. In other words, each user is served by the BS only once within an entire scheduling period of N_G timeslots.

The selection procedure is summarized as follows. Firstly, we identify a user (say, user-k) with the largest Frobenius norm from the updated candidate user pool. Then, the simplified selection-oriented criterion is executed by selecting a group $\mathcal{T}^{(g)}$ that has the largest incremental gain in sum rate capacity due to the enrollment of this user while satisfying the dimensionality constraint of BD (c.f., lines 13 and 14 of Table I). If the constraint cannot be satisfied, this user will be assigned to the next best group. As a remark, once the user is selected, a user shedding step [2], [20] is performed to remove it from the candidate user pool (c.f., line 21 of Table I).

C. Algorithm 2: Degree-of-Freedom Maximization

The idea of this algorithm is to fully utilize the total degrees of freedom for every group which, according to the dimensionality constraint of BD, is the number of transmit

[4]We take a $\{1,1,1,2,3,4\} \times 6$ MU-MIMO system again as our example again. By using this proposed algorithm, N_G is significantly reduced to 2.

Table I
Hybrid Scheduling Algorithm-1: Group Number Minimization

Initialization
1: $N_G = \lfloor \sum_{k=1}^{K} M_{R_k} / M_T \rfloor$
2: $\mathcal{C} = \{1, 2, \cdots, K\}$
3: $\mathcal{T}^{(1)} = \mathcal{T}^{(2)} = \cdots = \mathcal{T}^{(N_G)} = \emptyset$
4: **for** $g = 1 \rightarrow N_G$
5: $u_1^{(g)} = \arg\max_{u \in \mathcal{C}} \|\mathbf{H}_u\|$
6: $\mathcal{T}^{(g)} = \mathcal{T}^{(g)} \cup \{u_1^{(g)}\}$
7: $\mathcal{C} = \mathcal{C} \backslash \{u_1^{(g)}\}$
8: **end**
9: **While** $|\mathcal{C}| > 0$
10: **for** $g = 1 \rightarrow N_G$
11: $\mathbf{H}_s^{(g)} = \{\mathbf{H}_i, i \in \mathcal{T}^{(g)}\}$
12: **end**
13: $u_k = \arg\max_{k \in \mathcal{C}} \|\mathbf{H}_k\|$
14: $g_s = \arg\max_{1 \leq g \leq N_G} \rho_k \det(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^H) \sin^2 \measuredangle(\mathbf{H}_s^{(g)}, \mathbf{H}_k)$
s.t. $rank(\mathbf{H}_s^{(g)}) + rank(\mathbf{H}_k) \leq M_T$
15: **if** $\{g_s\} \neq \emptyset$
16: $\mathcal{T}^{(g_s)} = \mathcal{T}^{(g_s)} \cup \{u_k\}$
17: **else**
18: $N_G \leftarrow N_G + 1$
19: $\mathcal{T}^{(N_G)} = \mathcal{T}^{(N_G)} \cup \{u_k\}$
20: **end**
21. $\mathcal{C} = \mathcal{C} \backslash \{u_k\}$
21. **end**

Table II
Hybrid Scheduling Algorithm-2: Degree-of-Fredom Maximization

Initialization
1: $\mathcal{C} = \{1, 2, \cdots, K\}$
2: $\mathcal{T}^{(1)} = \emptyset$
3: $g = 1$
4: **while** $|\mathcal{C}| > 0$
5: **if** $\mathcal{T}^{(g)} == \emptyset$
6: $u_1 = \arg\max_{u \in \mathcal{C}} \|\mathbf{H}_u\|$
7: $\mathcal{T}^{(g)} = \mathcal{T}^{(g)} \cup \{u_1\}$
8: $\mathcal{C} = \mathcal{C} \backslash \{u_1\}$
9: **else**
10: $\mathbf{H}_s = \{\mathbf{H}_i, i \in \mathcal{T}^{(g)}\}$
11: $u_k = \arg\max_{u \in \mathcal{C}} \rho_k \det(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^H) \sin^2 \measuredangle(\mathbf{H}_s, \mathbf{H}_k)$
s.t. $rank(\mathbf{H}_s) + rank(\mathbf{H}_k) \leq M_T$
12: **if** $\{u_k\} \neq \emptyset$
13: $\mathcal{T}^{(g)} = \mathcal{T}^{(g)} \cup \{u_k\}$
14: $\mathcal{C} = \mathcal{C} \backslash \{u_k\}$
15: **else**
16: $g \leftarrow g + 1$
17: **end**
18: **end**
19: **end**

antennas at the BS, M_T. The algorithm is initialized by setting the group size to the number of transmit antennas, i.e., $G = M_T$. User selection is started at the first group $\mathcal{T}^{(1)}$ by choosing the first user out of the K total users in the candidate user pool $\mathcal{C}$ with the maximum Frobenius norm (c.f., lines 5–8 of Table II). Like the previous algorithm, a user shedding step is performed such that the selected user will no longer be considered again in next iterations.

