Asymptotically Optimal Simple User Scheduling for Massive MIMO Downlink with Two-Stage Beamforming

Gilwon Lee and Youngchul Sung
Dept. of Electrical Engineering
KAIST
Daejeon, South Korea 305-701
Email: {gwlee@ and ysung@ee.}kaist.ac.kr

***Abstract*—In this paper, a simple user-scheduling-and-beamforming method is proposed for massive multi-user multiple-input multiple-output (MU-MIMO) downlink adopting two-stage beamforming. The key ideas of the proposed scheduling-and-beamforming method are to divide users into several candidate subsets according to the level of alignment of user channels to the dominant directions of the channel covariance matrix and select the user in each candidate subset based on a certain channel quality indicator (CQI) and to apply post-selection zero-forcing beamforming (ZFBF) to the selected users based on their channel state information (CSI). It is proved that the proposed scheduling-and-beamforming method is asymptotically optimal as the number of users increases. Furthermore, the proposed method significantly reduces the feedback overhead and shows superior sum rate performance compared to existing scheduling methods for MU-MIMO downlink.**

I. Introduction

The massive MIMO technology is considered as one of the core technologies for next generation wireless communication. However, there are several challenges to realize the potential of massive MIMO in real world engineering. Practical precoding architecture design for massive MU-MIMO downlink is one of such challenges. Designing precoding vectors or matrices with very high dimensions without introducing an efficient structure requires heavy complexity. One feasible precoding solution to multi-user massive MIMO downlink is two-stage beamforming. Recently, Adhikary *et al.* proposed an efficient two-stage beamforming method named 'Joint Spatial Division and Multiplexing (JSDM)' for multi-user massive MIMO downlink [2]. The main ideas of JSDM are 1) to partition users in a cell into groups each of which has a distinguishable linear subspace spanned by the dominant eigenvectors of the group's channel covariance matrix and 2) to divide transmit beamforming into two stages: pre-beamforming that separates groups by designing a pre-beamforming matrix for each group filtering the dominant eigenvectors of each group's channel covariance matrix and following MU-MIMO precoding that separates the users within a group based on the effective channel formed as the product of the pre-beamforming matrix and the actual channel matrix. One major advantage of such two-stage beamforming is that the pre-beamforming matrices can be designed without much difficulty since the channel covariance matrix of a user terminal (UT) changes slowly compared with the CSI and can be estimated without knowing instantaneous CSI. Furthermore, the channel covariance matrix in a realistic environment has a much smaller rank than the size of the original physical channel and hence the dimension of the effective channel (whose state information should be acquired) is significantly reduced and conventional MU-MIMO techniques such as zero-forcing (ZF) or minimum mean-square error (MMSE) beamforming based on effective CSI can be applied to the second-stage beamforming of JSDM.

In this paper, we consider optimal user scheduling for such two-stage beamforming and propose a simple but asymptotically optimal user scheduling method for such two-stage beamforming. User scheduling based on beamforming or opportunistic beamforming for MU-MIMO systems has been investigated extensively for the past decade [3]–[7]. For example, two representative user selection schemes were proposed under random beamforming (RBF) [3] and ZFBF [5] for uncorrelated channels. Both schemes, [3] and [5], are asymptotically optimal but have significantly different performance in the practical case of finite users due to the difference in the amount of feedback required for user selection. Since the user selection scheme in [5], named 'semi-orthogonal user selection (SUS)', exploits full CSI from all users, a smart selection of beamforming directions is possible and SUS has fairly good performance under the ZFBF strategy. On the other hand, the RBF scheme in [3] chooses a group of users to be nearly matched to predetermined random beam directions, and requires only the feedback of the best beam direction index and the corresponding SINR value from each user. Thus, the feedback overhead can be reduced significantly for the RBF scheme. Due to such feedback advantage, the RBF scheme was extended to the single correlated channel case [6] and recently to JSDM with multiple correlated channel groups [7]. However, as we shall see in Section V, the RBF scheme shows poor performance in the practical case of finite users. In this paper, we propose a new simple user scheduling method for JSDM with multiple correlated channel groups that overcomes the disadvantages of SUS and RBF. The proposed method is asymptotically optimal and shows good performance with significantly reduced feedback overhead.

This research was supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2011-11913-04001). The journal version of this paper was submitted [1] and is available at http://arxiv.org/abs/1403.6931.

978-1-4799-4903-8/14/$31.00 ©2014 IEEE

Notation: Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^H$ and $[\mathbf{A}]_{i,j}$ indicate the conjugate transpose and the entry at the i-th row and j-th column of $\mathbf{A}$, respectively. $\text{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_n)$ denotes a diagonal matrix composed of diagonal elements $\mathbf{A}_1$, ... ,$\mathbf{A}_n$. For vector $\mathbf{a}$, $\|\mathbf{a}\|$ represents the 2-norm of $\mathbf{a}$. $\mathbf{I}_K$ is the $K \times K$ identity matrix. $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that random vector $\mathbf{x}$ is complex Gaussian distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\theta \sim \text{Unif}(a, b)$ means that θ is uniformly distributed for $\theta \in [a, b]$. $\mathbb{E}[\cdot]$ denotes statistical expectation.

