IEEE Transactions on Vehicular Technology, Jan 2010, Volume 59, pp 406-419

On the Performance, Complexity and Fairness of Suboptimal Resource Allocation for Multi-User MIMO-OFDMA Systems

Tarcisio F. Maciel *Member, IEEE*, Anja Klein, *Member, IEEE*,

***Abstract*—The combination of Multiple Input Multiple Output (MIMO) and Orthogonal Frequency Division Multiple Access (OFDMA) is a promising solution to the flexible and spectrally efficient provision of data services in future wireless communication systems. However, adaptive Resource Allocation (RA) in frequency, time, and space in Multi-User MIMO-OFDMA systems is very complex due to the inclusion of the space dimension and to the large number of resources to be managed. Indeed, an optimal RA to maximize the sum rate is usually too complex for practical application and suboptimal strategies are required. In this work, the performance, complexity, and fairness of suboptimal RA strategies aiming at the maximization of the sum rate are investigated. A model for suboptimal RA strategies is proposed and two new RA strategies are introduced. The proposed strategies are compared in terms of sum rate, complexity, and throughput fairness and are shown to present better performance-complexity and performance-fairness trade-offs than some existing suboptimal strategies, as well as to achieve almost the same sum rate obtained through an Exhaustive Search.**

***Index Terms*—Sum rate maximization, throughput fairness, resource allocation**

I. INTRODUCTION

FUTURE wireless communication systems are expected to provide data services with rate requirements ranging from a few kbps up to some Mbps and, due to the high frequency spectrum costs, these systems must also be highly spectrally efficient. Transmission schemes based on Multiple Input Multiple Output (MIMO) and Orthogonal Frequency Division Multiple Access (OFDMA) are considered as promising solutions to meet these requirements [1]–[3]. MIMO-OFDMA systems are flexible and spectrally efficient due to the large number of narrowband frequency channels that may be adaptively allocated and to the ability of reusing channels in space [4], [5]. However, adaptive Resource Allocation (RA) in frequency, time, and space is complex in such systems due to the large number of degrees of freedom to be handled [6].

Resources can be thought of as elements of a 3-dimensional structure with subcarriers, Time-Slots (TSs), and spatial layers corresponding to frequency, time, and space resources, respectively [3]. In the Downlink (DL), the Base Station (BS) decides on the basis of Channel State Information (CSI) which resources to allocate to which Mobile Stations (MSs) and sends data to the selected MSs on the allocated resources. Since each resource can be allocated to a different BS-MS link, a huge number of possible allocations exist even for relatively small numbers of resources and MSs.

Through adequate frequency and time synchronization, frequency and time resources can be made orthogonal by design. Thus, the RA is simplified because signals sent to MSs on orthogonal resources do not interference with each other. However, space resources result from the spatial reuse of a same frequency-time resource and signals transmitted by the BS to a group of MSs on the resource essentially interfere with each other. Thus, frequency-time resources are the real system resources, which are shared through Space Division Multiple Access (SDMA) by a group of MSs, namely an SDMA group.

Different aspects affect the system performance. Firstly, if the channels of the MSs in an SDMA group are highly spatially uncorrelated, signals sent to these MSs can be efficiently separated in space at the BS through precoding. These MSs are said to be spatially compatible. SDMA groups must contain spatially compatible MSs in order to obtain SDMA gains and improve capacity. Otherwise, the signals sent to the MSs may strongly interfere with each other and compromise the system performance [7]–[11]. Consequently, the SDMA group composition affects the system performance.

Secondly, there are different precoding techniques which suppress spatial interference totally, in part, or ignore it [12]–[16]. Thus, the selection of the precoding technique also affects the system performance.

Thirdly, DL spatial interference is a function of the power distribution among the signals sent by the BS to the MSs. For a given amount of power available for an SDMA group, allocating more power to the signal sent to a certain MS enhances its the receive signal quality, e.g., in terms of Signal-to-Interference plus Noise Ratio (SINR), but reduces the SINR perceived by the other MSs in the group. Analogously, allocating more power to a certain resource enhances the SINR perceived by the MSs sharing this resource, but reduces the SINR perceived by MSs to which other resources have been allocated [17]. An efficient power distribution among MSs and resources must be performed and it also affects the system performance.

Finally, because spatial compatibility is resource-dependent, the selection of the resources assigned to the SDMA groups

T. F. Maciel was with the Communications Engineering Lab, Technische Universität Darmstadt, Darmstadt, Germany. He is now with the Federal University of Ceará, Fortaleza, Brazil (e-mail: maciel@ufc.br).

A. Klein is with the Communications Engineering Lab, Technische Universität Darmstadt, Darmstadt, Germany (e-mail: a.klein@nt.tu-darmstadt.de).

also affects the system performance [6], [10], [11].

Each of the four aspects above relates to a subproblem of the RA in Multi-User (MU) MIMO-OFDMA systems, namely:

1) The SDMA grouping problem, which corresponds to building groups of spatially compatible MSs on each resource.
2) The precoding problem, which corresponds to determining precoding vectors and separating in space the signals sent to the MSs.
3) The power allocation problem, which corresponds to allocating power to MSs and resources.
4) The resource assignment problem, which corresponds to assigning resources to the best SDMA groups, e.g., those leading to the highest sum rate.

Since only whole resources can be assigned to whole MSs, the SDMA grouping and the resource assignment problems are integer problems. Integer problems are usually hard to be solved optimally due to their combinatorial nature. Because the domain of integer optimization problems is described by discrete points and consequently not convex, these problems might not admit a unique optimal solution and can not be solved using convex optimization methods [18], [19]. Integer optimization problems are often solved considering intelligent enumerations and relaxations, such as those commonly employed in branch-and-bound methods to solve integer linear problems [18]. Integer problems might even be Non-deterministic Polynomial time Complete (NP-C), such as the SDMA grouping problem [20], [21] and some resource assignment problems [22], and require an Exhaustive Search (ES) in order to be optimally solved.

Precoding and power allocation problems are not integer problems, as long as precoding vectors and allocated powers are not discretized.

Because frequency spectrum is a scarce and expensive resource, RA strategies that aim at maximizing the sum rate of the system are an important research topic. However, an optimal RA strategy to maximize the sum rate must jointly solve the above four subproblems and in most of the cases leads to a complex combinatorial and non-convex optimization problem [6], [23], [24]. Due to coupling between the four subproblems, even when a formulation as a convex optimization problem is possible, high-complexity algorithms are required [12], [14]. Usually, these algorithms need a considerable number of iterations to converge to a suitable solution and involve complex operations, such as matrix inversions or decompositions, inside each iteration. The complexity of such optimum solutions rapidly increases with the number of MSs and resources and is not affordable for many practical cases. Therefore, suboptimal RA strategies with low complexity and able to achieve high sum rates are desired. Moreover, because such strategies lead to potentially unfair throughput distributions among the MSs, it is also desirable that a good degree of throughput fairness be achieved by the RA strategies without substantially compromising the sum rate. Suboptimal RA strategies with these characteristics are proposed and investigated in this work. The proposed suboptimal RA strategies follow a new framework that combines solutions to the four aforementioned subproblems to define new suboptimal RA strategies providing

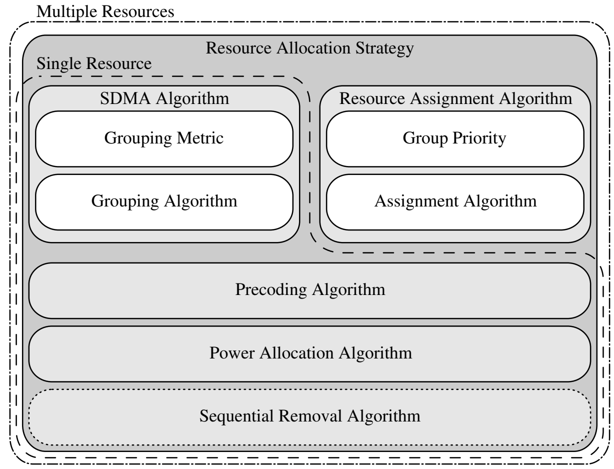


Figure 1. Framework for suboptimal RA strategies for MIMO-OFDMA systems.

good performance-complexity and performance-fairness trade-offs. The remainder of this work is organized as follows.