Then, the next best users u_k for $\mathcal{T}^{(1)}$ are chosen from the updated candidate user pool according to the simplified selection-oriented criterion given in (10) (c.f., lines 13 and 14 of Table II). This selection process for $\mathcal{T}^{(1)}$ is terminated when the sum of channel ranks of the existing users and the new incoming user is larger than the remaining degree of freedom, i.e., $rank(\mathbf{H}_s) + rank(\mathbf{H}_k) > M_T$. Finally, the whole scheduling procedure repeats for the second group $\mathcal{T}^{(2)}$ and so on until all of the K users have been assigned.

D. Advantages

Compared with the conventional user grouping algorithm [6], our algorithms have two main advantages. Firstly, *the average sum rate capacity is higher.* Since the proposed algorithms provide an efficient mechanism in minimizing the number of groups required and utilize greedy-based criteria in enrolling as many users in one group as possible. Therefore, the total degree of freedom per group is better exploited and the timeslots required for providing fairness are largely reduced, which results in a higher average sum rate capacity per group than [6]. The second advantage is a *lower computational complexity*. While the conventional user grouping algorithm requires about $L^2 C_2^{2L} C_2^{2L-2} \cdots C_2^2 / L!$ comparisons in finding an optimal grouping arrangement, these numbers are significantly required to approximately $N_G(2L - N_G)$ and $L(2L-1)$ for Algorithms-1 and 2, respectively.

V. Numerical Results

Monte Carlo simulations are provided to evaluate the effectiveness of our three proposed user scheduling criteria and the two proposed hybrid user scheduling algorithms in terms of the 10% outage capacity [21], which is defined as the rate that the channel can support with 90% probability.

For simplicity, we consider a general Kronecker Product Form channel model [22] for the simulation, i.e., $\mathbf{H}_k = \sqrt{\rho_k} \mathbf{R}_{r,k}^{1/2} \mathbf{H}_{k,w} \mathbf{R}_{t,k}^{1/2}$, where $\rho_k = P_T / M_T d_k^{\alpha}$ is the received power of user-k with P_T, d_k and α being the transmit power, the distance between BS and k-th user, and path loss exponent, respectively. In addition, $\mathbf{H}_{k,w} \in \mathcal{C}^{M_{R_k} \times M_T}$ is a zero-mean unit-variance i.i.d. complex Gaussian matrix between the BS and the k-th user, $\mathbf{R}_{r,k} \in \mathcal{C}^{M_{R_k} \times M_{R_k}}$ and $\mathbf{R}_{t,k} \in \mathcal{C}^{M_T \times M_T}$ are the receive and transmit correlation matrices of user-k, which can be modeled as $[\mathbf{R}_{r,k}]_{ij} = \gamma_{r,k}^{|i-j|^2}$ and $[\mathbf{R}_{t,k}]_{ij} = \tau_{t,k}^{|i-j|^2}$, respectively, with correlation coefficients $\gamma_{r,k}, \tau_{t,k}$.

Unless stated otherwise, the simulation configurations of some key parameters are listed as follows.

- We employ BD as the linear precoding algorithm. Water-filling policy is considered even though the theorems are derived by following an equal power allocation policy.
- M_{R_k} is randomly chosen from $\{1, \cdots, N\}$ with equal probability, where N is the largest number of receive antennas in the system.
- As for the received power ρ_k, α is set to 3 and d_k is randomly generated with range $[200m, 1000m]$. Then, ρ_k is normalized by the maximum possible received power when $d_k = 200m$.
- $\gamma_{r,k}$ and $\tau_{t,k}$ are modeled as uniformly distributed variables with range [0,1].

As the first example, we compare the performance of the geometrical angle (3) and grouping-oriented criterion (5) with the largest principal angle, subspace collinearity and chordal distance. Two baseline criteria, namely exhaustive search and random selection, are also considered. Fig. 2 shows the 10% outage capacity of these user grouping criteria

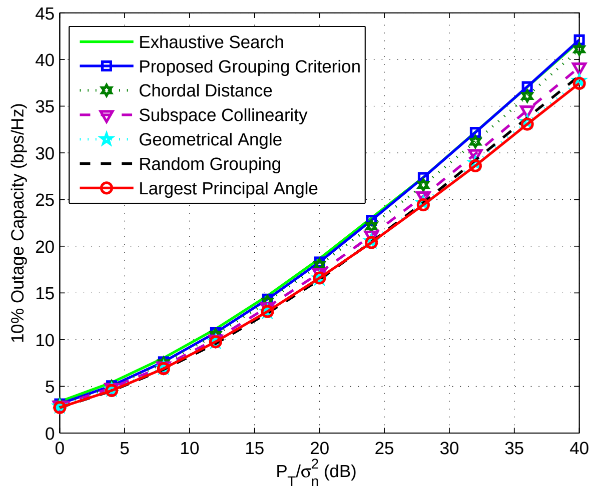

Fig. 2. 10% outage capacity performance of various user grouping criteria. A $\{1,1,1,2,3,4\}\times 6$ system is considered. The proposed criterion refers to the "Grouping-Oriented Criterion".