II. System Model and Preliminaries

We consider a single-cell MIMO downlink system consisting of a single base station (BS) employing M transmit antennas and K single-antenna UTs. We consider the large-number-of-users regime, i.e., $K \gg M$, and assume that the BS chooses S ($\leq M$) users among the K users within the cell and broadcasts independent data streams to the S selected users. We assume that the users in the cell are partitioned into G groups, and $\sum_{g=1}^{G} K_g = K$ and $\sum_{g=1}^{G} S_g = S$, where K_g and S_g are the number of users and the number of independent data streams in group g, respectively. We assume that each group has a different channel covariance matrix and every user in a group has the same channel covariance matrix, as in [2]. Then, the channel vector $\mathbf{h}_{g_k}$ of user k in group g can be expressed as $\mathbf{h}_{g_k} = \mathbf{U}_g \boldsymbol{\Lambda}_g^{1/2} \boldsymbol{\eta}_{g_k}$, where $\mathbf{R}_g = \mathbf{U}_g \boldsymbol{\Lambda}_g \mathbf{U}_g^H$ is the eigendecomposition of the channel covariance matrix $\mathbf{R}_g$ of group g, $\mathbf{U}_g$ is the $M \times r_g$ matrix composed of the orthonormal eigenvectors corresponding to the r_g non-zero eigenvalues of $\mathbf{R}_g$, $\boldsymbol{\Lambda}_g$ is the $r_g \times r_g$ diagonal matrix of non-zero eigenvalues, and $\boldsymbol{\eta}_{g_k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{r_g})$.

Denoting by $\mathbf{H}_g = [\mathbf{h}_{g_1}, \cdots, \mathbf{h}_{g_{K_g}}]^H$ the $K_g \times M$ channel matrix for the users in group g, we have the overall $K \times M$ channel matrix $\mathbf{H}$ constructed by stacking $\{\mathbf{H}_g\}$, i.e., $\mathbf{H} = [\mathbf{H}_1^H, \cdots, \mathbf{H}_G^H]^H$. Then, the received signal at all the users in the cell is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the $M \times 1$ transmitted signal vector at BS, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the noise vector, and the BS has an average power constraint $\mathbb{E}[\|\mathbf{x}\|^2] \leq P$. Here, in two-stage beamforming the transmitted signal vector $\mathbf{x}$ is the precoded version of the $S \times 1$ original data vector $\mathbf{d}$ by the product of a $M \times b$ pre-beamformer $\mathbf{V}$ and a $b \times S$ MU-MIMO precoder $\mathbf{W}$, i.e.,

$$\mathbf{x} = \mathbf{V}\mathbf{W}\mathbf{d},$$

where $\mathbf{d} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_S)$. As explained, the pre-beamforming matrix $\mathbf{V}$ is designed based on the channel *statistics* information $\{\mathbf{U}_g, \boldsymbol{\Lambda}_g\}$ but not on the instantaneous CSI. Let the pre-beamforming matrix for group g be the $M \times b_g$ submatrix $\mathbf{V}_g$ such that $\mathbf{V} = [\mathbf{V}_1, \cdots, \mathbf{V}_G]$. Then, the received signal in (1) can be expressed as [2]

$$\mathbf{y} = \mathbf{G}\mathbf{W}\mathbf{d} + \mathbf{n}, \tag{2}$$

where

$$\mathbf{G} := \mathbf{H}\mathbf{V} = \begin{bmatrix} \mathbf{H}_1\mathbf{V}_1 & \mathbf{H}_1\mathbf{V}_2 & \cdots & \mathbf{H}_1\mathbf{V}_G \\ \mathbf{H}_2\mathbf{V}_1 & \mathbf{H}_2\mathbf{V}_2 & \cdots & \mathbf{H}_2\mathbf{V}_G \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_G\mathbf{V}_1 & \mathbf{H}_G\mathbf{V}_2 & \cdots & \mathbf{H}_G\mathbf{V}_G \end{bmatrix}. \tag{3}$$

