Research Article

A Comparison of Scheduling Strategies for MIMO Broadcast Channel with Limited Feedback on OFDM Systems

Ermanna Conte, Stefano Tomasin, and Nevio Benvenuto

Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy

Correspondence should be addressed to Ermanna Conte, ermanna.conte@gmail.com

We consider a multiuser downlink transmission from a base station with multiple antennas (MIMO) to mobile terminals (users) with a single antenna, using orthogonal frequency division multiplexing (OFDM). Channel conditions are reported by a feedback from users with limited rate, and the base station schedules transmissions and beamforms signals to users. We show that an important set of schedulers using a general utility function can be reduced to a scheduler maximizing the weighted sum rate of the system. For this case we then focus on scheduling methods with many users and OFDM subcarriers. Various scheduling strategies are compared in terms of achieved throughput and computational complexity and a good tradeoff is identified in greedy and semiorthogonal user selection algorithms. In the greedy selection algorithm, users are selected one by one as long as the throughput increases, while in the semiorthogonal approach users are selected based on the channel correlation. An extension of these approaches from a flat-fading channel to OFDM is considered and simplifications that may be useful for a large number of subcarriers are presented. Results are reported for a typical cellular transmission of the long-term evolution (LTE) of 3GPP.

1. Introduction

Next generation wireless cellular systems are expected to support high-quality multimedia services; this motivates the interest in multiantenna (MIMO) systems, where both spatial diversity and multiplexing can be used to increase the achievable throughput. In fact, it has been shown that the downlink capacity of a MIMO system with perfect channel state information (CSI) scales as a linear function of the number of transmit antennas [1]. Although nonlinear dirty paper coding scheme achieves the system capacity, it has a high computational cost [2], and simpler solutions have been investigated. Linear beamforming has been shown [3] to achieve a large part of dirty paper coding capacity; in particular, zero forcing beamforming matched to an opportunistic scheduling is widely used [3].

However, benefits of MIMO are obtained only by a proper scheduling of transmissions, which opportunistically exploits channel conditions in order to increase throughput, while ensuring quality of service (QoS). Several scheduling techniques have been proposed for MIMO single carrier systems on flat fading channels based on various approaches, including clique search [4], maximization of the Frobenius norm of the composite channel matrix [5, 6], user channel orthogonality [7–9], single bit feedback [10], waterfilling [11], tree search [12], evolutionary algorithms [13], and greedy scheduling [14] extended to the case of limited feedback in [15]. In some cases, joint optimization of scheduling and power allocation is performed [4–6, 10, 11, 13], while in other cases only scheduling is considered [7–9, 14]. Moreover, QoS-oriented multiuser scheduling and beamforming have been investigated in [16], in order to conciliate the request of high throughput with low packet delays. An overview of research on cross-layer scheduling for multiuser MIMO single-carrier systems is given in [17]. A similar problem to multiuser MIMO scheduling can be found in other transmission systems, such as multicarrier code- or frequency-division multiple access [18].

In frequency selective channels, single carrier modulation is often replaced by orthogonal frequency division multiplexing (OFDM) due to its efficiency in overcoming multipath fading. In fact, the combination of MIMO and

OFDM seems to be the technology of future wireless cellular systems, as it has been proposed for downlink in the long term evolution (LTE) release of 3GPP standard [19, 20]. When MIMO OFDM is considered, scheduling becomes more complex, as the number of resources to be allocated, that is, the number of subcarriers, increases and only suboptimal approaches are viable [12]. Complexity is further increased in a frequency division duplexing system, where CSI is provided to the base station by each user (mobile terminal) through a feedback channel. In fact, due to the limited feedback rate, only a partial CSI is available at the base station and additional processing is required to compensate the channel uncertainty. Some of the scheduling techniques considered for single carrier transmissions can be extended to OFDM. For example, in [21] a scheduling algorithm has been proposed for MIMO OFDM systems which extends method [14] for single-carrier systems: the set of scheduled users on each subcarrier is built in a greedy fashion, by adding one user at a time with the aim of maximizing a weighted sum rate (WSR). In [22] this approach has been further simplified to avoid the need of computing a new beamforming matrix upon the insertion of a new candidate in the set of scheduled users. A further simplification of the scheduling is achieved by computing an estimate on their signal to interference ratio which is then used to exclude users that would not contribute to the WSR, by introducing a threshold to their signal to noise plus interference ratio.

In this paper, we first show that any scheduler maximizing a wide class of utility functions can be reduced to a scheduler maximizing the weighted sum rate, where the weights are suitably chosen according to the utility function. Then we revise scheduling techniques proposed in the literature that maximize the weighted sum rate for a multiuser MIMO OFDM system with limited feedback and compare them in a LTE 3GPP scenario in terms of (i) computational complexity, (ii) memory requirements, and (iii) achievable throughput.

The rest of the paper is organized as follows. In Section 2 we describe the downlink MIMO OFDM system model. In Section 3 a general scheduling method is derived and algorithms [14, 21] are revised. Sections 4 and 5 present, respectively, the user selection and user preselection strategies of [22]. In Section 6 the complexity of the various strategies is investigated. In Section 7 simulation results are illustrated and Section 8 outlines main conclusions.

Notation. Bold upper and lower letters denote matrices and vectors, respectively; $(\cdot)^H$ denotes Hermitian operation (transpose complex conjugate), while $(\cdot)^T$ denotes transpose; $\|\cdot\|$ is the vector norm, and $\mathbb{E}[\cdot]$ stands for expectation.

2. System Model

We consider the downlink of a cellular system based on OFDM [23] with N_C subcarriers. The base station has M transmit antennas while K users have one antenna each. Transmission is performed in time slots of L OFDM symbols

and in each time slot users feedback a partial CSI, which is used by the base station to schedule downlink transmissions. The transmission bandwidth is divided into N resource blocks, each composed of N_S subcarriers. At slot n, let $\mathcal{U}_c(n) = \{u_{1,c}(n), u_{2,c}(n), \ldots, u_{|\mathcal{U}_c(n)|,c}(n)\}$ be the set of $|\mathcal{U}_c(n)|$ users, $u_{i,c}(n) \in \{1, \ldots, K\}$, scheduled for downlink transmission on resource block $c \in \{1, \ldots, N\}$. We denote as *stream* the (user, resource block) pair (k, c). Let also $\mathcal{P}(n)$ be the set of streams scheduled at slot n, that is,

$$\mathcal{P}(n) = \{(k,c) \mid k \in \mathcal{U}_c(n), c \in \{1,\ldots,N\}\}. \quad (1)$$

In our analysis we model the channel as quasi static, that is, it is considered invariant for the duration of one OFDM symbol, and it has the same frequency response on all subcarriers of each resource block. Hence, the frequency response of the MIMO channel on resource block c of OFDM symbol t for all M transmit antennas and all $|\mathcal{U}_c(n)|$ users is described by the complex $|\mathcal{U}_c(n)| \times M$ channel matrix $\mathbf{H}_c(t) = [\mathbf{h}_{1,c}(t), \ldots, \mathbf{h}_{|\mathcal{U}_c(n)|,c}(t)]^T$, where the $M \times 1$ column channel vector $\mathbf{h}_{k,c}(t)$ collects the gains between the M antennas of the base station and stream (k, c). In general, for OFDM symbol t, $\mathbf{x}_c(t, m)$, and $\mathbf{y}_c(t, m)$ are, respectively, the $M \times 1$ and $|\mathcal{U}_c(n)| \times 1$ column vectors of the transmitted and received signals on subcarrier m of resource block c. The discrete-time complex baseband transmission model for subcarrier m of resource block c is given by

$$\mathbf{y}_c(t,m) = \mathbf{H}_c(t)\mathbf{x}_c(t,m) + \mathbf{n}_c(t,m), \quad m = 1, \ldots, N_S, \quad (2)$$

where $\mathbf{n}_c(t,m)$ is a $|\mathcal{U}_c(n)| \times 1$ complex Gaussian noise vector with i.i.d. components having zero mean and variance σ_n^2. The transmit signal is subject to the total power constraint $\mathbb{E}[\sum_{c=1}^{N}\sum_{m=1}^{N_S}\|\mathbf{x}_c(t,m)\|^2] \leq P$ where P is the available power. In order to exploit spatial diversity, the transmit signal is obtained from the $|\mathcal{U}_c(n)| \times 1$ data signal $\mathbf{d}_c(t, m)$ by applying the zero-forcing beamforming matrix $\mathbf{G}_c(nL)$, that is,

$$\mathbf{x}_c(t,m) = \mathbf{G}_c(nL) \, \mathrm{diag}\, \{\mathbf{p}_c(nL)\}^{1/2}\mathbf{d}_c(t,m),$$
$$nL \leq t < (n+1)L, \quad (3)$$

where $\mathbf{p}_c(nL)$ is the power normalization vector which enforces equal stream power as given by

$$P_S = \mathbb{E}\left[\left\|\mathbf{x}_{k,c}(t,m)\right\|^2\right] = \frac{P}{\sum_{c=1}^{N}|\mathcal{U}_c(n)|}, \quad (4)$$

where the expectation is taken only with respect to $\mathbf{d}_c(t,m)$, and $\mathbf{x}_{k,c}(t,m)$ is the kth entry of $\mathbf{x}_c(t,m)$.