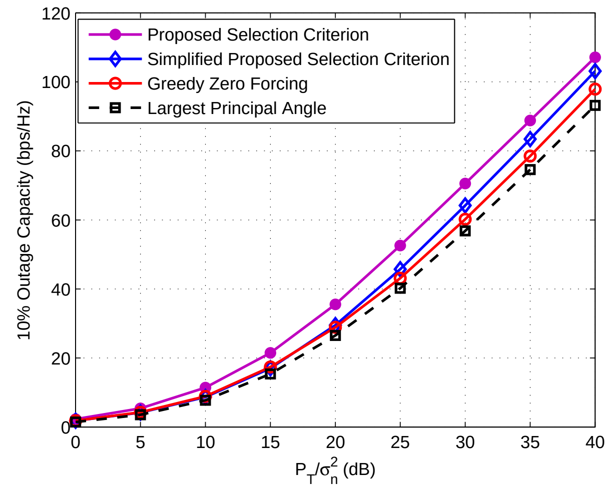

Fig. 3. 10% outage capacity performance of various user selection criteria. There are $M_T = 12$ transmit antennas at the base station and 20 users with either 1 or 2 receive antennas. The proposed criterion refers to the "Selection-Oriented Criterion".

for a $\{1,1,1,2,3,4\}\times 6$ system. It can be observed that our proposed criteria outperform the largest principal angle. Further, the grouping-oriented criterion achieves the average sum rate capacity per group of the exhaustive search method at high SNR (P_T/σ_n^2) with less computational complexities.

To illustrate the effectiveness of our third proposed metric, i.e., selection-oriented criterion, we apply it into a greedy user selection algorithm whose details are presented as follows. Let $\mathcal{C} = \{1, 2, \cdots, K\}$ and $\mathcal{T} = \emptyset$ be, respectively, the set of the candidate user pool and the scheduled user pool. The algorithm is initialized by specifying a maximum degree of freedom D available for interference-free transmission. Due to the dimensionality constraint of BD, it is usually an integer no larger than the total number of transmit antennas, i.e., $D = M_T$. User selection is started by choosing the first user with the maximum Frobenius norm. i.e., $u_1 = \arg\max_{u\in\mathcal{C}} ||\mathbf{H}_u||_F$. The two user pools are updated accordingly as $\mathcal{T} = \mathcal{T} \cup \{u_1\}$ and $\mathcal{C} = \mathcal{C}\backslash\{u_1\}$. Then, the next best users u_k, where $k \in \mathcal{C}$, are chosen from the updated candidate user pool according to the simplified selection-oriented criterion (10) as an illustrative example. The selection process will be terminated if the sum of channel ranks of the existing users in the subset and the new incoming user is larger than the remaining degree of freedom, i.e., $rank(\mathbf{H}_s) + rank(\mathbf{H}_k) > M_T$. Fig. 3 illustrates the effectiveness of our proposed selection-oriented criterion. We consider a heterogeneous MU-MIMO broadcast channel, where the BS has 12 transmit antennas and each of the 20 users equips with either 1 or 2 receive antennas. For comparison purpose, we have also considered (a) the greedy zero-forcing algorithm with BD precoding [3], [4] and (b) applying the largest principal angle into the user selection algorithm. As referred to the figure, it is clear that the largest principal angle does not perform well because of its incapability in reflecting users' spatial separability accurately. On the other hand, our proposed simplified selection-oriented criterion performs better than the greedy zero forcing algorithm and the performance difference increases with SNR (for example, from less than 5 bps/Hz at 20 dB to more than 10 bps/Hz at 40 dB). In addition to the simplified selection criterion, we have also shown the performance of the original version of our proposed criterion (9) that considers both C_{gain} and C_{loss}. It is seen that the capacity improvement is even higher despite an increase in computational complexity but the performance of the simplified criterion (10) would approach the original one (9) when the number of users is asymptotically large. It is because when the number of users in the candidate user pool increases, we can always find a user with the largest gain in sum rate capacity and a relatively smaller capacity loss.

Apart from comparing the performance of various user scheduling criteria from the SNR point of view, we also investigate into their performance in terms of the number of users available for scheduling. In order to demonstrate the effectiveness of (9), we also consider an optimal user selection by exhaustive search. As referred to Fig. 4, it is seen that at high SNR, the performance of our proposed selection-oriented criterion approaches that of the optimal user selection when the number of users is large enough (e.g., 30 users). There is also an interesting observation on geometrical angle. Namely, while its sum rate capacity is the lowest among all possible criteria when there are only a few users in the candidate user pool, its performance increases with the number of users and approaches that of our proposed selection-oriented criterion. This observation is consistent with our findings in Section III.B that the impact of the channel volume (i.e., $\rho_k \det\left(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^H\right)$ in (9) and (10)) on the sum rate capacity vanishes when the number of users is asymptotically large.