The MU-MIMO precoder $\mathbf{W}$ is now designed in a block-diagonal form as $\mathbf{W} = \text{diag}(\mathbf{W}_1, \cdots, \mathbf{W}_G)$ for simplicity, where $\mathbf{W}_g$ is the MU-MIMO precoder of size $b_g \times S_g$ and depends only on the *effective* channel $\mathbf{G}_g := \mathbf{H}_g\mathbf{V}_g$ for group g. Consequently, in the two-stage beamforming, the received signal at the users in group g is given by

$$\mathbf{y}_g = \mathbf{G}_g\mathbf{W}_g\mathbf{d}_g + \sum_{g' \neq g} \mathbf{H}_g\mathbf{V}_{g'}\mathbf{W}_{g'}\mathbf{d}_{g'} + \mathbf{n}_g, \tag{4}$$

where $\mathbf{d}_g$ and $\mathbf{n}_g$ are the data and noise vectors for group g, respectively. Let the row-wise decomposition of $\mathbf{G}_g$ and the column-wise decomposition of $\mathbf{W}_g$ be $\mathbf{G}_g = [\mathbf{g}_{g_1}, \cdots, \mathbf{g}_{g_{K_g}}]^H$ and $\mathbf{W}_g = [\mathbf{w}_{g_1}, \cdots, \mathbf{w}_{g_{K_g}}]$, respectively. Then, the received signal of user k in group g is given by

$$y_{g_k} = \mathbf{g}_{g_k}^H \mathbf{w}_{g_k} d_{g_k} + \sum_{k' \neq k} \mathbf{g}_{g_k}^H \mathbf{w}_{g_{k'}} d_{g_{k'}} + \sum_{g' \neq g} \mathbf{h}_{g_k}^H \mathbf{V}_{g'} \mathbf{W}_{g'} \mathbf{d}_{g'} + n_{g_k} \tag{5}$$

where $\mathbf{g}_{g_k}$, $\mathbf{w}_{g_k}$, d_{g_k} and n_{g_k} are the effective channel, the MU-MIMO precoding vector, the data and the noise symbols of user k in group g, respectively. The second and third terms in the right-hand side (RHS) of (5) are the intra-group and inter-group interference, respectively.

Regarding the inter-group interference, we assume that at least the approximate block diagonalization (BD) condition holds: [2]

- Exact BD: Each group has a sufficient signal space to transmit S_g multiple data streams that does not interfere with other groups, i.e., [2]

$$\dim\left(\text{span}(\mathbf{U}_g) \cap \text{span}^{\perp}(\{\mathbf{U}_{g'} : g' \neq g\})\right) \geq S_g. \tag{6}$$

- Approximate BD: When exact BD is not possible, approximate BD can be achieved by selecting a matrix $\mathbf{U}_g^*$ composed of the r_g^* ($\leq r_g$) dominant eigenvectors for each group g such that (r_g^* is a control parameter) [2]

$$\dim\left(\text{span}(\mathbf{U}_g^*) \cap \text{span}^{\perp}(\{\mathbf{U}_{g'}^* : g' \neq g\})\right) \geq S_g. \tag{7}$$

Note that in the case of approximate BD, the inter-group interference still remains in (5) due to the weakest $r_g - r_g^*$ eigenvectors not included in $\{\mathbf{U}_g^*\}$. Note that both $\mathbf{U}_g$ and $\mathbf{U}_g^*$ have column-wise submatrices of a unitary matrix. Thus, the average transmit power for user g_k is $\|\mathbf{w}_{g_k}\|^2$ when $\mathbf{V}_g = \mathbf{U}_g^*$.

III. The Proposed User Scheduling Algorithm

In this section, we propose a user scheduling algorithm for a given pre-beamformer $\mathbf{V} = [\mathbf{V}_1, \cdots, \mathbf{V}_G]$, assuming that ZFBF is used for the second-stage MU-MIMO precoder $\mathbf{W}_g$. For the sake of simplicity, we assume $b_g = S_g = r_g^*$ and $\mathbf{V}_g = \mathbf{U}_g^*$, $\forall\ g$. We also assume that each user g_k (not BS) knows its *effective* CSI $\mathbf{g}_{g_k}$ and the power allocated to group g is proportional to the number of supported users in group g.

The Proposed User Selection Method

0) $\alpha \in (0,\ 1)$ is a pre-determined parameter and is shared by the BS and all UTs. The BS initializes

$$\mathcal{W}_{g,i} = \emptyset, \text{ for } i = 1, \cdots, r_g^* \tag{8}$$

$$\mathcal{S}_g = \emptyset. \tag{9}$$

1) Each user g_k independently computes the following set:

$$\mathcal{I}_{g_k} := \left\{ i : \left| \mathbf{e}_{i_g}^H \frac{\mathbf{g}_{g_k}}{\|\mathbf{g}_{g_k}\|} \right| \geq \alpha,\ i = 1, \cdots, r_g^* \right\}, \tag{10}$$

where $\mathbf{e}_{i_g}$ is the i-th column of $\mathbf{I}_{r_g^*}$.
If user g_k has $\mathcal{I}_{g_k} \neq \emptyset$, then the user finds

$$i_{g_k}^* = \arg\max_{i \in \mathcal{I}_{g_k}} \left| \mathbf{e}_{i_g}^H \frac{\mathbf{g}_{g_k}}{\|\mathbf{g}_{g_k}\|} \right| \tag{11}$$

and feedbacks the pair $(i_{g_k}^*, \mathcal{R}(g_k))$ to the BS, where

$$\mathcal{R}(g_k) := \frac{\|\mathbf{g}_{g_k}\|^2}{\frac{1}{\rho} + \sum_{g' \neq g} \|\mathbf{h}_{g_k}^H \mathbf{V}_{g'}\|^2}, \tag{12}$$

where $\rho := \frac{P}{\sum_{g=1}^G r_g^*}$. If $\mathcal{I}_{g_k} = \emptyset$, user g_k does not feedback. After the feedback, the BS updates $\mathcal{W}_{g,i_{g_k}^*} \leftarrow \mathcal{W}_{g,i_{g_k}^*} \cup \{k\}$ and stores $\mathcal{R}(g_k)$.