In Section II, a framework for suboptimal RA strategies is proposed, which is used to classify some existing RA strategies representing the state of the art and to define the new RA strategies studied in this work. In Section III, the system model considered in this work is presented. In this section, the problem of maximizing the sum rate is presented and a new mixed-integer formulation for the problem is proposed, which characterizes the four above subproblems. Sections IV to VII introduce the algorithms applied to each subproblem. These algorithms are combined in Section VIII to define new suboptimal RA strategies whose performance, complexity, and fairness are analyzed in Section IX. Finally, Section X presents some conclusions.

II. Framework for Suboptimal RA Strategies

In this section, a framework is proposed to model suboptimal RA strategies. It divides the RA problem into the four subproblems introduced in Section I and is illustrated in Fig. 1. For each subproblem, existing or new algorithms oriented towards the maximization of the sum rate are employed and their combination defines an RA strategy.

In Fig. 1, two cases are defined regarding the number of resources considered by the RA strategy: a single-resource case, indicated by the dashed line, and a multiple-resource case, indicated by the dot-dashed line. In the single-resource case, resources are allocated one-by-one by the RA strategies and, consequently, the resource assignment problem loses relevance. In the multiple-resource case, RA strategies take all the multiple resources into account and the resource assignment problem must be considered.

In the following, the different blocks in Fig. 1 are described. An SDMA algorithm is used to solve the SDMA grouping problem. This problem is NP-C [20], [21] and would need an ES over all the possible groups to find the one maximizing the sum rate. Therefore, suboptimal SDMA algorithms with low-complexity are preferred. They are usually composed by two main elements, cf. Fig. 1:

- A grouping metric, which measures the spatial compatibility among MSs in a group.
- A grouping algorithm, which employs the grouping metric to build and compare groups while avoiding an ES.

A precoding algorithm and a power allocation algorithm are employed to solve the precoding problem and the power allocation problem, respectively, as illustrated in Fig. 1.

To determine which resource to assign to which SDMA group in the multiple-resource case, a resource assignment algorithm is used, which involves two elements, cf. Fig. 1:

- A group priority, which measures the efficiency of assigning a given resource to an SDMA group.
- An assignment algorithm, which employs the group priority to assign the resources to the SDMA groups.

The algorithms applied to each subproblem appear isolated in Fig. 1, but the exchange of information among them is allowed. Separating the SDMA and resource assignment algorithms from the precoding and power allocation algorithms simplifies the RA. Indeed, precoding and power allocation problems are easy to solve if the SDMA grouping and resource assignment problems, i.e., the combinatorial part of the RA problem, are solved beforehand. However, if the SDMA algorithm is aware of the actual precoding and power allocation, it may estimate better the performance of a group, e.g., in terms of group capacity, and avoid putting MSs into the group that do not contribute to enhance the sum rate [8]. Otherwise, resources might be allocated to MSs that do not contribute to improve the sum rate and, consequently, some SDMA groups might contain more MSs than they should do. In this case, choosing the size of the SDMA group becomes also a problem and group sizes may need to be adjusted by removing MSs. In order to adjust the size of SDMA groups and enhance the sum rate, a Sequential Removal (SR) algorithm is employed to remove MSs from the group [8]–[11], [21], [25]. Because a more reliable decision about which MSs to remove can be made considering the actual precoding and power allocation, the SR algorithm employs this information. The SR algorithm shown in Fig. 1 is only needed if the SDMA algorithm is unaware of precoding and power allocation.

In the following, some state-of-the-art strategies fitting into the framework of Fig. 1 are shortly discussed. They are listed in Table I, whose columns correspond to the algorithms (and their elements) previously discussed in this section.

In Table I, the RA strategies are grouped in terms of grouping metric and for each metric they are roughly ordered in terms of complexity. The RA strategies in [23], [24] aim at maximizing the sum rate and disregard fairness aspects. In [23], [24], the joint solution of the SDMA grouping, resource assignment, and power allocation problems based on convex optimization is considered and high sum rates are achieved, but with very high complexity.

More simple strategies are obtained considering a single resource [8], [9], [20], [21], [26], [27], [29]. Complexity reductions are also achieved by using less complex grouping metrics. For example, capacity-based metrics [26], [27] and metrics based on null-space Successive Projections (SPs) [15], [29] involve relatively more complex matrix operations than metrics based on the spatial correlation, which in spite of being more simple are able to capture the spatial compatibility among the MSs efficiently [7]–[11], [20], [21], [30]. Simple greedy grouping algorithms, such as the Best Fit (BF) algorithm of [20], also have considerably lower complexity than grouping algorithms based on convex optimization, as in [23]–[25], [31] and the Compatibility Optimization (CO) algorithm of [7]. The BF algorithm of [20] is employed in this work and is described in Section IV-B.

Most of the mentioned works concentrate on the maximization of the sum rate, give less attention to Quality of Service (QoS) aspects, and present quite variable complexity. In [20], [25], [31], minimum target SINR for each MS is also considered in order to ensure QoS. In [25], [31], complex convex optimization problems are formulated for SDMA grouping, precoding, and power allocation which are solved using SDP. Because predicting the feasibility of this problem is as hard as solving the problem itself, SR algorithms are employed in [25], [31] to adequately remove MSs until a feasible solution is obtained. While achieving high sum rates, the strategies in [25], [31] present very high complexity. In [20], [25], more simple algorithms have also been proposed, which admit MSs to an SDMA group only if the SINR of all MSs in the group becomes not lower than a given target SINR. Nevertheless, the RA strategies in [20], [25] remain more complex than other strategies employing more simple grouping metrics.

It can also be noted that for strategies considering multiple resources, grouping metrics and group priorities usually match each other, as well as grouping and assignment algorithms [7], [15], [23]–[25], [28], [31]. In these cases, SDMA grouping and resource assignment problems are solved simultaneously by the RA strategies with SDMA groups being built in parallel on the different resources. However, in some cases the obtained solution is equivalent to allocating resources one-by-one [15], [29].

Adaptive RA for MU MIMO-OFDMA systems has been an active research field in the last few years, cf. Table I, and a considerable number of investigations has already been conducted. Complexity plays a role in many previous works. However, often a detailed analysis of the performance-complexity trade-off of suboptimal RA strategies has not been considered. Most of the strategies also concentrated either on maximizing the sum rate or providing QoS. However, strategies able to provide a high degree of throughput fairness among the MSs at the expense of only small reductions of the sum rate have not been provided. Therefore, it is an objective in this work to propose suboptimal RA strategies having attractive performance-complexity and performance-fairness trade-offs, i.e., they should be able to provide a good degree of throughput fairness among the MSs despite the fact of being mainly oriented to sum rate maximization.

For these strategies, it is relevant to investigate whether they can achieve sum rates close to the sum rate obtained through an ES with considerably lower complexity. In this context, it is important to investigate whether SDMA algorithms using low-complexity grouping metrics not depending on precoding and power allocation can perform as good as other more complex algorithms employing metrics that depend on precoding and

Table I
RA STRATEGIES.

SDMA algorithm		Precoding algorithm*	Power allocation algorithm*	SR criterion*	Resource assignment algorithm		Ref.
Grouping metric	Grouping algorithm				Group priority	Assignment algorithm	
Group capacity	Convex optimization	ZF	WF	-	Group capacity	Convex optimization	[23]
	Convex optimization	ZF	WF	-	Group capacity	Convex optimization	[24]
	Best Fit	ZF	WF, EPA	-	Single resource		[26]
	Best Fit	ZF	WF	-	Single resource		[27]
Channel gains with SPs	Best Fit	ZF	WF	Fixed group size	Channel gains with SPs	Best Fit	[15], [28]
	Best Fit	ZF	WF	Fixed group size	Single resource		[29]
Total spatial correlation	Compatibility Optimization	-	-	Fixed group size	Total spatial correlation	Compatibility Optimization	[7]
	Best Fit, First Fit	GEP	EPA	-	Single resource		[20]
	Best Fit	ZF	WF, EPA	Channel gain	Single resource		[21]
	MS partitioning	ZF	Adap. bit loading	Fixed group size	Single resource		[30]
Minimum target SINR	Convex optimization	SDP		Max. SINR gap	Single resource		[31]
	Best Fit, First Fit, admit all	SDP		E.g., Random	Single resource		[25]
	Best Fit, First Fit	GEP	EPA	-	Single resource		[20]

*Acronyms: Zero-Forcing (ZF), Generalized Eigen-Precoding (GEP), Semidefinite Programming (SDP), Water Filling (WF), Equal Power Allocation (EPA), Sequential Removal (SR)

power allocation. RA strategies either optimize SDMA group sizes, e.g. in [23], [24], or adjust them using an SR algorithm, e.g. in [25], [31], or just fix their values, e.g., in [29], [30]. In particular, the impact of the the SDMA group size selection by an SR algorithm on the performance of the strategies deserves additional investigation. Additionally, the sensitiveness of RA strategies to imperfect CSI must also be considered in order to determine whether the high-complexity strategies offer some advantage compared to the low-complexity ones in this case. These aspects will be addressed in the following sections.