Lastly, we demonstrate the effectiveness of our hybrid user scheduling algorithms over [6] in a $\{1,1,1,2,3,4\}\times 6$ configuration. As discussed in Section IV.A, the group size of [6] is pre-determined and its average sum rate capacity per group is expected to be lower than that of our proposed algorithms.

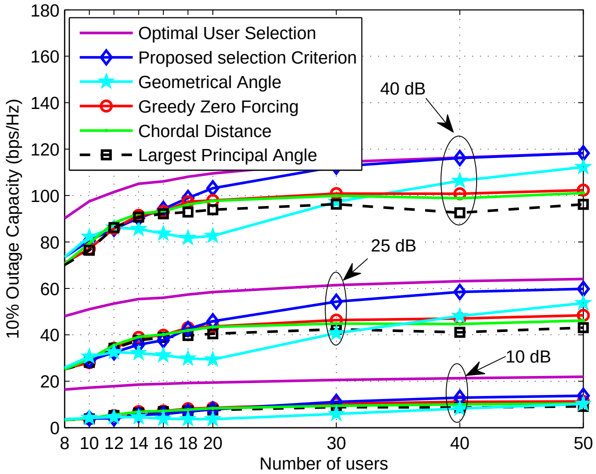

Fig. 4. The impact of the number of users on the 10% outage capacity performance of various user selection criteria. There are $M_T = 12$ transmit antennas at the base station and each user is equipped with either 1 or 2 receive antennas. The proposed criterion refers to the "Selection-Oriented Criterion".

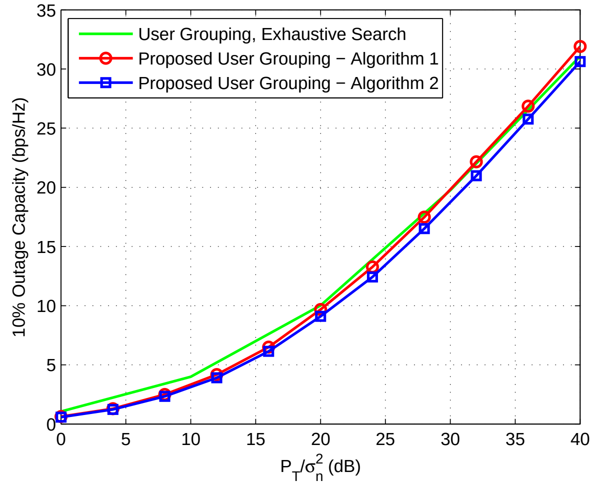

Fig. 5. 10% outage capacity performance of the two proposed hybrid user scheduling algorithms and the user grouping algorithm [6]. A $\{1, 1, 1, 2, 3, 4\} \times 6$ system is considered. "Selection-Oriented Criterion" is applied for the proposed algorithms while exhaustive search is applied for [6].

In view of these concerns, we apply an optimal user grouping strategy by exhaustive search in [6], while considering a sub-optimal yet simplified selection-oriented criterion (10) for our algorithms. It can be observed from Fig. 5 that though our algorithms perform slightly inferior than [6] at low SNR due to the asymptotic SNR approximation (14) for C_{gain}, they perform better at high SNR while requiring significantly less computational complexities.

VI. Conclusion

We have investigated into the design of user scheduling metrics for downlink MU-MIMO systems with heterogeneous users. We study users' channel characteristics in a subspace approach by representing mutual interference across users that are originated from interference non-orthogonality as the inter-user subspace correlation, and find that those conventional subspace-based user scheduling criteria that are commonly used in homogeneous users do not accurately reflect users' spatial separability. In response, we design from a geometric point of view three effective user scheduling metrics that aim at maximizing sum rate capacity while minimizing interference non-orthogonality among users. We also propose two hybrid user scheduling algorithms that can capture fairness among users while maximizing sum rate capacity in a greedy manner. When compared with the conventional user scheduling algorithm, our proposed approaches have lower computational complexities and shown to achieve a higher average sum rate capacity.

Appendix

A. Proof of Theorem 1

For BD, the precoding matrix of user-k is expressed as a product of two precoders $\mathbf{F}_{a_k}$ and $\mathbf{F}_{b_k}$, i.e., $\mathbf{F}_k = \beta \mathbf{F}_{a_k}\mathbf{F}_{b_k} = \beta\tilde{\mathbf{V}}_k^{(0)}\mathbf{F}_{b_k}$. in which the former is used for interference suppression and the latter for performance optimization. Here, the columns of $\tilde{\mathbf{V}}_k^{(0)}$ act as basis vectors that span the null space of the interference channel $\tilde{\mathbf{H}}_k$, and β is chosen such that the total transmit power is less than the maximum transmit power constraint P_T. Assuming equal power allocation ($\beta^2\mathbf{F}_{b_k}\mathbf{F}_{b_k}^H = \mathbf{I}$), the sum rate capacity of user-k is given by