2.1. Feedback Information. In a frequency division duplexing system, channel state information is provided through a feedback channel; therefore, we assume that matrix $\mathbf{H}_c(t)$ is not perfectly known at the base station while user k perfectly estimates the channel vectors once for time slot, that is, when $t = nL$, to obtain $\mathbf{h}_{k,c}(nL)$, $c = 1, 2, \ldots, N$. As in [24] we adopt a double feedback information for all the users at each slot. In particular, at slot n user k feeds back for each resource

block c: (i) a channel direction information (CDI) $\overline{\mathbf{h}}_{k,c}$, which ideally tracks the normalized channel vector $\widetilde{\mathbf{h}}_{k,c}(nL)$, namely

$$\widetilde{\mathbf{h}}_{k,c}(nL) = \frac{\mathbf{h}_{k,c}(nL)}{||\mathbf{h}_{k,c}(nL)||} \quad c = 1,\ldots,N, \quad (5)$$

and (ii) a channel quality information (CQI), based on the estimated signal-to-noise plus interference ratio (SNIR) at the receiver for M orthogonal scheduled users evaluated as [24]

$$\xi_{k,c} \triangleq \frac{||\mathbf{h}_{k,c}(nL)||^2 \left| \widetilde{\mathbf{h}}_{k,c}^T(nL)\overline{\mathbf{h}}_{k,c}^*(nL) \right|^2}{\sigma_n^2 + (P/M \cdot N)||\mathbf{h}_{k,c}(nL)||^2 \left(1 - \left| \widetilde{\mathbf{h}}_{k,c}^T(nL)\overline{\mathbf{h}}_{k,c}^*(nL) \right|^2 \right)}. \quad (6)$$

We assume that the feedback channel has a finite rate of N_b bits per slot and per user and allows zero-delay error-free transmission. The base station builds the matrix

$$\overline{\mathbf{H}}_c(nL) = \left[\overline{\mathbf{h}}_{u_{1,c}}(nL),\ldots,\overline{\mathbf{h}}_{u_{|\mathcal{U}_c(n)|,c}}(nL) \right]^T, \quad u_{i,c} \in \mathcal{U}_c(n) \quad (7)$$

containing the unit-norm *reconstructed* channel vectors $\overline{\mathbf{h}}_{k,c}(nL)$. Using the partial CSI, base station evaluates an estimate $\gamma_{k,c}(n)$ of the SNIR of stream (k,c) as will be seen in Section 4. Zero-forcing beamforming with equal power distribution among streams is implemented for each resource block, hence the beamforming matrix is

$$\mathbf{G}_c(nL) = \overline{\mathbf{H}}_c^H(nL)\left(\overline{\mathbf{H}}_c(nL)\overline{\mathbf{H}}_c^H(nL)\right)^{-1}. \quad (8)$$

An estimate of the normalized (with respect to the bandwidth) rate achieved by stream $(k,c) \in \mathcal{P}(n)$ at slot n is

$$R_{k,c}(n, \mathcal{P}(n)) = \log_2(1 + \gamma_{k,c}(n)). \quad (9)$$

Notation $R_{k,c}(n, \mathcal{P}(n))$ highlights the fact that rates achieved by different streams are mutually dependent, as (i) more streams allocated simultaneously on the same resource blocks yield interference, and (ii) the total power is distributed among active streams. Performance is evaluated in terms of WSR,

$$\mathcal{R}(\mathcal{P}(n)) = \sum_{(k,c)\in\mathcal{P}(n)} w_k(n)R_{k,c}(n, \mathcal{P}(n)), \quad (10)$$

with $w_k(n)$ suitable weights that take into account fairness and QoS constraints.

2.2. Exhaustive Search Scheduling.
At each slot, we aim at scheduling the set of streams that maximizes WSR.

This problem can be solved by considering all $(\sum_{i=1}^{M}(K!/i!(K-i)!))^N$ possible sets and evaluating the WSR achieved by each candidate set. Unfortunately, this *exhaustive search* (ES) scheduling has a high computational cost, which becomes infeasible for an increasing number of users and subcarriers. Simpler and suboptimal scheduling methods are investigated in Section 4.

3. Maximum Utility Scheduler

In order to balance the opportunistic use of channel resources with fairness among users, we consider a multiuser scheduler. We first consider in this section general criteria for the choice of weights of the WSR and we derive the optimum maximum utility scheduler weights for a general utility function. Then we specialize the result for the maximum sum rate scheduler and the proportional fair scheduler.

3.1. General Multiuser Scheduling.
Let $R_k(n, \mathcal{P})$ be the achievable rate associated with user k, that is, $R_k(n, \mathcal{P}) = \sum_{c=1}^{N} R_{k,c}(n, \mathcal{P})$. In the first slot, the average throughput achieved by user k is $T_k(1) = 0$, $k = 1, 2, \ldots, K$. The estimate of the average throughput achieved by user k can be updated as

$T_k(n+1)$

$$= \begin{cases} (1-\alpha_k)T_k(n) + \alpha_k R_k(n, \mathcal{P}(n)) & k \in \bigcup_{c=1}^{N}\mathcal{U}_c(n), \\ (1-\alpha_k)T_k(n) & k \notin \bigcup_{c=1}^{N}\mathcal{U}_c(n), \end{cases} \quad (11)$$

where $\bigcup_{c=1}^{N}\mathcal{U}_c(n)$ is the set of scheduled users at slot n. If we aim at achieving an average throughput ρ_k for user k, we can define the normalized averaged throughput at slot n as

$$B_k(n) = \frac{T_k(n)}{\rho_k}. \quad (12)$$

In [25], the following concave and differentiable utility function has been proposed to design schedulers

$$U_k(B_k(n)) = \frac{\rho_k(1-\alpha_k)^\kappa}{\alpha_k}\frac{1}{1-\kappa}\left(B_k^{1-\kappa}(n) - 1\right), \quad (13)$$

where $\kappa \in [0,1) \cup (\infty)$ is a fairness parameter to be chosen according to the desired scheduling policy. For example, for $\kappa \to 1$ we obtain the *proportional fair* scheduler (PFS). For $\kappa = 0$ we obtain the utility function of the *maximum sum rate* scheduler. When $\kappa \to \infty$, (13) becomes the utility function of the *max-min* scheduler.

In Appendix A, we derive the set of users $\mathcal{P}(n)$ that maximizes the sum utility

$$\sum_{k=1}^{K} U_k(B_k(n)). \quad (14)$$

The solution is

$$\mathcal{P}(n) = \arg\max_{\mathcal{l}\subseteq\mathcal{C}}\mathcal{R}(\mathcal{l}) = \arg\max_{\mathcal{l}\subseteq\mathcal{C}} \sum_{(k,c)\in\mathcal{l}} w_k(n)R_{k,c}(n, \mathcal{l}), \quad (15)$$

with weights

$$w_k(n) = \sum_{c:(k,c)\in\mathcal{l}} \frac{R_{k,c}(n, \mathcal{l})\rho_k^\kappa}{T_k(n)^\kappa}, \quad (16)$$

where $\mathcal{C} = \{(k,c) : k \in \{1,\ldots,K\}, c \in \{1,\ldots,N\}\}$ is the set of all possible streams. Note that for $K = 1$, (16) boils down to the maximum utility scheduler of [25].