2) For $i = 1, \cdots, r_g^*$, the BS finds

$$\kappa_{g,i} = \arg\max_{k \in \mathcal{W}_{g,i}} \mathcal{R}(g_k), \tag{13}$$

and updates

$$\mathcal{S}_g \leftarrow \mathcal{S}_g \cup \{\kappa_{g,i}\}. \tag{14}$$

3) The BS transmits a paging signal to notify that the users in $\mathcal{S}_g$ are scheduled and then, the corresponding UTs feedback their effective CSI to the BS. Finally, the BS constructs the MU-MIMO ZFBF precoder with water-filling power allocation for each group and transmits data streams to the scheduled UTs.

In step 1), each user checks the level of alignment of its channel to each of the r_g^* dominant channel directions in $\mathbf{U}_g^*$ of group g by computing (10). (Note that $\mathbf{g}_{g_k} = [\mathbf{h}_{g_k}^H \mathbf{u}_{g,1}^*, \cdots, \mathbf{h}_{g_k}^H \mathbf{u}_{g,r_g^*}^*]^H$, where $\mathbf{u}_{g,i}^*$ is the i-th column of $\mathbf{U}_g^* = \mathbf{V}$.) If user g_k has a non-empty set $\mathcal{I}_{g_k}$ of channel directions of alignment level α, then user g_k finds the most aligned direction in $\mathcal{I}_{g_k}$ and feedbacks the direction index $i_{g_k}^*$ and the quasi-SINR $\mathcal{R}(g_k)^*$ to the BS. If $\mathcal{I}_{g_k} = \emptyset$, then user g_k does not feedback any information to the BS.

After the quasi-SINR feedback period is over, the BS makes r_g^* candidate sets $\{\mathcal{W}_{g,i}\}$ for r_g^* channel directions for group g, based on the feedback information. Here, $\mathcal{W}_{g,i}$ represents the set of users whose channels are most aligned to the i-th dominant direction of $\mathbf{U}_g^*$ with level larger than or equal to α. In step 2), the BS chooses the user $\kappa_{g,i}$ having the largest quasi-SINR $\mathcal{R}(g_k)$ in each set $\mathcal{W}_{g,i}$, $i = 1, \cdots, r_g^*$, to construct the set $\mathcal{S}_g$ of scheduled users for each group g. As explained in [1], by the use of (10) with the threshold α and the quasi-SINR (12), *we can select semi-orthogonal users with large channel magnitude* as the SUS algorithm in [5].

Step 3) is a beam construction stage based on ZFBF after the selection of semi-orthogonal users with large channel magntigude by steps 1) and 2). Step 3) requires CSI feedback only from the selected users not from all users.

Remark 2. As seen, the proposed method is composed of two steps: The first is the user selection step based on scalar-valued quasi-SINR feedback and the second step is the ZFBF beam construction and data transmission to the selected users based on CSI feedback from the selected users. The amount of feedback required for the proposed method for group g is $\sum_{i=1}^{r_g^*} |\mathcal{W}_{g,i}|$ integers (for user index feedback) and $\sum_{i=1}^{r_g^*} |\mathcal{W}_{g,i}| + 2(r_g^*)^2$ real numbers (for quasi-SINR feedback and later effective CSI feedback). Note that only r_g^* users per group need to feedback their effective CSI of (complex) dimension r_g^* for one scheduling period for the proposed scheme. It can be shown that when $\alpha \leq 1/\sqrt{r_g^*}$, $\mathcal{I}_{g_k}$ is a non-empty set for all g_k and thus, every user feedbacks its quasi-SINR to the BS. Hence, in this case, $\sum_{i=1}^{r_g^*} |\mathcal{W}_{g,i}|$ reduces to K_g. When $\alpha > 1/\sqrt{r_g^*}$, on the other hand, $\mathcal{I}_{g_k} = \emptyset$ for some users and thus in this case, $\sum_{i=1}^{r_g^*} |\mathcal{W}_{g_i}|$ can be less than K_g. In Section V, numerical results show that many users does not feedback to the BS for optimally chosen α. Hence, the feedback overhead can be reduced drastically.

IV. Optimality of The Proposed Method

In this section, we prove the asymptotical optimality of the proposed method as $K \to \infty$. We start with the optimal capacity scaling law for the general K-user MISO broadcast channel consisting of multiple groups with each group's having the same channel covariance matrix.