$$\begin{aligned} C_k &= \log_2 \det\left(\mathbf{I}_{M_T} + \frac{1}{\sigma_n^2}\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{F}_k^H\right) \\ &= \log_2 \det\left(\mathbf{I}_{M_T} + \frac{\rho_k}{\sigma_n^2}\bar{\mathbf{V}}_k^{(1)}\bar{\mathbf{\Sigma}}_k^2\bar{\mathbf{V}}_k^{(1)H}\tilde{\mathbf{V}}_k^{(0)}\tilde{\mathbf{V}}_k^{(0)H}\right), \end{aligned}$$

where the first line is due to the zero-interference constraint of BD that ensures $\tilde{\mathbf{H}}_k\mathbf{F}_k = \mathbf{0}$ [1], and the second line is due to the definition of $\mathbf{H}_k = \sqrt{\rho_k}\bar{\mathbf{H}}_k$, the equal power allocation policy, and eigenvalue decomposition of $\bar{\mathbf{H}}_k$, with $\bar{\mathbf{\Sigma}}_k = diag\{\lambda_{k,1}, \cdots, \lambda_{k,M_{R_k}}\}$ being its singular matrix and the columns of $\bar{\mathbf{V}}_k^{(1)}$ being the basis vectors spanning its range space. Denote $\mathbf{T}_k = \tilde{\mathbf{V}}_k^{(0)H}\bar{\mathbf{V}}_k^{(1)}$. The sum rate capacity can be expressed in terms of the eigenvalue matrix of $\mathbf{T}_k\bar{\mathbf{\Sigma}}_k^2\mathbf{T}_k^H$:

$$C_k = \log_2 \det\left(\mathbf{I}_{M_T-\tilde{L}_k} + \frac{\rho_k}{\sigma_n^2}\Lambda(\mathbf{T}_k\bar{\mathbf{\Sigma}}_k^2\mathbf{T}_k^H)\right), \quad (11)$$

where $\tilde{L}_k$ is the rank of $\tilde{\mathbf{H}}_k$, and $\Lambda(\cdot)$ represents the corresponding diagonal matrix. Though (11) is exact, it is not easy to obtain any insight and therefore, we resort to develop upper and lower bounds of C_k by using the following propositions.

Proposition 1 (Upper bound on a matrix determinant [23]): For any positive definite matrix $\mathbf{M}_C$ and any positive integer m, the following relation holds $\det(\mathbf{M}_C) \le (tr(\mathbf{M}_C)/m)^m$. □

Proposition 2 (Trace inequality for matrix product [24]): For any two Hermitian positive semi-definite matrices $\mathbf{M}_D$ and $\mathbf{M}_E$, there holds $\sum_{i=1}^n \lambda_i(\mathbf{M}_D)\lambda_{n-i+1}(\mathbf{M}_E) \le tr(\mathbf{M}_D\mathbf{M}_E) \le \sum_{i=1}^n \lambda_i(\mathbf{M}_D)\lambda_i(\mathbf{M}_E)$, where $\lambda_i(\cdot)$ is the i-th singular value. □

By using Proposition 1, (11) can be upper-bounded as

$$C_k \leq M_{R_k} \log_2 \left(1 + \frac{\rho_k}{M_{R_k} \sigma_n^2} tr(\bar{\mathbf{\Sigma}}_k^2 \mathbf{T}_k^H \mathbf{T}_k) \right). \quad (12)$$

Denote $\lambda_i(\mathbf{T}_k) = \sin\theta_{k,\tilde{k},i}$ with $\theta_{k,\tilde{k},i}$ being the i-th principal angle of the two subspaces $\tilde{\mathbf{V}}_k^{(0)}$ and $\bar{\mathbf{V}}_k^{(1)}$. The sum rate capacity (12) can further be upper-bounded as the following closed-form expression by using Proposition 2.

$$C_k \leq M_{R_k} \log_2 \left(1 + \frac{\rho_k \lambda_{k,max}^2}{M_{R_k} \sigma_n^2} \sum_{i=1}^{M_{R_k}} \sin^2 \theta_{k,\tilde{k},i} \right),$$

where $\lambda_{k,max} = \lambda_{k,1}$ is the maximum eigenvalue of $\bar{\mathbf{H}}_k$.

Similarly, the capacity lower bound can be developed by using Proposition 2 as follows.

$$\begin{aligned} C_k &= \log_2 \det \left(\mathbf{I}_{M_T - \tilde{L}_k} + \frac{\rho_k}{\sigma_n^2} \Lambda(\mathbf{T}_k \bar{\mathbf{\Sigma}}_k^2 \mathbf{T}_k^H) \right) \\ &\geq \log_2 \left(1 + \frac{\rho_k}{\sigma_n^2} \sum_{i=1}^{M_{R_k}} \lambda_{k,n-i+1}^2 \sin^2 \theta_{k,\tilde{k},i} \right) \\ &\geq \log_2 \left(1 + \frac{\rho_k \lambda_{k,min}^2}{\sigma_n^2} \sum_{i=1}^{M_{R_k}} \sin^2 \theta_{k,\tilde{k},i} \right), \end{aligned}$$

with $\lambda_{k,mix} = \lambda_{k,M_{R_k}}$ being the minimum eigenvalue of $\bar{\mathbf{H}}_k$.