3.2. Maximum Sum Rate Scheduling. The *maximum sum rate* scheduler does not consider the fairness among users ($\kappa = 0$) and simply aims at maximizing the achievable sum rate (SR), providing $w_k = 1$, for $k = 1, \ldots, K$, and

$$\mathcal{P}(n) = \arg\max_{\mathcal{I} \subseteq \mathcal{C}} R(\mathcal{I}) = \arg\max_{\mathcal{I} \subseteq \mathcal{C}} \sum_{(k,c) \in \mathcal{I}} R_{k,c}(n, \mathcal{I}). \quad (17)$$

3.3. Proportional Fair Scheduling. The multiuser multicarrier *proportional fair* scheduling (MMPFS) algorithm [26] is an extension to the OFDM multiuser scenario of the PFS algorithm.

For MMPFS, the average throughput of user k is updated as in (11) with $\alpha_k = 1/\tau$, where τ is a parameter related to the time over which fairness should be achieved. In [27] it has been shown that proportional fairness, maximizing $\sum_k \log_2 T_k(n)$, is achieved by scheduling users as

$$\mathcal{P}(n) = \arg\max_{\mathcal{I} \subseteq \mathcal{C}} \sum_{(k,c) \in \mathcal{I}} \log_2\left(1 + \frac{R_{k,c}(n, \mathcal{I})}{(\tau - 1)T_k(n)}\right). \quad (18)$$

We observe that for $\tau \gg 1$ we can approximate

$$\log_2\left(1 + \frac{R_{k,c}(n, \mathcal{I})}{(\tau - 1)T_k(n)}\right) \approx \frac{R_{k,c}(n, \mathcal{I})}{(\tau - 1)T_k(n)} \quad (19)$$

and MPFS (18) coincides with the maximization of the WSR (15) with weights (16), $\rho_k = 1/(\tau - 1)$ and $\kappa = 1$.

4. Greedy Scheduling Strategies

Now that we have established that maximizing the weighted sum rate is equivalent to the maximization of a wide set of utility functions, we focus on methods that allow to achieve this goal. In the following we investigate suboptimal solutions to problem (15) for a small number of users K, when the probability of having a fully loaded system is small. In fact, in this scenario power distribution has an important role in selecting the optimal user set. In Section 4.3 we will consider the case of a high number of users K, and in this case a simplification of scheduling is possible. For ease of notation we drop both slot (n) and OFDM symbol (t) index in the remaining of the paper.

4.1. Multicarrier Greedy (MG). In [14], a greedy scheduling algorithm in a single-carrier flat-fading system has been proposed, where users are selected one by one as long as the throughput increases and it has been then extended to an OFDM system in [21] and denoted here multicarrier greedy (MG).

The MG algorithm comprises N_{step} steps, and at each step we select the stream that maximizes the increase of WSR. Let $\mathcal{S}^{(i)}$ be the set of streams scheduled for transmission at step i, ($i = 1, \ldots, N_{\text{step}}$), with the corresponding WSR $\mathcal{R}(\mathcal{S}^{(i)})$. Initially we have $\mathcal{S}^{(0)} = \emptyset$. The stream selected at step $i + 1$, is

$$\left(\bar{k}, \bar{c}\right) = \arg\max_{(k,c) \in \mathcal{C} \setminus \mathcal{S}^{(i)}} \mathcal{R}\left(\mathcal{S}^{(i)} \cup \{(k,c)\}\right), \quad (20)$$

and we set $\mathcal{S}^{(i+1)} = \mathcal{S}^{(i)} \cup \{(\bar{k}, \bar{c})\}$. The WSR $\mathcal{R}(\mathcal{S}^{(i)})$ increases at each step, since stream (k, c) is inserted under the condition that

$$\mathcal{R}\left(\mathcal{S}^{(i)} \cup \left\{\left(\bar{k}, \bar{c}\right)\right\}\right) \geq \mathcal{R}\left(\mathcal{S}^{(i)}\right). \quad (21)$$

When (21) does not hold, the algorithm is stopped, $N_{\text{step}} = i$ and $\mathcal{P} = \mathcal{S}^{(N_{\text{step}})}$. Hence, N_{step} is a random variable. Evaluation of the WSR in (20) for the current set of streams is based on the observation that for an equal power allocation across users and resource blocks under constraint (4), using (3), the allocated power to user k and resource block c turns out to be [15]

$$p_{k,c} = \frac{P}{(i + 1) \cdot \left\|\mathbf{g}_{j,c}^{(i)}\right\|^2}. \quad (22)$$

Therefore, a SNIR estimate for stream $(j, c) \in \mathcal{S}^{(i)} \cup \{(k, c)\}$ with $(k, c) \in \mathcal{C} \setminus \mathcal{S}^{(i)}$ can be computed by scaling with the newly computed transmit power the CQI value, that is,

$$\gamma_{j,c}^{(i+1)} = \frac{P}{(i + 1) \cdot \left\|\mathbf{g}_{j,c}^{(i)}\right\|^2} \xi_{j,c}, \quad (23)$$

where $\xi_{j,c}$ is given by (6) while $\mathbf{g}_{j,c}^{(i)}$ is the jth column of the beamforming matrix $\mathbf{G}_c^{(i)}$ for users scheduled at step i. Note that total power P has been divided by $|\mathcal{S}^{(i+1)}| = i+1$ in order to obtain the per stream power P_S.

4.2. Projection-Based Greedy (PBG). According to the MG algorithm, the introduction of a new candidate stream $(\bar{k}, \bar{c})$ into the set $\mathcal{S}^{(i)}$ at step $i + 1$ decreases the SNIRs (23) for two reasons:

(1) the power is redistributed among all streams;

(2) beamforming of streams already scheduled on the same resource block is modified.

Due to (a), it is beneficial to perform scheduling jointly among resource blocks rather than separately on each resource block. Due to (b), a new beamforming matrix must be computed for users scheduled on resource block $\bar{c}$ of the candidate stream. Hence, at each step many beamformers must be designed for each resource block to test (21) and only one candidate stream is then scheduled. In order to reduce the computational complexity, the projection-based greedy (PBG) algorithm [22] assumes that the insertion of a new stream does not significantly alter the SNIR of already scheduled streams. Indeed, this assumption holds as long as the channel vector of the candidate stream is almost orthogonal to channel vectors of previously scheduled streams. Therefore, we update the SNIR estimate of already scheduled streams as follows

$$\gamma_{p,q}^{(i+1)} = \frac{i}{i+1}\gamma_{p,q}^{(i)} \quad (p,q) \in \mathcal{S}^{(i)}, \quad (24)$$

for $i = 2, 3, \ldots, N_{\text{step}} - 1$, while for the first step we set $\gamma_{p,q}^{(1)} = \xi_{p,q}, (p,q) \in \mathcal{S}^{(1)}$. Furthermore, the evaluation of the SNIR

for the candidate streams requires only the computation of $\left\|\mathbf{g}_{k,c}^{(i)}\right\|^2$ instead of the entire beamformer. In particular, if we define

$$a_{k,c}\left(\mathcal{S}^{(i)}\right) = \frac{1}{\left\|\mathbf{g}_{k,c}^{(i)}\right\|^2}, \tag{25}$$

from (23) we have

$$\gamma_{k,c}^{(i+1)} = \frac{P}{i+1}\, \xi_{k,c} a_{k,c}\left(\mathcal{S}^{(i)}\right) \quad (k,c) \in \mathcal{C} \setminus \mathcal{S}^{(i)}. \tag{26}$$

In order to compute (25) and the corresponding SNIR (26) of the candidate stream (k,c), it can be observed that its beamforming vector is obtained by the orthogonalization of $\overline{\mathbf{h}}_{k,c}$ with respect to the normalized channel vector of already scheduled streams on the same resource block. Hence, an orthonormal basis $\mathcal{B}_c(i) = \{\overline{\mathbf{b}}_{j,c}\}$ is first constructed for the space generated by the channel vectors $\{\overline{\mathbf{h}}_{p,c}\}$ of streams in $\mathcal{S}^{(i)}$ on resource block c. Then the beamforming vector for stream (k,c) would be proportional to

$$\mathbf{g}_{k,c}^{(i)} \propto \overline{\mathbf{h}}_{k,c} - \sum_{j=1}^{|\mathcal{B}_c(i)|} \left(\overline{\mathbf{h}}_{k,c} \cdot \overline{\mathbf{b}}_{j,c}^H\right)\overline{\mathbf{b}}_{j,c}. \tag{27}$$

Now, by imposing $\mathbf{G}_c^{(i)}\overline{\mathbf{H}}_c = \mathbf{I}$, the identity matrix, it is $\overline{\mathbf{h}}_{k,c}^H \mathbf{g}_{k,c}^{(i)} = 1$ and we have

$$a_{k,c}\left(\mathcal{S}^{(i)}\right) = \left|1 - \sum_{j=1}^{|\mathcal{B}_c(i)|} \left|\overline{\mathbf{h}}_{k,c} \cdot \overline{\mathbf{b}}_{j,c}^H\right|^2\right|^2. \tag{28}$$

By using (26) and (28), there is no need to determine a new beamformer in correspondence of each candidate stream; instead, only the basis $\mathcal{B}_c(i)$ needs to be updated at each step, and this requires only few vector multiplications. Note that the computation of $a_{k,c}$ is based on the projection of the candidate vector on the basis, as from the acronym PBG. Once all streams have been scheduled, a beamformer is computed to perform transmission.