Theorem 1: [7] In a MU-MISO downlink system consisting of a BS with M antennas and total power constraint P and K single-antenna users divided into G groups of equal size $K' = K/G$, where the channel vector of each user in group g is independent and identically distributed (i.i.d.) from $\mathcal{CN}(\mathbf{0}, \mathbf{R}_g)$ for $g = 1, \cdots, G$, the sum-rate capacity (which is achieved by DPC) scales as

$$R_{DPC} = \beta \log\log(K') + \beta \log \frac{P}{\beta} + O(1) \tag{15}$$

where $\beta = \min\{M, \sum_{g=1}^G r_g\}$ and $O(1)$ denotes a constant, independent of K', as $K' \to \infty$.

Proof: See Theorem 1 in [7]. ■

The same scaling law is achieved by the proposed method under the approximate BD condition.

Theorem 2: In the system described in Theorem 1, the sum-rate of the scheduled sets $\{\mathcal{S}_g\}$ by the proposed method scales as

$$\mathbb{E}\left[\sum_{g=1}^G R_{ZF,g}(\mathcal{S}_g) \right] \sim R_{DPC}, \tag{16}$$

*We refer to $\mathcal{R}(g_k)$ as the quasi-SINR because $\mathcal{R}(g_k)$ is the SINR when there is no intra-group interference. This is valid under the ZF second-stage beamforming. Note that without inter-group interference, the quasi-SINR is simply the channel norm square. For the computation of $\mathcal{R}(g_k)$, please refer to [1], [8].

where $x \sim y$ indicates that $\lim_{K' \to \infty} x/y = 1$. Here, $R_{ZF,g}(\mathcal{S}_g)$ is the sum-rate of the users in $\mathcal{S}_g$ determined by the proposed method with ZF MU-MIMO second-stage precoding.

Proof: Proof of the optimality is by showing both the effective channel gain reduction and the multi-user diversity gain reduction associated with the proposed method become negligible as $K' \to \infty$.

1) Effective channel gain: Since ZFBF is assumed for the second-stage beamforming, we have $\mathbf{W}_g := \mathbf{W}_g(\mathcal{S}_g) = \mathbf{G}_g^H(\mathcal{S}_g)(\mathbf{G}_g(\mathcal{S}_g)\mathbf{G}_g^H(\mathcal{S}_g))^{-1}\mathbf{P}_g$,[†] and $\mathbf{P}_g = \mathrm{diag}(\sqrt{P_{\kappa_{g,1}}}, \cdots, \sqrt{P_{\kappa_{g,r_g^*}}})$, where $P_{\kappa_{g,i}}$ is the transmit power loading factor for the scheduled user $\kappa_{g,i} \in \mathcal{S}_g$. (Since the pseudo-inverse $\mathbf{G}_g^H(\mathcal{S}_g)(\mathbf{G}_g(\mathcal{S}_g)\mathbf{G}_g^H(\mathcal{S}_g))^{-1}$ is given for the given effective channel, we need $\mathbf{P}_g$ to control the user power.) Then, the effective channel gain $\gamma_{\kappa_{g,i}}$ for user $\kappa_{g,i}$ is given by [5]

$$\gamma_{\kappa_{g,i}} = \frac{1}{[(\mathbf{G}_g(\mathcal{S}_g)\mathbf{G}_g(\mathcal{S}_g)^H)^{-1}]_{i,i}}. \tag{17}$$

Since $[\mathbf{G}_g(\mathcal{S}_g)\mathbf{G}_g(\mathcal{S}_g)^H]_{i,j} = \mathbf{g}_{\kappa_{g,i}}^H \mathbf{g}_{\kappa_{g,j}}$, $\forall i, j$, we can decompose it as

$$\mathbf{G}_g(\mathcal{S}_g)\mathbf{G}_g(\mathcal{S}_g)^H = \mathbf{D}\tilde{\mathbf{G}}\mathbf{D}, \tag{18}$$

where $\mathbf{D} = \mathrm{diag}(\|\mathbf{g}_{\kappa_{g,1}}\|, \cdots, \|\mathbf{g}_{\kappa_{g,r_g^*}}\|)$ and

$$\tilde{\mathbf{G}} = \begin{bmatrix} 1 & \tilde{\mathbf{g}}_{\kappa_{g,1}}^H \tilde{\mathbf{g}}_{\kappa_{g,2}} & \cdots & \tilde{\mathbf{g}}_{\kappa_{g,1}}^H \tilde{\mathbf{g}}_{\kappa_{g,r_g^*}} \\ \tilde{\mathbf{g}}_{\kappa_{g,2}}^H \tilde{\mathbf{g}}_{\kappa_{g,1}} & 1 & & \vdots \\ \vdots & & \ddots & \tilde{\mathbf{g}}_{\kappa_{g,r_g^*-1}}^H \tilde{\mathbf{g}}_{\kappa_{g,r_g^*}} \\ \tilde{\mathbf{g}}_{\kappa_{g,r_g^*}}^H \tilde{\mathbf{g}}_{\kappa_{g,1}} & \cdots & \tilde{\mathbf{g}}_{\kappa_{g,r_g^*}}^H \tilde{\mathbf{g}}_{\kappa_{g,r_g^*-1}} & 1 \end{bmatrix} \tag{19}$$