This completes the proof of Theorem 1.

B. Proof of Theorem 2

For a K-heterogeneous user downlink MU-MIMO system, the sum rate capacity is updated by an amount ΔC when a new user is added in the selected user subset $\mathcal{T}$ [19], i.e.,

$$\Delta C = C_{gain} - C_{loss},$$

where C_{gain} refers to the gain in sum rate capacity due to this new user and it is given by

$$\begin{aligned} &C_{gain} \\ &= \log_2 \det \left(\mathbf{I}_{M_T} + \frac{1}{\sigma_n^2} \mathbf{H}_k^H \mathbf{H}_k \mathbf{F}_k \mathbf{F}_k^H \right) \\ &= \log_2 \det \left(\mathbf{I}_{M_{R_k}} + \frac{\rho_k}{\sigma_n^2} \bar{\mathbf{\Sigma}}_k \bar{\mathbf{\Sigma}}_k^T \bar{\mathbf{V}}_k^{(1)H} \mathbf{V}_s^{(0)} \mathbf{V}_s^{(0)H} \bar{\mathbf{V}}_k^{(1)} \right), \end{aligned}$$

with the precoder $\mathbf{F}_k = \beta \mathbf{V}_s^{(0)} \mathbf{F}_{b_k}$. Here, the columns of $\bar{\mathbf{V}}_k^{(1)}$ and $\mathbf{V}_s^{(0)}$ being the basis vectors that span, respectively, the range space of the incoming user-k's channel $\bar{\mathbf{H}}_k$ and the null space of the aggregated channels of the existing users in the subset $\mathbf{H}_s$. Denote $\mathbf{\Gamma}_k = \mathbf{V}_s^{(0)H} \bar{\mathbf{V}}_k^{(1)}$. We can asymptotically approximate C_{gain} with respect to SNR as

$$C_{gain} \approx \log_2 \left(\frac{\rho_k}{\sigma_n^2} \det \left(\bar{\mathbf{\Sigma}}_k \bar{\mathbf{\Sigma}}_k^T \right) \prod_{i=1}^{M_{R_k}} \sin^2 \theta_{k,s,i} \right), \quad (13)$$

with $\theta_{k,s,i}$ being the i-th principal angle of the two subspaces $\bar{\mathbf{V}}_k^{(1)}$ and $\mathbf{V}_s^{(0)}$. Following the definition of geometrical angle in Section III.B, (13) is written as

$$C_{gain} \approx \log_2 \left(\frac{\rho_k}{\sigma_n^2} \det \left(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^H \right) \sin^2 \psi_{k,s} \right). \quad (14)$$

Though the sum rate capacity is increased due to the incoming user-k, its presence in the subset induces interference and hence performance loss with the existing users. Denote, respectively, the sum rate capacity before and after enrolling user-k as C_{pre} and C_{post}, the loss in sum rate capacity C_{loss} can be quantified in the following way.

$$C_{loss} = C_{pre} - C_{post},$$

where

$$C_{pre} = \sum_{j \in \mathcal{T}} \log_2 \det \left(\mathbf{I}_{M_T} + \frac{1}{\sigma_n^2} \mathbf{H}_j^H \mathbf{H}_j \mathbf{V}_{\mathcal{T}\setminus j}^{(0)} \mathbf{V}_{\mathcal{T}\setminus j}^{(0)H} \right) \quad (15)$$

and

$$C_{post} = \sum_{j \in \mathcal{T}} \log_2 \det \left(\mathbf{I}_{M_T} + \frac{1}{\sigma_n^2} \mathbf{H}_j^H \mathbf{H}_j \mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)} \mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)H} \right),$$

with $\mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)}$ being the intersection of the null spaces of $\mathbf{H}_s \setminus \mathbf{H}_j$ and $\mathbf{H}_k$. In order to make $\mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)}$ tractable, we apply alternating projection algorithm [26] into $\mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)} \mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)H}$ such that the intersection of two subspaces is approximated by the infinite power of the product of their projection matrices, namely $\mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)} \mathbf{V}_{(\mathcal{T}\setminus j)\cap k}^{(0)H} \approx \left(\mathbf{V}_{\mathcal{T}\setminus j}^{(0)} \mathbf{V}_{\mathcal{T}\setminus j}^{(0)H} \mathbf{V}_k^{(0)} \mathbf{V}_k^{(0)H} \mathbf{V}_{\mathcal{T}\setminus j}^{(0)} \mathbf{V}_{\mathcal{T}\setminus j}^{(0)H} \right)^{\kappa}$, where $\kappa \to \infty$. Since our main focus is to investigate into the first-order capacity change from a geometrical viewpoint, rather than to come up with an exact closed-form expression, we consider $\kappa = 1$ for the ease of derivation. Then, C_{post} is approximated as (16), where the columns of $\mathbf{V}_{\mathcal{T}\setminus j}^{(0)}$ refer to the basis vector that span the null space of $\mathbf{H}_s \setminus \mathbf{H}_j$, i.e., the aggregated channels of the existing users except user-j.