4.3. Greedy Scheduling Strategies in the High K Scenario. If $K \gg M$, multiuser diversity provides M orthogonal streams on each resource block with very high probability, thus we will have almost always a fully loaded system, that is, $|\mathcal{U}_c| = M$. In this case, both MG and PBG algorithms can be simplified without redistributing the available power at each new insertion, and the per stream power (4) becomes

$$P_S = \frac{P}{N \cdot M}. \tag{29}$$

Scheduling can then be simplified by operating independently on each resource block.

4.4. Multicarrier Semiorthogonal User Selection Algorithm (MSUS). The semiorthogonal user selection (SUS) scheme [9] can be easily generalized to the OFDM scenario and is here denoted as multicarrier SUS (MSUS). The generalization includes also the maximization of the WSR instead of the SR as considered in [9]. MSUS proceeds by steps now applied separately on each resource block. For resource block c, let $\mathcal{A}_c^{(1)} = \{1,\dots,K\}$ be the initial set containing the indexes of all users. The scheduled stream at step 1 is characterized by having maximum CQI, that is,

$$k_1 = \arg\max_{k \in \mathcal{A}_c^{(1)}} w_k \cdot \log_2\left(1 + \xi_{k,c}\right). \tag{30}$$

After selecting i streams, the $(i+1)$th stream k_{i+1} is chosen within the set

$$\mathcal{A}_c^{(i+1)} = \left\{k \in \mathcal{A}_c^{(i)} \setminus k_{(i)}, c: \left|\hat{\mathbf{h}}_{k,c}\hat{\mathbf{h}}_{k_{j,c},c}^H\right| \le \epsilon, 1 \le j \le i\right\}$$
$$i = 2,\dots,M \tag{31}$$

as

$$k_{i+1,c} = \arg\max_{k \in \mathcal{A}_c^{(i+1)}} w_k \log_2\left(1 + \xi_{k,c}\right), \tag{32}$$

where ϵ is a design parameter that sets the maximum correlation allowed between the quantized channel vectors of the selected users. We note that in MSUS we apply N single carrier SUS in parallel, one for each resource block. Also in this case the number of steps is random as the algorithm ends when set $\mathcal{A}_c^{(i)}$ is empty. Once users have been scheduled, the total power is equally distributed among the scheduled streams according to (4).

5. Preselection Methods

In the MG algorithm the WSR $\mathcal{R}(\mathcal{S}^{(i)})$ increases at each step and using (9) and (10), condition (21) becomes

$$\sum_{(p,q)\in\mathcal{S}^{(i+1)}} w_p \log_2\left(1 + \gamma_{p,q}^{(i+1)}\right) \ge \sum_{(p,q)\in\mathcal{S}^{(i)}} w_p \log_2\left(1 + \gamma_{p,q}^{(i)}\right). \tag{33}$$

From (23) we obtain that this condition is satisfied only if the SNIR is high enough to compensate for losses incurred by the insertion of a new scheduled stream, that is, the power redistribution and the beamforming modification, as described by conditions (1) and (2) of Section 4.2. This observation suggests a further simplification of the PBG algorithm, by a-priori excluding the streams whose SNIR is below a certain threshold. Indeed, as for each candidate stream the SNIR (26) must be evaluated, by excluding streams that could never be inserted, the scheduling procedure can be fastened [22].

Note that the idea of preselecting users has been first introduced in [28], by letting users feeding back their CSI and rate request only if the quality of their channel is above a threshold. On the other hand, we use preselection as a technique to simplify scheduling rather than reducing the feedback rate. Moreover, in our case the preselection is not based only on the channel quality but also on the correlation with other users' channels.

5.1. Preselection PBG (PPBG). We first observe from (28) that $a_{k,c}(\delta^{(i)}) \le 1$ and from (25) we obtain

$$\gamma_{k,c}^{(i+1)} \le \xi_{k,c} \frac{P}{i+1}. \tag{34}$$

Therefore, at step i of PBG there is a minimum value of $\xi_{k,c}$ that satisfies (33), denoted $A(i+1)$, and we consider for scheduling only streams having SNIR

$$\xi_{k,c} > A(i+1). \tag{35}$$

As shown Appendix B, at high SNR we have

$$A(i+1) \approx \sqrt[w_k]{\prod_{(p,q)\in\delta^{(i)}} \left(\frac{i+1}{i}\right)^{w_p} \frac{(i+1)}{P}}. \tag{36}$$

Then by considering only streams (k,c) satisfying (35), we decrease the number of comparisons and SINR updates at each step of PBG. In the high K scenario the preselection technique is not feasible; in fact, as illustrated in Appendix B, $A(i) \to 0$ for $K \to \infty$, and therefore (35) is verified by all streams.

We further note that $A(i)$ is an increasing function of i; hence, streams whose CQI is below the threshold $A(i)$ at step i can be neglected also in the next steps.

5.2. Simplified Preselection PBG (S-PPBG). A further simplification in preselection is achieved by neglecting weights w_p on evaluating $A(i)$, that is, by considering $w_k \approx w_p \approx 1$ in (B.2) to yield

$$A(i+1) \approx \frac{(i+1)}{P}\left(\frac{i+1}{i}\right)^i. \tag{37}$$

Within PBG methods, we note that this approach becomes optimal when the scheduling objective coincides with the maximization of the SR. However, for the maximization of the WSR, S-PPBG is in general suboptimal.

6. Complexity Analysis

We analyze the worst case complexity of the various approaches, in terms of both computational complexity and memory requirement.

6.1. Computational Complexity. We assume that a comparison yields a computational complexity equal to λ complex multiplications (CMUX), while the inversion of an $M \times M$ matrix performed by Gaussian elimination methods has complexity $M \cdot (M^2 - 1)/3$ CMUX. The beamforming and $\|\mathbf{g}_{k,c}\|^2$ evaluation have therefore complexity

$$\text{BFC} = 2|\mathcal{U}_c| \cdot M^2 + \frac{M \cdot (M^2 - 1)}{3} + |\mathcal{U}_c| \cdot M. \tag{38}$$

We first observe that all considered algorithms select one stream per step, until at most M streams are allocated on each resource block, thus in general $N_{\text{step}} \le N \cdot M$. At step i, $|\mathcal{C} \setminus \delta^{(i)}| = K \cdot N - i$ streams are considered for insertion in $\delta^{(i+1)}$. Furthermore, at each step, the per stream power P_S is adapted, due to the insertion of a candidate stream in $\delta^{(i+1)}$.

6.1.1. MG Complexity. Complexity of the MG algorithm in the low K scenario is given by

$$\begin{aligned}
\mathcal{C}_{\text{MG-low}K} = &\; \lambda \cdot N \cdot K \\
&+ \sum_{i=2}^{N_{\text{step}}} (i-1) \\
&+ \left\{ \left(K - \left|\mathcal{U}_{\zeta(i-1)}^{(i)}\right|\right) \cdot (\text{BFC} + 2) \right. \\
&\left. + \lambda(K \cdot N - i + 1) + \lambda \right\},
\end{aligned} \tag{39}$$

where $\zeta(i-1)$ denotes the resource block of the stream selected at step $i-1$. The first term in (39) accounts for the selection of the stream with maximum CQI. The remaining terms account for step 2 through N_{step}, with (a) update of SNIR estimate of the $(i-1)$ already scheduled streams, (b) computation of a new beamformer for each of the $(K - |\mathcal{U}_{\zeta(i-1)}^{(i)}|)$ candidate streams on subcarrier $\zeta(i-1)$, (c) evaluation of $\|\mathbf{g}_{k,\zeta(i-1)}\|^2$, (d) update of the SNIR estimates, and (e) evaluation of the WSR. Lastly, the algorithm determines the stream which maximizes the WSR at step i and checks condition (21).