III. SYSTEM MODEL

In this section, the system model used in this work is presented. The DL of a single BS located at the corner of a hexagonal sector is considered. The BS has an M-element Antenna Array (AA) and serves a number K of single-antenna MSs. It is assumed that the BS's transmit power P can be arbitrarily distributed among the MSs and that CSI about the DL channels to the MSs is available at the BS.

Gaussian signaling is considered and the data symbols transmitted by the BS to the MSs are assumed to be uncorrelated with unit average power. Inter-cell interference is assumed to be Gaussian-distributed and is incorporated in the Additive White Gaussian Noise (AWGN) perceived in the system. Only fast fading is considered, which is a common assumption. Indeed, most of the works referred to in Section II consider only fast fading. Additionally, low MS mobility is assumed because it is well-known that adaptive RA fits well for low mobility scenarios [4].

A frequency block composed of Q_{sub} adjacent subcarriers is considered the minimum allocable resource unit in frequency. Frequency blocks are expected to have an almost flat channel transfer function.

Frames composed of T TSs are considered. A TS is the minimum allocable resource unit in time and transports several OFDMA symbols. The channel transfer function over a whole frame is not expected to vary considerably, which holds for low MS mobility and short frame durations [4].

A resource is defined as a frequency-time resource unit described by one frequency block and one TS. These resources are also called Physical Resource Blocks (PRBs) [1], slots [2], or chunks [3].

On each resource, SDMA is used to multiplex up to M data streams separated in space through linear precoding [13], i.e., on each of the M spatial layers of a resource the BS can transmit a data stream to a different MS [3].

Denoting by σ^2 the average AWGN power per subcarrier, the average Signal-to-Noise Ratio (SNR) γ in the system is defined as

$$\gamma = \frac{P}{\sigma^2}. \tag{1}$$

In this work, the geometric-based stochastic MIMO channel model of the Wireless World Initiative New Radio (WINNER) project is employed. The WINNER Phase I Channel Model (WIM) captures space and time characteristics of the channel for realistic scenarios and their parameters have been determined from measurement campaigns [32], [33].

In this work, channel modeling in the frequency domain is adopted. The channel transfer function of all the Q_{sub} subcarriers of a frequency block $b, b = 1, \ldots, B$, can be efficiently represented by that of the middle subcarrier of the frequency block. The channel coefficient $h_{k,b,m}$ denotes the sampled frequency response of the channel between the m^{th} antenna of the BS and the k^{th} MS on the middle subcarrier of the frequency block b. The channel coefficients $h_{k,b,m}$ are obtained using the WIM [32], [33] and are organized in a vector

$$\mathbf{h}_{k,b} = \begin{bmatrix} h_{k,b,1} & h_{k,b,2} & \ldots & h_{k,b,M} \end{bmatrix} \tag{2}$$

for the channel between the BS and an MS k on the frequency block b. Denoting vector/matrix transposition by $(\cdot)^{\text{T}}$ and using (2), the channel matrix $\mathbf{H}_b$ of all MSs on frequency block b is obtained by stacking the channel vectors $\mathbf{h}_{k,b}$ as

$$\mathbf{H}_b = \begin{bmatrix} \mathbf{h}_{1,b}^{\text{T}} & \mathbf{h}_{2,b}^{\text{T}} & \ldots & \mathbf{h}_{K,b}^{\text{T}} \end{bmatrix}^{\text{T}}. \tag{3}$$

An estimated channel matrix $\hat{\mathbf{H}}_b$ of $\mathbf{H}_b$ is used to describe the Channel State Information at the Transmitter (CSIT) available at the BS on a frame basis. Perfect CSI is assumed at the MSs.

A model for imperfect CSIT is also considered to investigate the performance of the RA strategies. Imperfections in the CSIT might originate, e.g., from the AWGN in the system, suboptimal channel estimation, inherent processing or feedback delays, among others. For MS k and frequency block b, both estimation errors and imperfections due to processing or feedback delays can be modeled by an additive Zero Mean Circularly Symmetric Complex Gaussian (ZMCSCG) error term $\mathbf{e}_{k,b} \in \mathbb{C}^{1\times M}$ [34]–[38]. Let $0 \leq \nu \leq 1$ be a parameter controlling the amounts of the true channel $\mathbf{h}_{k,b}$ and error term $\mathbf{e}_{k,b}$ in the estimated channel $\hat{\mathbf{h}}_{k,b}$, which can be expressed as

$$\hat{\mathbf{h}}_{k,b} = \sqrt{1-\nu}\mathbf{h}_{k,b} + \sqrt{\nu}\mathbf{e}_{k,b}. \tag{4}$$

In order to obtain a normalization of $\hat{\mathbf{h}}_{k,b}$, the variance $\sigma^2_{\mathbf{e}}$ of the entries of the error term $\mathbf{e}_{k,b}$ can be modeled to be equal to the variance $\sigma^2_{\mathbf{h}}$ of the entries of $\mathbf{h}_{k,b}$. Note that by dividing (4) by $\sqrt{1-\nu}$ one obtains a standard "nominal plus perturbation" model for imperfect CSIT, as in [35]. Also note that the model in (4) matches the model for delayed CSIT [29], [36]–[38]. According to [34], the model in (4) allows to draw only a lower bound on the training-based capacity for Minimum Mean Square Error (MMSE) estimation.

Using (4), the different amounts of imperfection in the CSIT can be obtained by varying ν. Denoting by $|\cdot|$ the absolute value of a complex number and using (2), the model in (4) allows to describe the quality of the CSIT as

$$\gamma_{\text{CSI}} = \frac{\mathcal{E}\{(1-\nu)\,|h_{k,b,m}|^2\}}{\mathcal{E}\{\nu\,|e_{k,b,m}|^2\}} = \frac{1-\nu}{\nu}, \tag{5}$$

which expresses the relationship between the expected magnitudes of the terms due to $\mathbf{h}_{k,b}$ and $\mathbf{e}_{k,b}$ present in $\hat{\mathbf{h}}_{k,b}$.

In the following, the sum rate maximization problem is formulated. Because CSI is available at the BS on a frame basis, the problem can be formulated for the resources described by the frequency blocks $b, b = 1,\ldots,B$, and the first TS of each frame. A solution obtained for these resources applies to the resources associated with the remaining TSs. Further on, the frequency-time resource units present during each TS will be indexed by $b, b = 1,\ldots,B$. Let $p_{k,b}$ and $\mathbf{w}_{k,b}$ denote the allocated power and the precoding vector of MS k on resource b, respectively, and let $\|\cdot\|_2$ denote the 2-norm of a vector. The DL SINR $\gamma_{k,b}$ is given by

$$\gamma_{k,b} = \frac{p_{k,b}|\hat{\mathbf{h}}_{k,b}\mathbf{w}_{k,b}|^2}{\sigma^2 + \sum\limits_{j=1,j\neq k}^{K} p_{j,b}|\hat{\mathbf{h}}_{k,b}\mathbf{w}_{j,b}|^2}, \tag{6}$$

and the rate of MS k on resource b becomes

$$R_{k,b} = \log_2(1+\gamma_{k,b}). \tag{7}$$

Using (7), the maximization of the sum rate can be formulated as

$$\{p^\star_{k,b}, \mathbf{w}^\star_{k,b}\} = \underset{\{p_{k,b},\mathbf{w}_{k,b}\}}{\arg\max} \left\{\sum_{b=1}^{B}\sum_{k=1}^{K} R_{k,b}\right\} \tag{8a}$$

subject to

$$p_{k,b} \geq 0, \forall k, b, \tag{8b}$$
$$\sum_{b=1}^{B}\sum_{k=1}^{K} p_{k,b} = P, \tag{8c}$$
$$\|\mathbf{w}_{k,b}\|_2 = 1, \forall k, b, \tag{8d}$$

where constraint (8b) ensures non-negative powers, constraint (8c) limits the total transmit power, and constraint (8d) implies unit-norm precoding vectors.