Given (15), (16) and the fact that $\log_2 \det(\mathbf{I} + \mathbf{M}_F + \mathbf{M}_G) - \log_2 \det(\mathbf{I} + \mathbf{M}_F) = \log_2 \det(\mathbf{I} + \mathbf{M}_G)$ holds for any two matrices $\mathbf{M}_F$ and $\mathbf{M}_G$ that are orthogonal to each other, C_{loss} is given as (17). Further denote $\mathbf{\Upsilon}_{js} = \mathbf{V}_{\mathcal{T}\setminus j}^{(0)H} \bar{\mathbf{V}}_j^{(1)}$ and $\mathbf{\Upsilon}_{ks} = \mathbf{V}_{\mathcal{T}\setminus j}^{(0)H} \mathbf{V}_k^{(1)}$. We asymptotically approximate (17) as

$$\begin{aligned} &C_{loss} \\ &\approx \sum_{j \in \mathcal{T}} \log_2 \det \left(\frac{\rho_j}{\sigma_n^2} \bar{\mathbf{\Sigma}}_j \bar{\mathbf{\Sigma}}_j^T \mathbf{\Upsilon}_{js}^H \mathbf{\Upsilon}_{ks} \mathbf{\Upsilon}_{ks}^H \mathbf{\Upsilon}_{js} \right) \\ &= \sum_{j \in \mathcal{T}} \log_2 \left(\frac{\rho_j}{\sigma_n^2} \det \left(\bar{\mathbf{\Sigma}}_j \bar{\mathbf{\Sigma}}_j^T \right) \prod_{i=1}^{M_{R_j}} \lambda_i^2(\mathbf{\Upsilon}_{js}) \prod_{i=1}^{M_{R_k}} \lambda_i^2(\mathbf{\Upsilon}_{ks}) \right) \\ &= \sum_{j \in \mathcal{T}} \log_2 \left(\frac{\rho_j}{\sigma_n^2} \det(\bar{\mathbf{\Sigma}}_j \bar{\mathbf{\Sigma}}_j^T) \prod_{i=1}^{M_{R_j}} \sin^2 \theta_{j,s\setminus j,i} \prod_{i=1}^{M_{R_k}} \sin^2 \theta_{k,s\setminus j,i} \right), \end{aligned}$$

where $\lambda_i(\mathbf{\Upsilon}_{js}) = \sin\theta_{j,s\setminus j,i}$ with $\theta_{j,s\setminus j,i}$ being the i-th principal angle of the two subspaces $\mathbf{V}_{\mathcal{T}\setminus j}^{(0)H}$ and $\bar{\mathbf{V}}_j^{(1)}$, and similar definition holds for $\lambda_i(\mathbf{\Upsilon}_{ks})$. Finally, we follow the definition of geometrical angle and rewrite the approximated loss in sum rate capacity as

$$C_{loss} \approx \sum_{j \in \mathcal{T}} \log_2 \left(\frac{\rho_j}{\sigma_n^2} \det \left(\bar{\mathbf{H}}_j \bar{\mathbf{H}}_j^H \right) \sin^2 \psi_{j,s\setminus j} \sin^2 \psi_{k,s\setminus j} \right).$$

This completes the proof of Theorem 2.

$$C_{post} \approx \sum_{j \in \mathcal{T}} \log_2 \det \left(\mathbf{I}_{M_T} + \frac{1}{\sigma_n^2} \mathbf{H}_j^H \mathbf{H}_j \mathbf{V}_{\mathcal{T} \setminus j}^{(0)} \mathbf{V}_{\mathcal{T} \setminus j}^{(0)H} \mathbf{V}_k^{(0)} \mathbf{V}_k^{(0)H} \mathbf{V}_{\mathcal{T} \setminus j}^{(0)} \mathbf{V}_{\mathcal{T} \setminus j}^{(0)H} \right), \tag{16}$$

$$C_{loss} \approx \sum_{j \in \mathcal{T}} \log_2 \det \left(\mathbf{I}_{M_T} + \frac{1}{\sigma_n^2} \mathbf{H}_j^H \mathbf{H}_j \mathbf{V}_{\mathcal{T} \setminus j}^{(0)} \mathbf{V}_{\mathcal{T} \setminus j}^{(0)H} \mathbf{V}_k^{(1)} \mathbf{V}_k^{(1)H} \mathbf{V}_{\mathcal{T} \setminus j}^{(0)} \mathbf{V}_{\mathcal{T} \setminus j}^{(0)H} \right). \tag{17}$$

REFERENCES

[1] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[2] T. Yoo and A. J. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[3] G. Dimic and N. 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] S. Sigdel and W. A. Krzymien, “Simplified fair scheduling and antenna selection algorithms for multiuser MIMO orthogonal space-division multiplexing downlink,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1329–1344, Mar. 2009.