In the high K scenario complexity becomes

$$\begin{aligned}
\mathcal{C}_{\text{MG-high}K} = &\; \lambda \cdot N \cdot K + N \cdot \sum_{i=2}^{M} \{(K - i + 1) \cdot (\text{BFC} + 2) \\
&+ \lambda(K - i + 1) + \lambda\},
\end{aligned} \tag{40}$$

since now $N_{\text{step}} = M$ and no power update is necessary at each step.

6.1.2. PBG Complexity. Complexity of the PBG in the low K scenario is

$$\begin{aligned}
\mathcal{C}_{\text{PBG}} = &\; \lambda \cdot N \cdot K + N \cdot \text{BFC} \\
&+ \sum_{i=2}^{N_{\text{step}}} \left\{ (i-1) + \left(K - \left|\mathcal{U}_{\zeta(i-1)}^{(i)}\right|\right) \cdot (M+2) \right. \\
&\left. + \left|\mathcal{U}_{\zeta(i-1)}^{(i)}\right| + 2 \cdot M + \lambda \cdot (K \cdot N - i + 1) + \lambda \right\}.
\end{aligned} \tag{41}$$

In fact, the PBG algorithm for each candidate stream on resource block $\zeta(i-1)$ (a) performs the projection of channel vector on the orthogonal basis and (b) updates the SNIR estimate. At each step, the basis is also updated according to the channel vector of last scheduled stream. At the end, the beamforming matrix is computed according to the set of scheduled streams.

In the high K scenario we have

$$\mathcal{C}_{\text{PBG-high}K} = \lambda \cdot N \cdot K + N \cdot \text{BFC}$$

$$+ N \cdot \sum_{i=2}^{M} \{(M+2) \cdot (K-i+1) + i + 2 \cdot M$$

$$+ \lambda \cdot (K-i+1) + \lambda\}, \quad (42)$$

since scheduling can be performed in parallel on all resource blocks.

6.1.3. PPBG Complexity. The complexity of the PPBG in the low K scenario is given by

$$\mathcal{C}_{\text{PPBG}} = \lambda \cdot N \cdot K + N \cdot \text{BFC}$$

$$+ \sum_{i=2}^{N_{\text{step}}} \left\{ (i-1) + \left(K - \left| \mathcal{U}_{\zeta(i-1)}^{(i)} \right| \right) \cdot (M+2) \right.$$

$$+ \left| \mathcal{U}_{\zeta(i-1)}^{(i)} \right| + 2 \cdot M + 1 \quad (43)$$

$$\left. + \lambda \cdot (K \cdot N - i + 1) + \lambda \right\}.$$

It only differs from PBG in the evaluation of $A(i+1)$ at each step, since it depends on the set of scheduled streams. Similarly, in the high K scenario we have

$$\mathcal{C}_{\text{PBG-high}K} = \lambda \cdot N \cdot K + N \cdot \text{BFC}$$

$$+ N \cdot \sum_{i=2}^{M} \{(M+2) \cdot (K-i+1) + i + 2 \cdot M + 1$$

$$+ \lambda \cdot (K-i+1) + \lambda\}. \quad (44)$$

6.1.4. S-PPBG Complexity. Applying the S-PPBG algorithm, we have an additional cost due to (35); on the other hand, on resource block c, at each step i we exclude a number of streams $Q_c(i)$ from the set of possible streams. $Q_c(i)$ takes into account also the scheduled streams. Then at step i we have $J_{i,c} = K - \sum_{j=1}^{i} Q_c(i)$ candidate streams on resource block c and in total $J_i = |\mathcal{C} \setminus \mathcal{S}^{(i)}| = \sum_{c=1}^{N} J_{i,c}$. Complexity becomes

$$\mathcal{C}_{\text{SPPBG}} = \lambda \cdot N \cdot K + N \cdot \text{BFC}$$

$$+ \sum_{i=2}^{N_{\text{step}}} \left\{ (i-1) + \left(J_{i,\zeta(i-1)} - \left| \mathcal{U}_{\zeta(i-1)}^{(i)} \right| \right) \cdot (M+2) \right.$$

$$\left. + \left| \mathcal{U}_{\zeta(i-1)}^{(i)} \right| + 2 \cdot M + \lambda \cdot (2J_i + 1) \right\}. \quad (45)$$

Note that $Q_c(i)$ is a random variables depending on the channel realization. In the high K scenario we still consider power adjustment; otherwise, from (B.2), we could never

exclude streams, and then S-PPBG would become PBG. Complexity of S-PPBG in the high K scenario becomes

$$\mathcal{C}_{\text{SPPBG-high}K} = \lambda \cdot N \cdot K + N \cdot \text{BFC}$$

$$+ \sum_{i=2}^{N_{\text{step}}} \{(i-1) + (J_{i,\zeta(i-1)} - i + 1) \cdot (M+2) \quad (46)$$

$$+ i + 2 \cdot M + \lambda \cdot (2J_i + 1)\}.$$

The MSUS algorithm is equivalent to N SUS algorithms working in parallel. We remind that at each step SUS considers $|\mathcal{A}_c^{(i)}| = K - i - \sum_{j=1}^{i} Q(i)$ candidate users, where $Q(i)$ is the number of users excluded at step i. It is

$$\mathcal{C}_{\text{MSUS}} = N \cdot \text{BFC}$$

$$+ \sum_{c=1}^{N} \left(\lambda \cdot K + \sum_{i=2}^{M} \left(\left| \mathcal{A}_c^{(i-1)} \right| \cdot i + \lambda \cdot \left| \mathcal{A}_c^{(i-1)} \right| \right) \right.$$

$$\left. + \left| \mathcal{A}_c^{(i)} \right| (\lambda + 1) \right). \quad (47)$$

6.2. Asymptotic Complexity Analysis. According to complexities required by various scheduling algorithms, we investigate their asymptotic behavior as a function of K. For MG we have

$$\mathcal{C}_{\text{MG-}\infty} \approx K[\lambda \cdot N + N(M-1)(\text{BFC} + 2 + \lambda)] + \mathcal{O}(K). \quad (48)$$

For PBG and PPBG we have

$$\mathcal{C}_{\text{PBG-}\infty} = \mathcal{C}_{\text{PPBG-}\infty}$$
$$\approx K[\lambda \cdot N + N(M-1)(M+2+\lambda)] + \mathcal{O}(K). \quad (49)$$

Both S-PPBG and MSUS perform the exclusion of worse streams. Let β_i be the percentage of streams excluded at step i, for S-PPBG it is $J_{i,c} = (K-i) \cdot (1-\beta_i)$ while for MSUS $\mathcal{A}_c^{(i)} = (K-i) \cdot (1-\beta_i)$. Asymptotic expressions are

$$\mathcal{C}_{\text{SPPBG-}\infty} \approx K[\lambda \cdot N + D \cdot (M+2) + \lambda \cdot 2 \cdot D \cdot N]$$
$$+ \mathcal{O}(K),$$

$$\mathcal{C}_{\text{MSUS-}\infty} \approx K[N \cdot \lambda + B + \lambda \cdot (M-1) \cdot C + (\lambda + 1) \cdot D]$$
$$+ \mathcal{O}(K), \quad (50)$$

where $B = \sum_{i=2}^{M} i(1 - \beta_{i-1})$, $C = \sum_{i=2}^{M}(1 - \beta_{i-1})$, and $D = \sum_{i=2}^{M}(1 - \beta_i)$.