Note that the SDMA grouping and resource assignment problems are implicit in problem (8). In the following, a new formulation of problem (8) as a mixed-integer optimization problem is introduced and this new formulation explicitly characterizes the four subproblems of Section I. Let the binary variable $u_{k,b}$ indicate whether resource b is assigned to MS k. Of course, if $p_{k,b} > 0$, then $u_{k,b} = 1$, otherwise power would be wasted. Then, $u_{k,b}$ is defined as

$$u_{k,b} = \begin{cases} 1, & \text{for } p_{k,b} > 0, \\ 0, & \text{for } p_{k,b} = 0. \end{cases} \tag{9}$$

There is a maximum number $L = \sum_{l=1}^{M}\binom{K}{l}$ of SDMA groups that can be defined using $u_{k,b}$ and each resource is shared in space by one of the groups. Let $l, l = 1,\ldots,L$, indicate the SDMA groups and the binary variable $v_{l,b}$ indicate whether resource b is assigned to the SDMA group $\mathcal{G}_{l,b}$. Note that, if resource b is assigned to the SDMA group $\mathcal{G}'_{l,b}$ and $\exists k \notin \mathcal{G}'_{l,b}$ for which $p_{k,b} > 0$, then there is a group $\mathcal{G}_{l,b} \neq \mathcal{G}'_{l,b}$ for which $p_{k,b} > 0 \Leftrightarrow k \in \mathcal{G}_{l,b}$. Thus, the resource b can be seen as effectively assigned to SDMA group $\mathcal{G}_{l,b}$ and $v_{l,b}$ can be defined as

$$v_{l,b} = \begin{cases} 1, & \text{if } p_{k,b} > 0 \Leftrightarrow k \in \mathcal{G}_{l,b}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Using $u_{k,b}$ and $v_{l,b}$, problem (8) can be reformulated as

$$\left\{\begin{array}{l} p^\star_{k,b}, \mathbf{w}^\star_{k,b}, \\ u^\star_{k,b}, v^\star_{l,b} \end{array}\right\} = \underset{\left\{\begin{array}{l} p_{k,b},\mathbf{w}_{k,b}, \\ u_{k,b},v_{l,b} \end{array}\right\}}{\arg\max} \left\{\sum_{b=1}^{B}\sum_{l=1}^{L} v_{l,b}\sum_{k=1}^{K} u_{k,b}R_{k,b}\right\} \tag{11a}$$

subject to

$$p_{k,b} \geq 0, \forall k, b, \tag{11b}$$
$$\sum_{b=1}^{B}\sum_{k=1}^{K} p_{k,b} = P, \tag{11c}$$
$$\|\mathbf{w}_{k,b}\|_2 = 1, \forall k, b, \tag{11d}$$
$$u_{k,b} \in \{0,1\}, \forall k, b, \tag{11e}$$
$$\sum_{l=1}^{L} v_{l,b} \leq 1, \forall b, \tag{11f}$$
$$v_{l,b} \in \{0,1\}, \forall l, b. \tag{11g}$$

The new formulation in (11) clearly characterizes problem (8) as a mixed-integer optimization problem. Another mixed-integer formulation of problem (8) can be found in [24]. In the new formulation of (11), $u_{k,b}$ and $v_{l,b}$ together with $\mathbf{w}_{k,b}$ and $p_{k,b}$ allow to explicitly characterize the four subproblems in (8) as follows.

- The variables $u_{k,b}$ are related to the SDMA grouping problem, in which up to M of the K MSs must be selected on each resource b, so that $u_{k,b} = 1, \forall k \in \mathcal{G}_{l,b}$ and $u_{k,b} = 0, \forall k \notin \mathcal{G}_{l,b}$.
- The variables $\mathbf{w}_{k,b}$ are related to the precoding problem, as in problem (8).

- Similarly, the powers $p_{k,b}$ are related to the power allocation problem.
- The variables $v_{l,b}$ are related to the resource assignment problem, in which each resource b is assigned to no more than one SDMA group according to constraint (11f).

Problem (11) provides some insight into the elements of problem (8). Firstly, it can be noted that the power distribution is the element that keeps the four subproblems interdependent. If the power is distributed a priori among resources, e.g., using EPA, the resource assignment can be performed on a resource-by-resource basis. Additionally, the precoding vector $\mathbf{w}_{k,b}$ of MS k on resource b plays no role if the power $p_{k,b}$ is zero. Moreover, the binary variables $u_{k,b}$ and $v_{l,b}$ depend only on $p_{k,b}$, as shown in (9) and (10), respectively.

Secondly, the SDMA grouping problem is responsible for yielding (11) combinatorial, since the combinatorial increase in the number L of candidate SDMA groups is due to $u_{k,b}$, which affects $v_{l,b}$ subsequently. If SDMA groups are already defined on each resource, the problem is no longer NP-C.

Thirdly, dividing problem (11) into subproblems allows to adapt algorithms to each subproblem individually and to combine them into suboptimal but efficient RA strategies. This formulation leads to high flexibility and strategies providing interesting trade-offs between the sum rate maximization and the throughput fairness can be obtained.

IV. SDMA Algorithms

A. Grouping Metrics

In this section, the grouping metrics used by the SDMA algorithms are described. Grouping metrics are functions of the CSIT that map the spatial properties of the MSs' channels to a scalar value quantifying how efficiently the MSs can be separated in space. In this work, the following ones are considered:

- The group capacity $f_{\text{CAP}}(\mathcal{G})$.
- The sum of channel gains with null-space SPs $f_{\text{SP}}(\mathcal{G})$.
- The convex combination of the total spatial correlation and channel gains $f_{\text{CC}}(\mathcal{G})$.

For simplicity of notation, the index b is omitted in the sequel and the above metrics are described considering a single resource.

In the following, some additional definitions are made and the group capacity [8], [26], [27] is described. Let $\mathcal{G}$ denote an SDMA group containing a number G of MSs. The channel matrix $\hat{\mathbf{G}} \in \mathbb{C}^{G\times M}$ for the SDMA group $\mathcal{G}$ is obtained from $\hat{\mathbf{H}}$ in (3) by taking the rows corresponding to the channels of the MSs belonging to $\mathcal{G}$. For example, if MSs 1, 2, and K belong to $\mathcal{G}$, $\hat{\mathbf{G}}$ contains the 1$^{\text{st}}$, 2$^{\text{nd}}$, and K^{th} rows of $\hat{\mathbf{H}}$. The channel of the i^{th} MS in $\mathcal{G}$, with $i = 1, \dots, G$, is given by the i^{th} row $\hat{\mathbf{g}}_i$ of $\hat{\mathbf{G}}$. Let p_i and $\mathbf{w}_i$ denote the allocated power and the precoding vector of MS i in $\mathcal{G}$, respectively. Then, using (7) the capacity of the SDMA group $\mathcal{G}$ is written as

$$f_{\text{CAP}}(\mathcal{G}) = \sum_{i=1}^{G} R_i. \tag{12}$$

The higher $f_{\text{CAP}}(\mathcal{G})$ is, the more spatially compatible the MSs in $\mathcal{G}$ are. Since $f_{\text{CAP}}(\mathcal{G})$ reflects the effective capacity of the group considering precoding and power allocation, it is a reliable metric [8], [26], [27]. However, because precoding and power allocation must be computed for all MSs in $\mathcal{G}$ whenever the group composition changes, the complexity of SDMA algorithms using $f_{\text{CAP}}(\mathcal{G})$ might become high if a large number of groups is considered.