with $\tilde{\mathbf{g}}_{\kappa_{g,i}} = \frac{\mathbf{g}_{\kappa_{g,i}}}{\|\mathbf{g}_{\kappa_{g,i}}\|}$, $\forall i$. Substituting (18) into (17), we have

$$\gamma_{\kappa_{g,i}} = \|\mathbf{g}_{\kappa_{g,i}}\|^2 / [\tilde{\mathbf{G}}^{-1}]_{i,i}. \tag{20}$$

First, consider the term $[\tilde{\mathbf{G}}^{-1}]_{i,i}$. Since $\kappa_{g,i} \in \mathcal{W}_{g,i}$, it is easy to show that $|\tilde{\mathbf{g}}_{\kappa_{g,i}}^H \tilde{\mathbf{g}}_{\kappa_{g,j}}| \le 2\alpha\sqrt{1-\alpha^2}$ for $i \neq j$ when $\alpha > 1/\sqrt{2}$. By the Gershgorin circle theorem [9], every eigenvalue of the Hermitian matrix $\tilde{\mathbf{G}}$ is in a Gershgorin disk

$$\lambda(\tilde{\mathbf{G}}) \in \{z \in \mathbb{R} : |z-1| \le (r_g^*-1)2\alpha\sqrt{1-\alpha^2}\}, \tag{21}$$

provided that $\sum_{j \neq i} |\tilde{\mathbf{g}}_{\kappa_{g,i}}^H \tilde{\mathbf{g}}_{\kappa_{g,j}}| \le (r_g^*-1)2\alpha\sqrt{1-\alpha^2} < 1$, equivalently, $\alpha > \sqrt{\left(1+\sqrt{(r_g^*-2)/(r_g^*-1)}\right)/2}$. Therefore, we have

$$[\tilde{\mathbf{G}}^{-1}]_{i,i} \le [\lambda_{min}(\tilde{\mathbf{G}})]^{-1} \overset{(a)}{\le} \frac{1}{1-(r_g^*-1)2\alpha\sqrt{1-\alpha^2}}, \tag{22}$$

where (a) follows from (21), and $\lambda_{min}(\tilde{\mathbf{G}})$ is the minimum eigenvalue of $\tilde{\mathbf{G}}$. Thus, from (20), the effective channel gain $\gamma_{\kappa_{g,i}}$ is lower bounded by

$$\gamma_{\kappa_{g,i}} > \frac{\|\mathbf{g}_{\kappa_{g,i}}\|^2}{\frac{1}{1-(r_g^*-1)2\alpha\sqrt{1-\alpha^2}}}. \tag{23}$$

[†] $\mathbf{W}_g(\mathcal{S}_g) = [\{\mathbf{w}_{g_k}\}_{k \in \mathcal{S}_g}]$ and $\mathbf{G}_g(\mathcal{S}_g) = [\{\mathbf{g}_{g_k}\}_{k \in \mathcal{S}_g}]^H$ are respectively the submatrices of $\mathbf{W}_g$ and $\mathbf{G}_g$ according to the scheduler output $\mathcal{S}_g$.

As $K' \to \infty$, α can be made arbitrarily close to 1 such that $\mathcal{W}_{g,i}$ is not empty for each i in group g since $\mathbf{h}_{g_k} \overset{i.i.d}{\sim} \mathcal{CN}(0, \mathbf{R}_g)$. As $\alpha \uparrow 1$, the denominator in the RHS of (23) converges to one and $\gamma_{\kappa_{g,i}} \to \|\mathbf{g}_{\kappa_{g,i}}\|^2$, i.e., there is no loss in the effective channel gain.

2) Multi-user diversity gain: Define

$$\phi_{g_k}^i = \begin{cases} \mathcal{R}(g_k), & k \in \mathcal{W}_{g,i}, \\ 0, & \text{Otherwise} \end{cases} \tag{24}$$

for $k = 1 \cdots, K_g$. Then, for a given i, the random variable $\phi_{g_k}^i$ is i.i.d. across k in the same group g. Note that

$$\kappa_{g,i} = \arg\max_{k \in \mathcal{W}_{g,i}} \mathcal{R}(g_k) = \arg\max_{k \in \{1, \cdots, K_g = K'\}} \phi_{g_k}^i.$$

The multi-user diversity gain results from choosing the best user among all users with i.i.d. channel realizations. However, with the proposed method, for each data stream, the best user within $\mathcal{W}_{g,i}$ is chosen, and thus there exists some loss in the multi-user diversity gain. However, for the proposed method, with the approximate BD condition we can show for each i

$$\Pr\{\phi_{\kappa_{g,i}}^i > u_g^i\} \ge 1 - O(1/K'), \tag{25}$$

where $u_g^i = (\lambda_{g,1}\log K' - \lambda_{g,1}\log\log K' + a_i)/(1/\rho + c)$, $\lambda_{g,1}$ is the maximum eigenvalue of $\mathbf{R}_g$, and a_i and c are constants independent of K'. Proof (25) is based on the extreme value theory applied to $\phi_{g_k}^i$ and is omitted due to the lack of space. (Please see [1] for detail.)