6.3. Memory Occupation. Lastly we investigate memory requirements of the scheduling algorithms in terms of complex location (CLS) units. We first note that all algorithms store (a) CDI and CQI of all streams, (b) the set of selected streams, and (c) the final beamformers; then a memory occupation of $M_{\text{COMM}} = N \cdot M \cdot K + K \cdot N + N \cdot M^2 + M \cdot N$ CLS is common to all algorithms. For MG we have

$$\mathcal{M}_{\text{MG}} = M_{\text{COMM}} + K \cdot N + M \cdot N + K + M^2 \cdot K + 2 \quad (51)$$

since MG stores (a) $\gamma_{j,c}$ (or, equivalently, $\|\mathbf{g}_{j,c}\|^2$), requiring $K \cdot N$ CLS, (b) per user rates ($N \cdot M$ CLS as worst case), (c) new beamformer ($K \cdot M^2$ CLS), (d) total rate provided by each candidate (K CLS), and (e) current and last final rates (2 CLS). For PBG and PPBG we have

$$
\begin{aligned}
\mathcal{M}_{\text{PPBG}} &= \mathcal{M}_{\text{PBG}} \\
&= M_{\text{COMM}} + K \cdot N + M \cdot N + K + M^2 \cdot N + 2;
\end{aligned}
\tag{52}
$$

as PBG stores (a) the value $\sqrt{a_{k,c}}$, (b) total rate provided by each candidate stream (K CLS), and (c) orthogonal basis ($M^2 \cdot N$ CLS).

The S-PPBG memory requirement is given by

$$
\begin{aligned}
\mathcal{M}_{\text{SPPBG}} &= M_{\text{COMM}} + \cdot K \cdot N + M \cdot N + K \\
&\quad + M^2 \cdot N + 2 + M \cdot N;
\end{aligned}
\tag{53}
$$

with respect to PBG it needs to store also $A(i)$ ($M \cdot N$ CLS as worst case).

Finally, for MSUS we have

$$
\mathcal{M}_{\text{MSUS}} = M_{\text{COMM}} + M \cdot N + N \cdot K + 1 + K \cdot N
\tag{54}
$$

as MSUS stores (a) correlations of candidate streams and last inserted stream ($N \cdot K$ CLS), (b) the value of ϵ (1 CLS), and (c) the set of total rates of each candidate ($K \cdot N$ CLS as worst case).

7. Simulation Results

We compare the scheduling algorithms in terms of average sum rate (SR) and complexity requirements. All users are uniformly distributed in a cell of radius 500 m, as in [29]; we consider an average *SNR* of 15 dB per resource block at the cell border and path loss is included in the channel model. We assume also a realistic MIMO channel with time, frequency, and spatial correlation among the elements of $\mathbf{H}_c(t)$. The channel is modeled as slowly time-variant, frequency selective Rayleigh fading as from the spatial channel model (SCM) [30]. According to the LTE release, we set transmission bandwidth to 2.5 MHz, divided into $N = 12$ resource blocks and centered at the carrier frequency of 2 GHz. The base station is equipped with $M = 4$ antennas spaced by 10 wavelength. Scheduling and beamforming are performed once a slot, and each slot is composed of 7 adjacent OFDM symbols. CSI feedback is performed with a variable number of bits using an optimized codebook, as detailed in [31].

7.1. Performance Comparison. We first compare the SR achieved by MG with ES scheduling using as optimization criterion the maximum sum rate. For complexity reasons simulations have been limited to $N = 4$ resource blocks. To simplify simulations in the ES method, results of both MG and ES in the high K scenario, $K = 18N, 20N$, refer to $N = 1$. In fact, we verified that for high K the system is fully loaded with a probability higher then 95%; in this scenario the power granted to each carrier is P/N, and
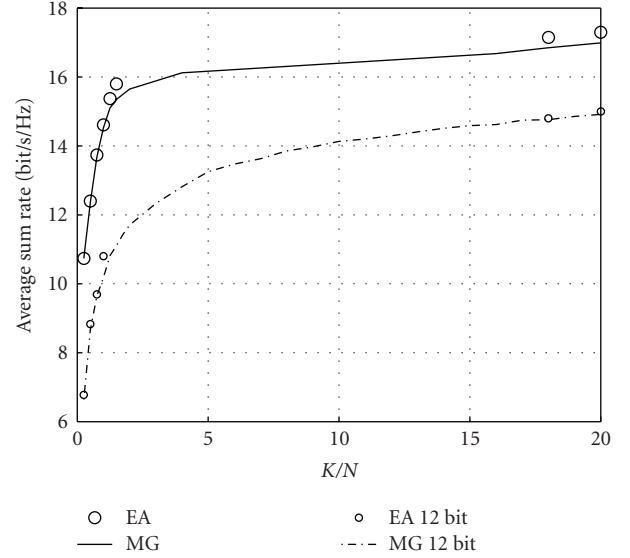

FIGURE 1: Average sum rate of MG and ES versus of total number of users K.

then user selection can be performed independently on each carrier. We consider both the case of perfect CSI at the transmitter and the case of partial CSI obtained by feedback from the receiver, with a feedback rate of 12 bit/user/resource block/slot. We observe that partial CSI provides a loss on SR of 2 up to 3.5 bit/user/resource block/slot, but it does not affect the general behavior of the two algorithms. As we can see from Figure 1, both MG and ES have a very close sum rate for all K. Hence, in the following we consider MG as performance bound.

Figure 2 illustrates the average SR achieved by the scheduling algorithms as a function of the number of users K in the low K scenario for a feedback rate of 12 bit/user/resource block/slot. We note a negligible loss in performance of the simplified methods. Similarly, simulations in the high K scenario show that MG, PBG, and S-PPBG achieve a SR of 16.40 bit/s/Hz, while MSUS provides 15.40 bit/s/Hz. Overall we observe that the simplified algorithms do not provide SR loss for all K. This is mainly due to the fact that all scheduling methods are based on an opportunistic approach, so they all aim at selecting the best set of orthogonal users. We also note that all algorithms always select the same first stream, whose channel vector in turn determines the choice of the other streams. We underline that the average SR of S-PPBG is very close to that of PBG and MG; moreover, since S-PPBG is an approximation of PPBG, we deduce that also PPBG provides the same SR of S-PPBG. Figure 3 confirms this behavior also with a PFS.

We note also in Figure 3 that preselection applied to PBG provides slightly better performance, despite the fact that it considers a lower number of candidate sets. In fact, preselection aims at excluding from scheduling streams that would not increase the WSR and prevents the scheduler from inserting them for fairness reasons.
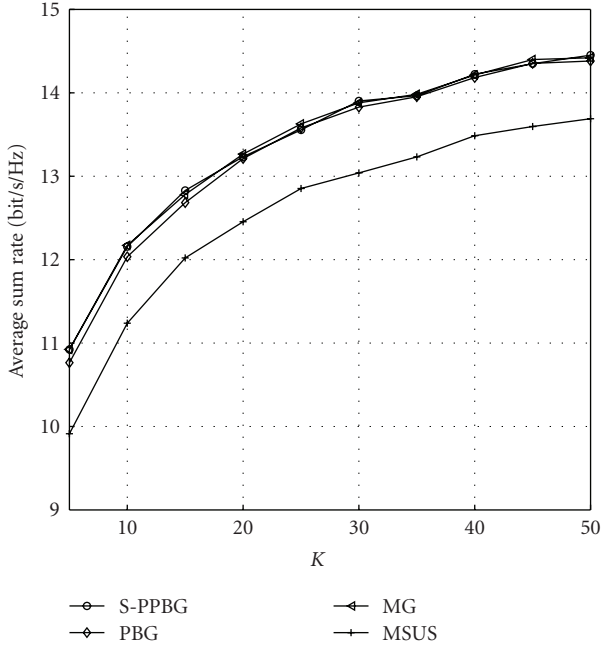
FIGURE 2: Average sum rate as versus the total number of users K. Scheduling objective is the maximization of the sum rate.
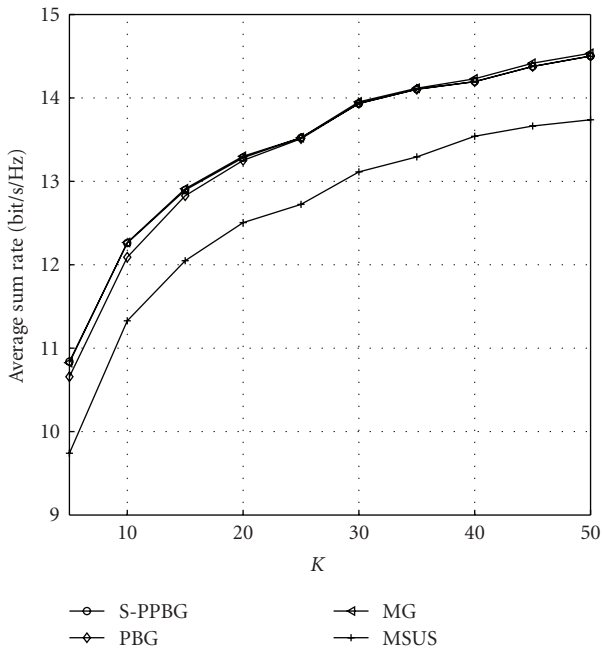

FIGURE 3: Average sum rate versus the total number of users K. Scheduling objective is proportional fair scheduling.