The sum of channel gains with null-space SPs [10], [11], [15], [27], [29] is described in the sequel. Let $\mathbf{I}_M$ denote an $M \times M$ identity matrix and assume an admission order for the MSs in $\mathcal{G}$. The channel $\hat{\mathbf{g}}_i$ of MS i is projected onto the null-space of the channels $\hat{\mathbf{g}}_{i'}$ of all MSs $i', i' = 1, 2, \dots, i-1$ previously admitted to $\mathcal{G}$ using a projection matrix $\mathbf{T}_i$ given by

$$\mathbf{T}_i = \begin{cases} \mathbf{I}_M, & \text{if } i = 1, \\ \mathbf{T}_{i-1} - \frac{\mathbf{T}_{i-1}^{\text{H}} \hat{\mathbf{g}}_{i-1}^{\text{H}} \hat{\mathbf{g}}_{i-1} \mathbf{T}_{i-1}}{\|\hat{\mathbf{g}}_{i-1}\mathbf{T}_{i-1}\|_2^2}, & \text{if } 2 \le i \le G. \end{cases} \tag{13}$$

Using (13), the sum of channel gains with null-space SPs is written as

$$f_{\text{SP}}(\mathcal{G}) = \sum_{i=1}^{G} \|\hat{\mathbf{g}}_i \mathbf{T}_i\|_2^2 . \tag{14}$$

The higher the channel gain $\|\hat{\mathbf{g}}_i\|_2^2$ of MS i is and the more spatially uncorrelated with respect to the MSs $i' \in \mathcal{G}$ the MS i is, the higher $\|\hat{\mathbf{g}}_i \mathbf{T}_i\|_2^2$ might become and the more spatially compatible MS i and the MSs i' are considered to be. Consequently, $f_{\text{SP}}(\mathcal{G})$ favors SDMA groups whose MSs have high channel gain and are highly spatially uncorrelated.

$f_{\text{SP}}(\mathcal{G})$ in (14) depends neither on precoding nor on power allocation and, consequently, has lower complexity than $f_{\text{CAP}}(\mathcal{G})$ [10], [11], [39]. By taking care of the admission order in the SDMA algorithm, $f_{\text{SP}}(\mathcal{G})$ can be efficiently used as grouping metric [10], [11], [15], [27], [29].

The convex combination of the total spatial correlation and channel gains has been proposed by the authors in [9] and is discussed in the sequel. Given the channel vectors $\hat{\mathbf{h}}_j$ and $\hat{\mathbf{h}}_k$ of MSs j and k, respectively, the spatial correlation among them is given by the maximum normalized scalar product

$$\rho_{j,k} = \frac{|\hat{\mathbf{h}}_j \hat{\mathbf{h}}_k^{\text{H}}|}{\|\hat{\mathbf{h}}_j\|_2 \|\hat{\mathbf{h}}_k\|_2}. \tag{15}$$

The lower $\rho_{j,k}$ is, the less spatially correlated MSs j and k are. Because $\rho_{j,k}$ is a pairwise metric, the sum of $\rho_{j,k}$ for every pair of MSs in $\mathcal{G}$ must be used to measure the total spatial correlation among the MSs in the group. Moreover, MSs with high channel gain should be preferred since it is well-known that allocating power to an MS with high channel gain is more efficient than giving the same power to an MS with low channel gain. However, effective channel gains of MSs in a group depend on the spatial compatibility among them. Therefore, an efficient SDMA group must provide an adequate trade-off between total spatial correlation and channel gains.

Let the attenuation vector $\mathbf{a}$ be defined as

$$\mathbf{a} = \left[\|\hat{\mathbf{h}}_1\|_2^{-2} \quad \|\hat{\mathbf{h}}_2\|_2^{-2} \quad \dots \quad \|\hat{\mathbf{h}}_K\|_2^{-2}\right]^{\text{T}}, \tag{16}$$

which contains the inverse of the channel gains, and let $\mathcal{D}\{\cdot\}$ denote a diagonal matrix whose diagonal is given in the vector argument. Then, using (15) and (16), the spatial correlation matrix $\mathbf{C}$ can be defined as

$$\mathbf{C} = \left| \sqrt{\mathcal{D}\{\mathbf{a}\}}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}}\sqrt{\mathcal{D}\{\mathbf{a}\}} \right|, \tag{17}$$

where $|\cdot|$ is applied element-wise. $\mathbf{C}$ contains $\rho_{j,k}$ for each pair of MSs in the system.

Using (9), let the binary vector $\mathbf{u}$ be defined as

$$\mathbf{u} = [u_1 \quad u_2 \quad \dots \quad u_K]^{\mathrm{T}}, \tag{18}$$

and let $\|\cdot\|_{\mathrm{F}}$ denote the Frobenius norm of a matrix/vector. Then, using (16), (17), and (18), the convex combination of the total spatial correlation and channel gains is defined as

$$f_{\mathrm{CC}}(\mathcal{G}) = \frac{(1-\beta)}{\|\mathbf{C}\|_{\mathrm{F}}}\mathbf{u}^{\mathrm{T}}\mathbf{C}\mathbf{u} + \frac{\beta}{\|\mathbf{a}\|_{\mathrm{F}}}\mathbf{a}^{\mathrm{T}}\mathbf{u}, \tag{19}$$

where $0 \leq \beta \leq 1$ is a parameter controlling the trade-off between spatial correlation and channel gain [9]–[11].

The lower the value that $f_{\mathrm{CC}}(\mathcal{G})$ assumes, the more spatially compatible the MSs in $\mathcal{G}$ are considered to be. $f_{\mathrm{CC}}(\mathcal{G})$ depends neither on precoding and power allocation nor on complex matrix operations and, consequently, is less complex than $f_{\mathrm{CAP}}(\mathcal{G})$ and $f_{\mathrm{SP}}(\mathcal{G})$. The grouping metric $f_{\mathrm{CC}}(\mathcal{G})$ will be used later as part of new RA strategies.

B. Grouping Algorithm

In this section, the grouping algorithms employed by the SDMA algorithms are described. The task of the grouping algorithm is to build an efficient SDMA group on a given resource with acceptable performance compared to an ES. The following grouping algorithms are considered here:

- The Exhaustive Search (ES) algorithm, which performs an ES for the SDMA group that maximizes the grouping metric.
- The Random Grouping (RG) algorithm, which just randomly builds an SDMA group of specific size.
- The Convex Grouping (CG) algorithm, which is a new grouping algorithm proposed by the authors in [9] and formulated as a quadratic optimization problem.
- The Best Fit (BF) algorithm, which is a greedy algorithm that builds an SDMA group by sequentially adding MSs spatially compatible to the MSs already in the group [20].

The ES algorithm finds the group $\mathcal{G}^{\star}$ that maximizes the grouping metric. However, it might be too complex because it compares all the L SDMA groups and L increases combinatorially with K.

The RG algorithm is the most simple algorithm. Given a target group size $G_t, 1 \leq G_t \leq M$, which is the number of MSs that the group $\mathcal{G}^{\star}$ must contain, the RG algorithm just selects G_t among the K MSs randomly. A target group size is used by many SDMA algorithms to simplify the search for the best group $\mathcal{G}^{\star}$, cf. Table I.

The CG algorithm has been proposed by the authors in [9] together with the metric $f_{\mathrm{CC}}(\mathcal{G})$ of (19), which can be easily expressed as a function of $\mathbf{u}$ since there is a unique mapping between $\mathbf{u}$ and a group $\mathcal{G}$. Let $\tilde{u}_k$ and $\tilde{\mathbf{u}}$ denote the continuous relaxed versions of u_k and $\mathbf{u}$ in (18), respectively. Denoting by $\mathbf{1}_K$ a $K \times 1$ vector of ones and using (19), the CG algorithm is formulated as the following quadratic optimization problem:

$$\tilde{\mathbf{u}}^{\star} = \arg\min_{\tilde{\mathbf{u}}} \left\{ \frac{(1-\beta)}{\|\mathbf{C}\|_{\mathrm{F}}}\tilde{\mathbf{u}}^{\mathrm{T}}\mathbf{C}\tilde{\mathbf{u}} + \frac{\beta}{\|\mathbf{a}\|_{\mathrm{F}}}\mathbf{a}^{\mathrm{T}}\tilde{\mathbf{u}} \right\}, \tag{20a}$$

subject to

$$\mathbf{1}_K^{\mathrm{T}}\tilde{\mathbf{u}} = G_t, \tag{20b}$$

$$0 \leq \tilde{u_k} \leq 1, \forall k, \tag{20c}$$

$$\tilde{u}_{k'} = 1. \tag{20d}$$

Problem (20) is the relaxed version of the equivalent integer optimization problem when $\mathbf{u}$ is binary [9]. The CG algorithm also considers a target group size G_t in the constraint (20b). By solving problem (20) and rounding to one the G_t largest components and to zero the other $K - G_t$ components of $\tilde{\mathbf{u}}^{\star}$, the CG algorithm finds a group $\mathcal{G}^{\star}$ composed of spatially uncorrelated MSs with low channel attenuation. Constraint (20d) forces an initial MS k' to belong to $\mathcal{G}^{\star}$ and can be used for scheduling purposes.