3) Finally, similarly to [5], from (5) the general sum-rate formula for the ZF MU-MIMO broadcast channel consisting of users $\{\kappa_{g,1}, \cdots, \kappa_{g,r_g^*}\}$ with power distribution $\{P_{\kappa_{g,1}}, \cdots, P_{\kappa_{g,r_g^*}}\}$ can be derived as

$$R_{ZF,g}(\mathcal{S}_g) = \max_{\{P_{\kappa_{g,i}}\}} \sum_{i=1}^{r_g^*} \log\left(1 + \frac{P_{\kappa_{g,i}}}{1 + \sum_{g' \neq g} \|\mathbf{h}_{g_{\kappa_{g,i}}}^H \mathbf{V}_{g'}\mathbf{W}_{g'}\|^2}\right)$$
$$\text{s.t.} \quad \sum_{i=1}^{r_g^*} \gamma_{g_{\kappa_{g,i}}}^{-1} P_{\kappa_{g,i}} \le r_g^* \cdot \rho, \tag{26}$$

where $\|\mathbf{w}_{\kappa_{g,i}}\|^2 = \gamma_{\kappa_{g,i}}^{-1} P_{\kappa_{g,i}}$. Now, we show (16) :

$$\begin{aligned}
&\mathbb{E}\left[\sum_{g=1}^{G} R_{ZF,g}(\mathcal{S}_g)\right] \\
&\overset{(a)}{\ge} \mathbb{E}\left[\sum_{g=1}^{G}\sum_{i=1}^{r_g^*} \log\left(1 + \frac{\rho\gamma_{\kappa_{g,i}}}{1 + \sum_{g' \neq g} \|\mathbf{h}_{\kappa_{g,i}}^H \mathbf{V}_{g'}\mathbf{W}_{g'}\|^2}\right)\right] \\
&\overset{(b)}{\ge} \mathbb{E}\left[\sum_{g=1}^{G}\sum_{i=1}^{r_g^\star} \log\left(1 + \frac{\|\mathbf{g}_{\kappa_{g,i}}\|^2\left(1-(r_g^\star-1)2\alpha\sqrt{1-\alpha^2}\right)}{\frac{1}{\rho} + \sum_{g' \neq g} \|\mathbf{h}_{\kappa_{g,i}}^H \mathbf{V}_{g'}\|^2}\right)\right] \\
&\overset{(c)}{\ge} \sum_{g=1}^{G}\sum_{i=1}^{r_g^\star} \Pr\{\phi_{\kappa_{g,i}}^i > u_g^i\} \log\left(1 + u_g^i(1-(r_g^\star-1)2\alpha\sqrt{1-\alpha^2})\right) \\
&\overset{(d)}{\ge} \sum_{g=1}^{G}\sum_{i=1}^{r_g^\star} \left[1 - O\left(\frac{1}{K'}\right)\right] \log\left(1 + u_g^i(1-(r_g^\star-1)2\alpha\sqrt{1-\alpha^2})\right) \\
&\overset{(e)}{\sim} \sum_{g=1}^{G}\sum_{i=1}^{r_g^*} \log\left(1 + \left(\frac{1-(r_g^*-1)2\alpha\sqrt{1-\alpha^2}}{1/\rho + c}\right)\lambda_{g,1}\log K'\right) \\
&\overset{(f)}{\sim} \left(\sum_{g=1}^{G} r_g^*\right)\log\rho + \sum_{g=1}^{G} r_g^* \log\lambda_{g,1} + \left(\sum_{g=1}^{G} r_g^*\right)\log\log K'
\end{aligned} \tag{27}$$

where (a) follows from the suboptimal equal power allocation (simple equal power allocation for each data stream means $\rho = \frac{P}{\sum_{g=1}^{G} r_g^*} = \|\mathbf{w}_{\kappa_{g,i}}\|^2 = \gamma_{\kappa_{g,i}}^{-1} P_{\kappa_{g,i}}, \ \forall g, i.$); (b) is obtained by (23) and submultiplicativity of norm; (c) holds by $\mathbb{E}f(X) = \int_0^\infty f(x)p(x)dx \geq \Pr(X \geq u)f(u)$ for a monotone increasing function f; (d) holds by (25); and (e) and (f) are obtained by simple manipulation. Hence, in both the cases of $\sum_{g=1}^{G} r_g < M$ and $\sum_{g=1}^{G} r_g \geq M$, we can choose r_g^* such that $\sum_{g=1}^{G} r_g^* = \min\{M, \sum_{g=1}^{G} r_g\} = \beta$. Then, (27) is the same as (15). ■