FIGURE 4: Average sum rate versus per stream feedback rate.

FIGURE 5: Complexity versus K.

Figure 4 reports the average SR versus the feedback rate; we observe that the simplified methods are also robust to quantization error; in fact, for all considered values of feedback rate, PBG and S-PPBG provide the same SR of MG. Note that a feedback rate of 12 bit/user/RB/slot would result in an extremely large feedback overhead for the cases with a high number of users (960 bit/RB/slot for $K = 80$ users),
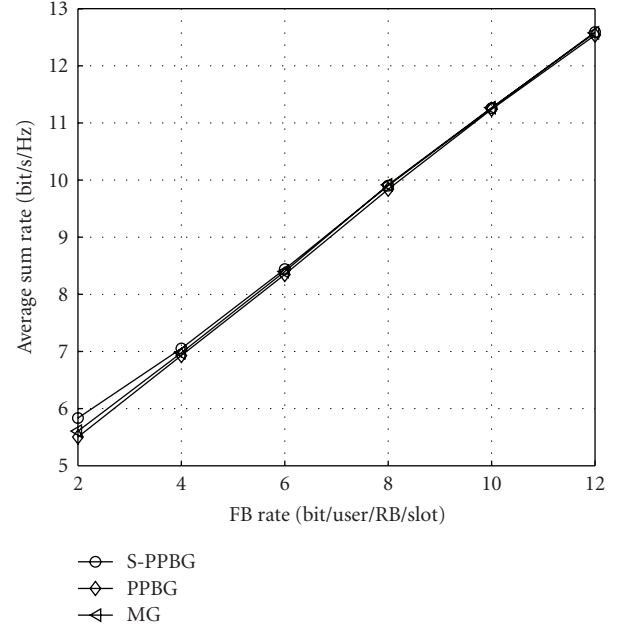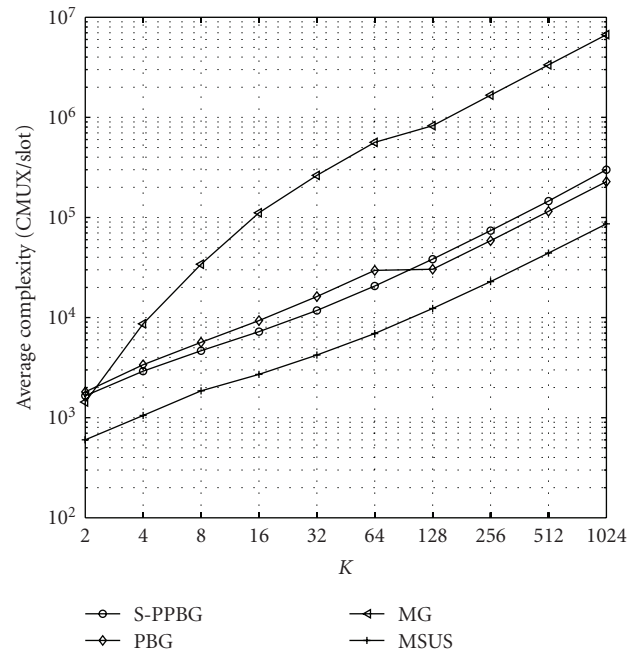
while performance decreases markedly at lower feedback rates.

7.2. Complexity Comparison. Figure 5 shows complexity versus K. For $K = 2$ to 64 the low K complexity expressions are used, while from $K = 128$ to 1024 we use the high K complexity expressions. We first observe that the complexity

ratio between the scheduling algorithms is nearly the same both in the low K and high K regime. As expected, MSUS and S-PPBG complexity trend is not influenced by the value of K. From Figure 5 we note that for $K = 5 \div 50$, with corresponding fully load probability in the range from 1% to 95%, the computational cost of MG is from 2.2 to 18.5 times the cost of PBG, with a factor increasing in K; as expected, the preselection technique further reduces complexity by a factor 1.2–1.4 with respect to PBG. We note also that complexity of S-PPBG is only 2.4–2.9 times the complexity of MSUS. As complexity of PPBG is bounded between that of PBG and S-PPBG and these two are very close, we omitted to show PPBG in Figure 5.

In the high K scenario, simulations confirm the analysis; in fact, for $K = 400$ we have $\mathcal{C}_{\mathrm{MG}} = 2.61 \cdot 10^6$, $\mathcal{C}_{\mathrm{PBG}} = 9.4 \cdot 10^4$, $\mathcal{C}_{\mathrm{MSUS}} = 3.49 \cdot 10^4$, and $\mathcal{C}_{\mathrm{SPPBG}} = 11.9 \cdot 10^4$. We underline that in the high K regime S-PPBG complexity is higher than that of PBG because of the required power distribution; indeed simplification of preselection does not compensate the need of redistributing the total power. On the other hand, we note that the high complexity required by MG is mainly due to the evaluations of the beamformer at each step.

Memory requirements, investigated in Section 6, do not prefigure large differences between different methods; for $K = 400$ required memory locations are 35890 for MG, 29682 for PBG, 29730 for S-PPBG, and 33841 for MSUS. Hence, the simplified techniques achieve a reduction of memory requirement with respect to existing algorithms.

8. Conclusions

This paper has provided an overview of scheduling problems for multiuser downlink MIMO OFDM systems. We first have shown that scheduling according to a wide class of utility functions can be reduced to a scheduling problem aiming at maximizing the weighted sum rate of the system, under a proper choice of the weighting function. Then we have compared scheduling algorithm having as objective the maximization of the weighted sum rate, including greedy algorithms, based on throughput maximization and algorithms based on the semiorthogonality among MIMO channels. Extensions to a OFDM scenario of algorithms originally devised for flat-fading single-carrier systems have been investigated. The comparison has been carried out both in terms of computational complexity and in terms of achievable throughput.

Several insights on the performance of the state of the art scheduling algorithms can be highlighted from the numerical results. Firstly, the MG approach achieves an average sum rate which is very close to the maximum value achieved by ES, over a wide range of cell loads. When compared against MSUS, the proposed MG technique has a gain of about 50% in terms of average sum rate in most network conditions. Moreover, MG requires a significantly lower complexity than that of ES and only 30% additional CMUXs than MSUS. Hence, we believe that MG provides a good trade-off between performance and complexity.

Lastly, limitations in the feedback rate have a severe impact on the performance of all scheduling approaches. Indeed we have seen that all schedulers yield an average sum rate that increases linearly with the number of bits used to feedback the CSI with an increase of about 1 bit/s/Hz for each additional feedback bit.

Appendices

A. Proof of (15)

We aim at solving

$$\mathcal{P}(n) = \arg\max_{\boldsymbol{\mathit{l}} \subseteq \mathcal{C}} \sum_{k=1}^{K} U_k(B_k(n)). \tag{A.1}$$

From (12) and (11), the problem (A.1) can be rewritten as

$$\mathcal{P}(n) = \arg\max_{\boldsymbol{\mathit{l}} \subseteq \mathcal{C}} \sum_{k:(k,c)\in\boldsymbol{\mathit{l}}} U_k\left[\frac{1}{\rho_k}\delta_k(\boldsymbol{\mathit{l}})\alpha_k R_k(n,\boldsymbol{\mathit{l}}) \right.$$
$$\left. + \frac{1}{\rho_k}(1-\alpha_k)T_k(n)\right], \tag{A.2}$$

where δ_k indicates that user k is scheduled, that is, $\delta_k(\boldsymbol{\mathit{l}}) = 1$ if $(k,c) \in \boldsymbol{\mathit{l}}$ and $\delta_k(\boldsymbol{\mathit{l}}) = 0$ otherwise.