The BF algorithm has been proposed in [20]. Similarly to the CG algorithm, the BF algorithm also considers an initial MS. The BF algorithm starts with an SDMA group containing only this initial MS. Then, the BF algorithm sequentially extends the group by admitting to it the MS that most improves the grouping metric. Let $\mathcal{G} = \{k'\}$ be the initial group containing only the MS k' and let G be the size of $\mathcal{G}$. Then, the BF algorithm temporarily admits one MS $k \notin \mathcal{G}$ to the group and computes the grouping metric $f_{(\cdot)}(\mathcal{G} \cup \{k\})$. This is done for each MS $k \notin \mathcal{G}$ and the MS that has led to the best metric value when temporarily admitted to $\mathcal{G}$ is permanently inserted into the group. Then, this procedure is repeated for the extended group until a group of size G_t be built or until no more MSs able to improve the grouping metric be found. Because the BF algorithm tests only a small number of candidate SDMA groups and relies on a simple heuristic, it is less complex than the ES and CG algorithms.

The CG and BF algorithms will be used later as part of the new RA strategies proposed in this work.

V. Precoding and Power Allocation Algorithms

In this work, only linear ZF precoding is considered [13], which will be simply termed ZF further on. Because ZF suppresses spatial interference completely, the effective MSs' channels are no longer coupled through interference and power allocation does not affect precoding anymore. Consequently, the RA is simplified. Moreover, it has been shown that the maximum sum rate of the system can be efficiently approximated using ZF [27], [29] and WF for power allocation [40]. WF is a suitable choice since it maximizes the sum rate of the set of independent channels obtained after applying ZF [21], [40].

Despite of its simplicity, ZF precoding is quite sensitive to the quality of the CSIT with its performance degrading rapidly when imperfect CSIT is considered. Linear MMSE-based precoding is more robust to imperfect CSIT. However, it inherently couples precoding and power allocation and, when aiming at the maximization of the sum rate, its adoption asks for the use of iterative algorithms incurring additional complexity. Therefore, linear MMSE-based precoding is not considered in this work, but only linear ZF precoding.

VI. Sequential Removal Algorithm

In this section, an SR algorithm is proposed, which intends to increase the capacity of the groups built by SDMA algorithms unaware of precoding and power allocation. The SR algorithm removes the MSs that do not contribute to enhance the sum rate from the group $\mathcal{G}$. Using ZF and WF, if zero power is allocated to an MS, it does not contribute to enhance the group capacity anymore. On the contrary, since the channels of the others MSs in $\mathcal{G}$ are projected onto the null-space of the channel of this one MS due to ZF, its removal can only improve the group capacity.

The SR algorithm removes one MS from the SDMA group $\mathcal{G}$ according to the effective channel gain of the MSs [21]. This is a reasonable criterion since the lower the effective channel gain of an MS is, the lower its achievable capacity is. Anyway, other criteria may be used [9]. After removing an MS, the SR algorithm computes and stores the capacity for the resulting SDMA group using (12). Then, the process is repeated and another MS is removed, and so on. At the end, the SDMA group with the highest capacity is kept as the best SDMA group $\mathcal{G}^\star$.

For an initial group $\mathcal{G}$ of size G, the SR algorithm needs to compute G group capacities using (12). Because G is relatively small and because the size of $\mathcal{G}$ is sequentially reduced, these computations add only slightly to the complexity of RA strategies using the SR algorithm. Nevertheless, the proposed SR algorithm can provide considerable gains to the system in terms of sum rate.

VII. Resource assignment algorithm

A. Group priority

In this section, the group priorities used by the resource assignment algorithm are described. The concept of MS priorities has been often used in time-scheduling algorithms to manage the QoS of the MSs and, e.g., to provide throughput fairness [41]. Because SDMA groups may contain several MSs, the concept of MS priorities is extended to group priorities later in this section. Priorities are defined according to:

- A Capacity Maximization (CM) criterion, which aims at maximizing the sum rate.
- A Proportional Fair (PF) criterion, which finds a trade-off between the QoS of the MSs and the sum rate.

Let $\mathfrak{u}_{i,b}$ denote the priority of MS i in the SDMA group $\mathcal{G}_{l,b}$. For the CM criterion, $\mathfrak{u}_{i,b}$ is defined as

$$\mathfrak{u}_{i,b} = R_{i,b}. \tag{21}$$

For the PF criterion, let R_i^c and $\bar{R}_i$ denote the contracted and the perceived average throughputs of MS i in $\mathcal{G}_{l,b}$, respectively. Thus, the throughput ratio $\frac{R_i^c}{\bar{R}_i}$ measures how well the MS has met its QoS requirements [42]. For the PF criterion, $\mathfrak{u}_{i,b}$ is defined as

$$\mathfrak{u}_{i,b} = \frac{R_i^c}{\bar{R}_i} R_{i,b}. \tag{22}$$

Let $\mathfrak{v}_{l,b}$ denote the group priority of the group $\mathcal{G}_{l,b}$, which quantifies the efficiency of assigning the resource b to it. It is proposed here to define the group priority $\mathfrak{v}_{l,b}$ simply as the sum of the priorities $\mathfrak{u}_{i,b}$ of the MSs in $\mathcal{G}_{l,b}$, i.e.,

$$\mathfrak{v}_{l,b} = \sum_{i \in \mathcal{G}_{l,b}} \mathfrak{u}_{i,b}, \tag{23}$$

which leads to

$$\mathfrak{v}_{l,b} = \begin{cases} f_{\mathrm{CAP}}(\mathcal{G}_{l,b}), & \text{for the CM criterion, and} \quad (24a) \\ \displaystyle\sum_{i \in \mathcal{G}_{l,b}} \frac{R_i^c}{\bar{R}_i} R_{i,b}, & \text{for the PF criterion.} \quad (24b) \end{cases}$$

According to (21), if CSI is available on a frame basis all the TSs of a frame are assigned to the same SDMA group $\mathcal{G}_{l,b}$. In order to improve fairness, RA is considered on a TS basis with the PF criterion in order to assign resources to potentially different groups during each TS. Anyway, the same CSIT is considered for all the TSs of a frame.

For the CM criterion, the higher the group capacity of an SDMA group on a resource b is, the higher its priority on this resource is and, consequently, the higher the chances of the group getting this resource assigned. For the PF criterion, the rates R_i of the MSs in a group are scaled by the throughput ratio $R_i^c/\bar{R}_i$. Thus, SDMA groups containing MSs achieving high rates R_i or MSs whose QoS requirements have not been fulfilled will have high priority [42] and, consequently, there will be higher chances of assigning resources to these groups.

B. Assignment algorithm

In this section, the assignment algorithms considered in this work are described. Using the group priorities, the assignment algorithm has to determine which resource to assign to which SDMA group. Two assignment algorithms will be considered:

- A sequential algorithm, which assigns resources one-by-one to SDMA groups.
- A resource-to-group algorithm, which is proposed here and is formulated as a standard assignment problem. It assigns at once the B resources to B SDMA groups selected from a set of $L > B$ candidate groups.

In fact, the sequential algorithm corresponds to the single-resource case of Section II. It just assigns the considered resource to the group built by the SDMA algorithm.

The proposed resource-to-group algorithm considers the multiple-resource case and is described in the sequel. Initially, a set of K candidate SDMA groups is built on each resource b. The k^{th} group is built by selecting the k^{th} MS as initial MS k' and applying an SDMA algorithm. Consequently, $\tilde{L} = K \cdot B$ groups are built. Precoding, power allocation, and the SR algorithms are applied for each group considering EPA among resources. In order to improve fairness, groups of same composition built on different resources are considered only once, i.e., only a number $L \leq K \cdot B$ of unique groups from the $\tilde{L}$ groups is considered. Then, it is proposed here to formulate the assignment of resources to SDMA groups as a standard assignment problem based on the group priorities, which is solved using Munkres' algorithm [43]. The referred formulation is described in the sequel.