V. Numerical Results

We evaluated the performance of the proposed method. We considered a MISO broadcast system where a BS was equipped with a ULA of $M = 32$ antenna elements and the covariance matrices of K UTs were modelled according to the one-ring model [2], [8]. Each user k had angle-of-arrival (AoA) θ_k, angle-spread (AS) Δ_k and $\mathbf{R}_k$. We generated AoA and AS according to $\theta_k \sim \text{Unif}[-\pi/3, \pi/3]$ and $\Delta_k \sim \text{Unif}[\pi/36, \pi/12]$ for each user independently. We set $G = 8$, $\frac{M}{G} = r_g^* = b_g \geq S_g$, $P = 15$ [dB], and the pre-beamforming matrix $\mathbf{V}_g$ is the $M \times b_g$ matrix such that $\mathbf{V} = [\mathbf{V}_1, \cdots, \mathbf{V}_G]$ is the $M \times M$ DFT matrix [2], [8]. Each user k was associated with the group $g^\star$ such that $g^\star = \arg\min_g \|\mathbf{U}_k^* \mathbf{U}_k^{*H} - \mathbf{V}_g \mathbf{V}_g^H\|_F^2$[‡]. After partitioning users into groups, we evaluated the sum rate performance (averaged over 20 iterations) of original SUS [5], RBF in [7], and the proposed method, as K varied, and the result is shown in Fig. 1. As shown in Fig. 1, the proposed method shows better performance than the two existing user scheduling methods: RBF and orignal SUS. Note that original SUS has a worse slope than the others. This is because SUS does not take inter-group interference into account. Thus, we modified SUS by using the metric in (12) to take the inter-group interference into account. As expected, the modified SUS shows the best performance since it exploits full CSI at the BS. The reason why RBF in [7] shows worse performance is explained in [1] in detail. Briefly speaking, there are mainly two reasons for the bad performance of RBF. The first is the metric used for RBF itself and the second is the lack of post-user-selection beam refinement. The two drawbacks of RBF were effectively corrected for the proposed method by using the metric (12) and applying post-user-selection ZFBF. Fig. 2 shows the amount of feedback (the number of real numbers) required for SUS, RBF, and the proposed method for the same setting as that for Fig. 1. The feedback overhead of the proposed method is far less than that of SUS and less than that of RBF.

VI. Conclusion

In this paper, we have proposed an efficient user scheduling method for massive MIMO downlink systems. We have shown that the proposed method is asymptotically optimal and yields comparable performance of SUS (with the modified metric) and far better performance than RBF with feedback overhead less than RBF. Although the proposed method is derived for two-stage beamforming with multiple groups in a cell, the same method can be applied to single-stage conventional small-scale MU-MIMO too. Furthermore, the proposed method can easily be modified to implement fairness among users by adopting round-robin or proportional fairness principle.

[‡]Details about user grouping are described in [7].

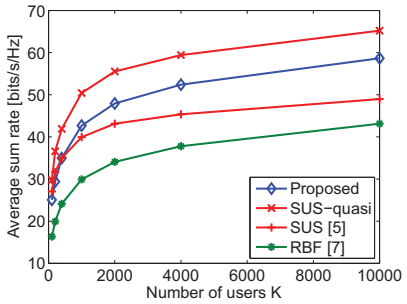

Fig. 1. Average sum rate performance (optimal values of α were used for each K for the proposed method)

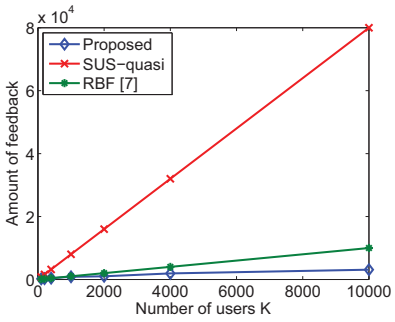

Fig. 2. Amount of feedback

References

[1] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user MIMO broadcast channels," *submitted to IEEE Trans. Inf. Theory*, Mar. 2014.

[2] A. Adhikary, J. Nam, J. Ahn and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, pp. 6441–6463, Oct. 2013.

[3] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Transactions on Information Theory*, vol. 51, pp. 506–522, Feb. 2005.

[4] G. Dimíc and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, pp. 3857 – 3868, Oct. 2005.

[5] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 528–541, Mar. 2006.

[6] T. Al-Naffouri, M. Sharif and B. Hassibi, "How much does transmit correlation affect the sum-rate scaling of MIMO Gaussian broadcast channels?," *IEEE Transactions on Communications*, vol. 57, pp. 562–572, Feb. 2009.

[7] A. Adhikary and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming and user grouping," *arXiv preprint arXiv:1305.7252*, 2013.

[8] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *accepted to IEEE Journal of Selected Topics in Signal Processing*, Dec. 2013.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.