Following the derivations of [25] we observe that for all but the scheduled users, the allocated rate at slot n is zero, therefore we have

$$\mathcal{P}(n) = \arg\max_{\boldsymbol{\mathit{l}} \subseteq \mathcal{C}} \sum_{k:(k,c)\in\boldsymbol{\mathit{l}}} U_k\left[\frac{1}{\rho_k}\alpha_k R_k(n,\boldsymbol{\mathit{l}}) \right.$$
$$\left. + \frac{1}{\rho_k}(1-\alpha_k)T_k(n)\right] \tag{A.3}$$
$$- U_k\left[\frac{1}{\rho_k}(1-\alpha_k)T_k(n)\right].$$

Under the assumption $(1-\alpha_k)T_k(n) \gg \alpha_k R_k(n,\boldsymbol{\mathit{l}})$, the following approximation holds

$$U_k\left[\frac{1}{\rho_k}\alpha_k R_k(n,\boldsymbol{\mathit{l}}) + \frac{1}{\rho_k}(1-\alpha_k)T_k(n)\right]$$
$$- U_k\left[\frac{1}{\rho_k}(1-\alpha_k)T_k(n)\right] \tag{A.4}$$
$$\approx \frac{\alpha_k}{\rho_k}R_k(n,\boldsymbol{\mathit{l}})\left.\frac{\partial U_k(x)}{\partial x}\right|_{(1-\alpha_k)T_k(n)/\rho_k}.$$

In turn, from (13) the derivative can be written as

$$\left.\frac{\partial U_k(x)}{\partial x}\right|_{(1-\alpha_k)T_k(n)/\rho_k} = \frac{\rho_k}{\alpha_k \rho_k^{-\kappa}}T_k(n)^{-\kappa}. \tag{A.5}$$

Therefore, by inserting (A.5) into (A.4) we obtain (15).

B. Proof of (36)

We observe that condition (33) is equivalent to

$$
\left[1 + \frac{P}{i+1} \xi_{k,c} a_{k,c}\left(\mathcal{S}^{(i)}\right) \right]^{w_k} \prod_{(p,q)\in\mathcal{S}^{(i)}} \left[1 + \gamma_{p,q}^{(i+1)} \right]^{w_p}
$$
$$
> \prod_{(p,q)\in\mathcal{S}^{(i)}} \left[1 + \gamma_{p,q}^{(i)} \right]^{w_p} \tag{B.1}
$$

where (k,c) is the generic candidate stream.

In the high SNR scenario, with $\xi_{p,q} \gg 1$, we have $1 + \gamma_{p,q}^{(i)} \approx \gamma_{p,q}^{(i)}$ and from (24), condition (B.1) becomes

$$
\left(\frac{P}{i+1} \xi_{k,c} a_{k,c}\left(\mathcal{S}^{(i)}\right) \right)^{w_k} \prod_{(p,q)\in\mathcal{S}^{(i)}} \left(\frac{P}{i+1} \right)^{w_p} > \prod_{(p,q)\in\mathcal{S}^{(i)}} \left(\frac{P}{i} \right)^{w_p}. \tag{B.2}
$$

Hence from (B.2), (36) follows. We note that, in the high K scenario, (B.1) becomes

$$
\left[1 + P_S \xi_{k,c} a_{k,c}(\mathcal{S}^{(i)}) \right]^{w_k} > 1 \tag{B.3}
$$

and $A(i) = 0$.

Acknowledgment

References

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.

[2] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[3] 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, no. 3, pp. 528–541, 2006.

[4] T. Yoo and A. Goldsmith, "Sum-rate optimal multi-antenna downlink beamforming strategy based on clique search," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 3, pp. 1510–1514, St. Louis, Mo, USA, November-December 2005.

[5] 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 Transactions on Signal Processing*, vol. 54, no. 9, pp. 3658–3663, 2006.

[6] J. Wang, D. J. Love, and M. D. Zoltowski, "User selection for the MIMO broadcast channel with a fadiness constraint," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 3, pp. 9–12, Honolulu, Hawaii, USA, April 2007.

[7] H. Viswanathan and K. Kumaran, "Rate scheduling in multiple antenna downlink wireless systems," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Allerton, Ill, USA, October 2001.

[8] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Allerton, Ill, USA, October 2004.

[9] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, 2007.

[10] J. Diaz, O. Simeone, and Y. Bar-Ness, "Sum-rate of MIMO broadcast channels with one bit feedback," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1944–1948, Seattle, Wash, USA, July 2006.

[11] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of Gaussian MIMO-BC," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1640–1646, 2006.

[12] M. Fuchs, G. D. Galdo, and M. Haardt, "Low-complexity space-time-frequency scheduling for MIMO systems with SDMA," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2775–2784, 2007.

[13] M. Jiang, F. Rubio, Y. Wang, J. Gomez, and D. Yuan, "User selection for maximum sum-rate in multi-user and MISO system with evolutionary algorithm," in *Proceedings of the 1st International Workshop on Cross Layer Design (IWCLD '07)*, pp. 74–77, Jinan, China, September 2007.

[14] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, 2005.

[15] M. Trivellato, F. Boccardi, and F. Tosato, "User selection schemes for MIMO broadcast channels with limited feedback," in *Proceedings of the 65th IEEE Vehicular Technology Conference (VTC '07)*, pp. 2089–2093, Dublin, Ireland, April 2007.

[16] A. Alexiou, J. Reis, and A. Gameiro, "QoS-based multiuser scheduling in MIMO systems," in *Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 2, pp. 837–841, Berlin, Germany, September 2005.

[17] C. Anton-Haro, P. Svedman, M. Bengtsson, A. Alexiou, and A. Gameiro, "Cross-layer scheduling for multi-user MIMO systems," *IEEE Communications Magazine*, vol. 44, no. 9, pp. 39–45, 2006.

[18] P. Trifonov, E. Costa, A. Filippi, and E. Schulz, "Adaptive coding in MC-CDMA/FDMA systems with adaptive sub-band allocation," *European Transactions on Telecommunications*, vol. 15, no. 3, pp. 207–214, 2004.

[19] "Overview of the 3GPP long term evolution physical layer," Motorola, January 2003, http://www.freescale.com/files/wireless_comm/doc/white_paper/3GPPEVOLUTIONWP.pdf.

[20] H. Ekstrom, A. Furuskar, J. Karlsson, et al., "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, vol. 44, no. 3, pp. 38–45, 2006.

[21] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Joint low-rate feedback and channel quantization for the MIMO broadcast channel," in *Proceedings of Tyrrhenian International Workshop on Digital Communication*, Ischia Island, Italy, September 2007.

[22] E. Conte, S. Tomasin, and N. Benvenuto, "Scheduling strategies for multiuser MIMO OFDM systems with limited feedback," in *Proceedings of the 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, Cannes, France, September 2008.

[23] L. Hanzo, M. Münster, B. Chei, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANS and Broadcasting*, John Wiley & Sons, New York, NY, USA, 2003.

[24] N. Jindal, "Finite rate feedback MIMO broadcast channels," in *Proceedings of Workshop on Information Theory and Its Applications (ITA '06)*, San Diego, Calif, USA, February 2006.

[25] P. Svedman, S. K. Wilson, L. J. Cimini Jr., and B. Ottersten, "Opportunistic beamforming and scheduling for OFDMA systems," *IEEE Transactions on Communications*, vol. 55, no. 5, pp. 941–952, 2007.

[26] M. Kountouris and D. Gesbert, "Memory-based opportunistic multi-user beamforming," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '05)*, pp. 1426–1430, Adelaide, SC, USA, September 2005.

[27] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proceedings of the 51st Vehicular Technology Conference (VTC '00)*, vol. 3, pp. 1854–1858, Tokyo, Japan, May 2000.

[28] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?" in *Proceedings of IEEE International Conference on Communications*, vol. 1, pp. 234–238, Paris, France, June 2004.

[29] "SCM micro cell and urban canyon model," Motorola, January 2003, http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/3GPP_3GPP2_SCM/ConfCall-10-20030130/.

[30] J. Salo, G. D. Galdo, J. Salmi, et al., "Matlab implementation of the 3GPP spatial channel model," Tech. Rep. 3GPP TR 25.996, January 2005, http://www.tkk.fi/Units/Radio/scm/.

[31] N. Benvenuto, E. Conte, S. Tomasin, and M. Trivellato, "Predictive channel quantization and beamformer design for MIMO-BC with limited feedback," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 3607–3611, Washington, DC, USA, November 2007.