Let $\mathfrak{V}$ denote a group priority matrix containing the group priorities $\mathfrak{v}_{b,l}$ of each SDMA group on each resource. Let $\mathbf{V}$ denote a resource-to-group assignment matrix whose binary entries indicate whether the b^{th} resource is assigned to the l^{th} group. The matrices $\mathfrak{V}$ and $\mathbf{V}$ are written as

$$\mathfrak{V} = \begin{bmatrix} \mathfrak{v}_{1,1} & \dots & \mathfrak{v}_{1,L} \\ \mathfrak{v}_{2,1} & \dots & \mathfrak{v}_{2,L} \\ \vdots & \ddots & \vdots \\ \mathfrak{v}_{B,1} & \dots & \mathfrak{v}_{B,L} \end{bmatrix}, \text{ and } \mathbf{V} = \begin{bmatrix} v_{1,1} & \dots & v_{1,L} \\ v_{2,1} & \dots & v_{2,L} \\ \vdots & \ddots & \vdots \\ v_{B,1} & \dots & v_{B,L} \end{bmatrix}, \tag{25}$$

respectively. Denoting by $\odot$ the Hadamard product and using (25), the assignment of resources to groups is formulated as

$$\mathbf{V}^{\star} = \arg\max_{\mathbf{V}} \left\{ \mathbf{1}_B^{\mathrm{T}} \left(\mathfrak{V} \odot \mathbf{V} \right) \mathbf{1}_L \right\} \tag{26a}$$

subject to

$$\mathbf{V}\mathbf{1}_L = \mathbf{1}_B \tag{26b}$$

$$\mathbf{V}^{\mathrm{T}}\mathbf{1}_B \leq \mathbf{1}_L, \tag{26c}$$

which is a standard assignment problem that can be efficiently solved using Munkres' algorithm [43]. After solving (26), the resources assigned to the groups are determined by the non-zero entries of $\mathbf{V}^{\star}$ and the B resources are assigned to B out of the $L > B$ candidate groups. Constraints (26b) and (26c) impose that no more than one resource be assigned to the same group, which may increase the fairness into the system. Moreover, to avoid assigning an unsuitable resource to an SDMA group, the group priority is computed only for the resources on which the SDMA group has been built and is set to zero on the other resources.

In both the sequential and resource-to-group algorithms, the SDMA grouping problem is solved beforehand, i.e., first the groups are built and then resources are assigned to them, which keeps SDMA and resource assignment algorithms relatively separated from each other.

VIII. RA Strategy Definition

In this section, the suboptimal RA strategies investigated in this work are defined by combining the algorithms introduced in Sections IV to Section VI.

Table II defines the RA strategies, whose names are given in the first column. The subsequent columns specify the SDMA, resource assignment, precoding, power allocation algorithms employed by each strategy. The last column indicates whether the SR algorithm is used by the strategy.

The ES strategy searches exhaustively for SDMA the group that maximizes the sum rate on each resource, while the RG strategy just builds randomly an SDMA group on each resource. These two strategies are considered to bound the performance of the other strategies in Table II above and below, respectively.

RA strategies whose SDMA algorithms combine $f_{\text{CAP}}(\mathcal{G})$ and $f_{\text{SP}}(\mathcal{G})$ with the BF algorithm are defined in Table II, namely the CAP-BF and the SP-BF strategies. They are used for comparison with the new CC-BF and CC-CG strategies proposed here, whose SDMA algorithms combine f_{CC} with the BF and CG algorithms, respectively. SDMA algorithms combining f_{CAP} and the BF algorithm have been studied, e.g., in [8], [26], [27]. SDMA algorithms combining f_{SP} and the BF have been studied, e.g., in [10], [11], [27], [29]. Combinations of f_{CAP} and f_{SP} with the CG algorithm are not considered because they do not permit an adequate formulation of the SDMA grouping problem as a quadratic optimization problem.

The CM criterion and the sequential algorithm are considered in the ES and RG strategies. The performance of these strategies is the same in both single- and multiple-resource cases since EPA among resources is used. For the remaining strategies in Table II, both the sequential and resource-to-group algorithms are considered. The sequential algorithm related to the single-resource case is considered in combination with the CM criterion only. The resource-to-group algorithm will be considered with both the CM and PF criteria. In particular for the PF criterion, it is of interest to investigate whether throughput fairness can be considerably enhanced at the expense of only small reductions of the sum rate.

For all the strategies, ZF and WF are considered for precoding and power allocation, respectively. The ES and CAP-BF strategies employ the group capacity in their SDMA algorithms, which are consequently aware of the actual precoding and power allocation. Therefore, the SR algorithm is no longer necessary in these strategies and is disabled.

IX. Analysis and Results

In this section, the performance of the RA strategies of Table II is investigated. The BS is equipped with a Uniform Linear Array (ULA) with $M = 4$ omnidirectional elements separated by half wavelength. A total number $K = 16$ of single-antenna MSs is served by the BS.

A center frequency $f_0 = 5$ GHz is considered. A total number $B = 8$ of frequency blocks composed of $Q_{\text{sub}} = 6$ adjacent subcarriers of bandwidth $\Delta_f \approx 9.766$ kHz are considered. Fast fading is generated using the WIM considering the urban macro-cell scenario C2 with Non Line Of Sight (NLOS) [32]. The channel has a coherence bandwidth $B_{\text{c}} \approx 250$ kHz, so that the frequency block bandwidth $Q_{\text{sub}}\Delta f < 0.25B_{\text{c}}$. An average MS speed $v_{\text{MS}} \approx 2.78$ m/s is assumed, which leads to a coherence time $T_{\text{c}} \approx 11$ ms. Frames of duration $T_{\text{FRM}} = 1$ ms are considered, so that $T_{\text{FRM}} < 0.1T_{\text{c}}$. Each frame is composed of $T = 4$ TSs.

Because the larger the SDMA group, the higher the potential SDMA gains, a target group size $G_t = M = 4$ corresponding to the maximum admissible group size is considered [8]–[11], [15], [29]. For the strategies employing the BF and the CG algorithms, the initial MS k' is selected as the one with the highest channel gain. For the CC-CG strategy, $\beta = 0.5$ is used, which has been experimentally adjusted as in [9]. The BS is assumed to always have data to send to the MSs. All MSs are assumed to have the same contracted average throughput R^c. The most relevant simulation parameters are listed in Table III.

Initially, the performance of the RA strategies of Table II is studied considering the single-resource case, the CM criterion, and the sequential algorithm, so that each resource is assigned to the group built by the SDMA algorithm. After

Table II
RA STRATEGIES DEFINITION.

RA	SDMA Algorithm		Resource Assignment Algorithm		Precoding	Power Allocation	SR
Strategy	Grouping Metric	Grouping Algorithm	Group priority	Assignment Algorithm	Algorithm	Algorithm	
ES	f_{CAP}	Exhaustive Search	Cap. Maximization	Sequential	Linear Zero-Forcing	Water Filling	Off
RG	-	Random Grouping					On
CAP-BF	f_{CAP}	Best Fit	Cap. Maximization, Proportional Fair	Sequential, Resource-to-Group			Off
SP-BF	f_{SP}						On
CC-BF	f_{CC}						
CC-CG	f_{CC}	Convex Grouping					

Table III
SIMULATION PARAMETERS.

Parameter	Symbol	Value	Unit
BS's ULA size	M	4	-
Number of MSs	K	16	-
Center frequency	f_0	5.0	GHz
Number of resources	B	8	-
Subcarriers / freq. block	Q_{sub}	6	-
Subcarrier bandwidth	Δf	9.766	kHz
Fast fading model	-	WIM, scenario C2	-
Average MSs' speed	v_{MS}	$\approx$2.78	m/s
Frame duration	T_{FRM}	1	ms
TSs / frame	T	4	-
Target group size	G_t	4	MSs
Initial MS	k'	$\arg\max_k \left\{\|\hat{\mathbf{h}}_k\|_2^2\right\}$	-
Parameter for $f_{\text{CC}}(\mathcal{G})$	β	0.5	-

that, precoding and power allocation are applied, as well as the SR algorithm if necessary. Fig. 2 shows the average sum rate achieved by the RA strategies of Table II as a function of the average SNR γ.