[5] N. H. Dawod, I. D. Marsland, and R. H. M. Hafez, “Improved transmit null steering for MIMO-OFDM downlinks with distributed base station antenna arrays,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 419–426, Mar. 2006.

[6] C. Wang and R. D. Murch, “Adaptive downlink multi-user MIMO wireless systems for correlated channels with imperfect CSI,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2435–2446, Sep. 2006.

[7] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, Jr., and B. L. Evans, “Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization,” *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[8] R. Chen, Z. Shen, J. G. Andrews, and R. W. Heath, Jr., “Multimode transmission for multiuser MIMO systems with block diagonalization,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3294–3302, Jul. 2008.

[9] L. N. Tran and E. K. Hong, “Multiuser diversity for successive zero-forcing dirty paper coding: greedy scheduling algorithms and asymptotic performance analysis,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3411–3416, Jun. 2010.

[10] I. Sohn, J. G. Andrews, and K. B. Lee, “MIMO broadcast channels with spatial heterogeneity,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2449–2454, Aug. 2010.

[11] L. Sun and M. R. McKay, “Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection”, *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5246–5261, Oct. 2010.

[12] A. Bjorck and G. H. Golub, “Numerical methods for computing angles between linear subspaces,” *Math. Comput.*, vol. 27, no. 123, pp. 579–594, Jul. 1973.

[13] P.-A. Absil, A. Edelman, and P. Koev, “On the largest principal angle between random subspaces,” *Linear Algebra Appl.*, vol. 414, no. 1, pp. 288–294, April 2006.

[14] N. Czink, B. Bandemer, G. Vazquez-Vilar, L. Jalloul, C. Oestges, and A. Paulraj, “Spatial separation of multi-user MIMO channels,” in *Proc. IEEE PIMRC*, Tokyo, Japan, Sep. 2009, pp. 1059–1063.

[15] G. Golub and C. F. van Loan, *Matrix Computations, 3rd edition*. London: The Johns Hopkins University Press, 1996.

[16] K. Ko and J. Lee, “Low complexity multiuser MIMO scheduling with chordal distance,” in *Proc. IEEE CISS*, Princeton, NJ, United States, Mar. 2008, pp. 80–84.

[17] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc.: packings in Grassmannian space,” *Experimental Mathematics*, vol. 5, pp. 139–159, 1996.

[18] H. Gunawan, O. Neswan, and W. Setya-Budhi, “A formula for angles between two subspaces of inner product spaces,” *Beitrage zur Algebra und Geometrie*, vol. 46, no. 2, pp. 311–320, 2005.

[19] B. O. Lee, O.-S. Shin, and K. B. Lee, “Uplink user selection scheme for multiuser MIMO systems in a multicell environment,” in *Proc. IEEE ICC*, Cape Town, South Africa, May 2010.

[20] A. Razi, D. J. Ryan, I. B. Collings, and J. Yuan, “Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 356–365, Jan. 2010.

[21] V. Stankovic, *Multi-user MIMO wireless communications*. Ph.D. Dissertation, Ilmenau University of Technology, Nov. 2006.

[22] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, “A stochastic MIMO radio channel model with experimental validation,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[23] A. Bucur, “Properties regarding the trace of a matrix,” *General Mathematics*, vol. 14, no. 4, pp. 11–14, Oct. 2006.

[24] N. Komaroff, “Bounds on eigenvalues of matrix products with an application to the algebraic Riccati equation,” *IEEE Trans. Autom. Control*, vol. 35, no. 3, pp. 348–350, Mar. 1990.

[25] J. Miao and A. Ben-Israel, “Product cosines of angles between subspaces,” *Linear Algebra Appl.*, vol. 237/238, pp. 71–81, Apr. 1996.

[26] I. Halperin, “The product of projection operators,” *Acta Scientiarum Mathematicarum (Szeged)*, vol. 23, pp. 96–99, 1962.

Xinping Yi received his B.S. degree in 2005 from Huazhong University of Science and Technology and his M.Sc. degree in 2009 from University of Electronic Science and Technology of China, both in Electrical Engineering. Currently, He is pursuing the Ph.D. degree at Mobile Communication Department, EURECOM, Sophia Antipolis, France. From 2009 to 2011, he was a Research Engineer in Huawei Technologies, Shenzhen, China. His current research interests include multicell/multiuser MIMO systems and interference management.

Edward Au (S'02-M'07) received his Ph.D. degree in Hong Kong University of Science and Technology. At his capacity as a Principal Engineer in Huawei Technologies, he is taking care of WLAN standardization activity. He is currently an Associate Editor of IEEE Transactions on Vehicular Technology, and he has served as a leading guest editor for a feature topic of the IEEE Communications Magazine.