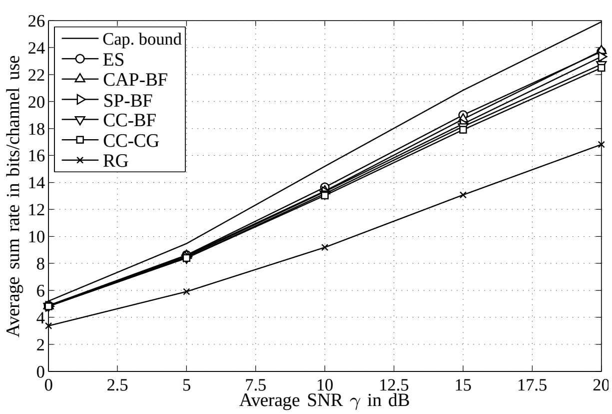


Figure 2. Average sum rate of the RA strategies of Table II.

It can be seen in Fig. 2 that the CAP-BF, SP-BF, CC-BF, and CC-CG strategies achieve over 95% of the average sum rate achieved by the ES strategy for all the considered average SNR values and that the performance of the ES strategy is only about 10% lower than the capacity upper bound [14] achiveable using Dirty Paper Coding (DPC). The performance gap between the ES strategy and the other strategies increases only slightly for higher values of γ. Since the ES strategy is much more complex than the other RA strategies, it can be noted that suboptimal RA strategies are able to efficiently approach the maximum sum rate of the system. As expected, the RG strategy performs worst and obtains about 70% of average sum rate achieved by the ES strategy.

Since quite different RA strategies are considered in Fig. 2, it is important to verify whether some of them are particularly more robust or sensitive to imperfections in the CSIT. For this purpose, the erroneous CSIT model of (4) is employed. Fig. 3 shows the average sum rate achievable by the RA strategies considering an average SNR γ of 10 dB and varying quality of the CSIT, given by γ_{CSI} in (5). The presented results correspond to the maximum rates that the RA strategies could ideally achieve, i.e., considering perfect feedback and rate adjustments while assuming that the erroneous CSI is the actual CSI.

If the quality of the CSIT is somehow compromised, schemes relying on Channel Quality Indicators (CQIs), such as SNR values, might become interesting alternatives. In order to illustrate this fact, the performance of a Grid of Beams (GoB) with four beams of equal power formatted using Chebyshev filtering with a Sidelobe Level (SLL) attenuation of 20 dB [44] is also included in Fig. 3 and is indicated by the horizontal dashed line. Since CQI values represent a small amount of information to be fed back to the BS by the MSs, it is assumed in Fig. 3 that they are correctly received and that the four MSs with the best SNR values are served by the BS.

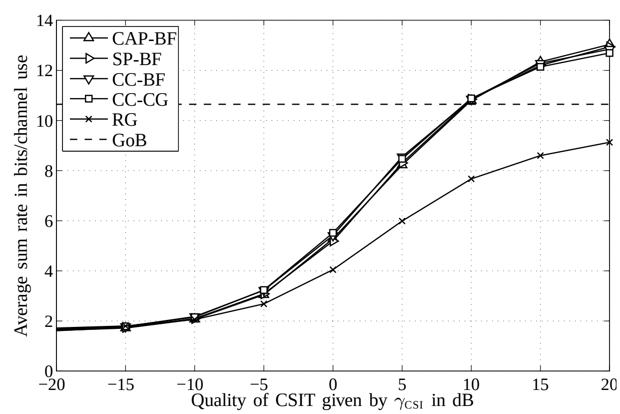


Figure 3. Average sum rate of the RA strategies considering imperfect CSIT. Average SNR $\gamma = 10$ dB

From Fig. 3, it can be verified that none of the RA strategies is particularly more robust against imperfections in the CSIT. It can also be noted that the performance of the strategies rapidly degrades when γ_{CSI} decreases. In order to obtain at least 60% of the sum rates shown in Fig. 2 for different average SNR values, the quality of the CSIT given by γ_{CSI} should not be lower than 10 dB. For a γ_{CSI} value of -10 dB, the

RA strategies attain only about 15% of their sum rates with perfect CSIT. Considering the parameter values in Table III and assuming that CSIT imperfections are only due to delays, this γ_{CSI} value corresponds to a delay of approximately 50% of the channel coherence time T_c [29], [37], [38]. Thus, the performance of considered strategies strongly depends on the quality of the CSIT.

Regarding the use of GoB, it can be seen in Fig. 3 that the achieved sum rate is only about 15% lower than those obtained by the considered RA strategies. Indeed, the performance of the GoB is even better than that of RG strategy. For the other RA strategies, a γ_{CSI} value of, at least, 10 dB is required in order to ensure better performance. Therefore, the use of GoB can represent an efficient alternative to more sophisticated RA when the quality of the CSIT is compromised. A more detailed investigation of this topic, e.g., when CQIs are also imperfect, is left for future studies.

In Fig. 2 and Fig. 3, the RG strategy achieves quite good sum rates in spite of being considerably more simple than the remaining strategies. However, the sum rates achieved by the RG strategy are mainly due to the use of the proposed SR algorithm of Section VI. Indeed, the SP-BF, CC-BF, CC-CG, and RG strategies employ the SR algorithm, which provides considerable gains in terms of average sum rate. In order to show the impact of the SR algorithm, the average sum rates achieved by SP-BF, CC-BF, CC-CG, and RG strategies when the SR algorithm is switched off has been evaluated and compared to the values in Fig. 2. The percentual reduction of the average sum rate of the referred strategies when the SR algorithm is switched off is shown in Fig. 4.

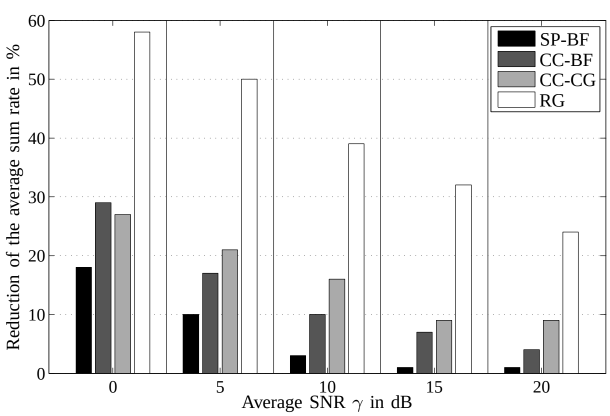


Figure 4. Percentual reduction of the average sum rates when the SR algorithm is switched off.

It can be noted that sum rate values lower than those presented in Fig. 2 are obtained if the SR algorithm is switched off. In particular for the RG strategy, losses are quite large and surpass 50% for low average SNR values. Indeed, for low average SNR values, the ideal SDMA group size is much smaller than M. However, because the group capacity is not a monotonic function in the group size, G_t cannot be determined a priori and the SR algorithm must be employed. Alternatively, a value $1 \leq G_t \leq M$ or a small set of values in this range could be used in the RA strategy and the complexity of the SR algorithm could be considerably reduced [9], [26]. Anyway, comparing Fig. 2 and Fig. 4, it can be seen that the proposed SR algorithm considerably improves the performance of the RA strategies.

From Fig. 2, Fig. 3, and Fig. 4, it can be concluded that the CAP-BF, SP-BF, CC-BF, and CC-CG strategies have very similar performance and approximate quite well the average sum rate of the ES strategy. However, these RA strategies have quite different complexities. Indeed, previous works lack on more precise analysis of the complexity of the RA strategies. Herein, the complexity of each strategy has been estimated in terms of the required number of complex multiplications and is given in Table IV as a function of the number K of MSs, the number M of transmit antennas, and the target group size G_t. Moreover, their complexity orders $\mathcal{O}(\cdot)$ assuming $G_t = M$, cf. Table II, are also given in Table IV.

Because the number L of groups considered by the ES strategy combinatorially increases with K and because precoding and power allocation must be computed for each group, this strategy has the highest complexity, which is non-polynomial. The RG strategy has the lowest complexity, which does not depend on the number of MSs. However, it presented the worst performance in terms of average sum rate.

Observing the complexity orders of the RA strategies, it can be seen that the CAP-BF strategy is more complex than the SP-BF strategy, which on its turn is more complex than the proposed CC-BF strategy. The complexity order of the proposed CC-CG strategy is quadratic in K but linear in M and, consequently, it might be more or less complex than the CAP-BF and SP-BF strategies depending on the number of MSs and on the size of the BS array. Moreover, the complexity of the CC-CG strategy depends on the number of iterations I_{CG} required by the CG algorithm to converge.

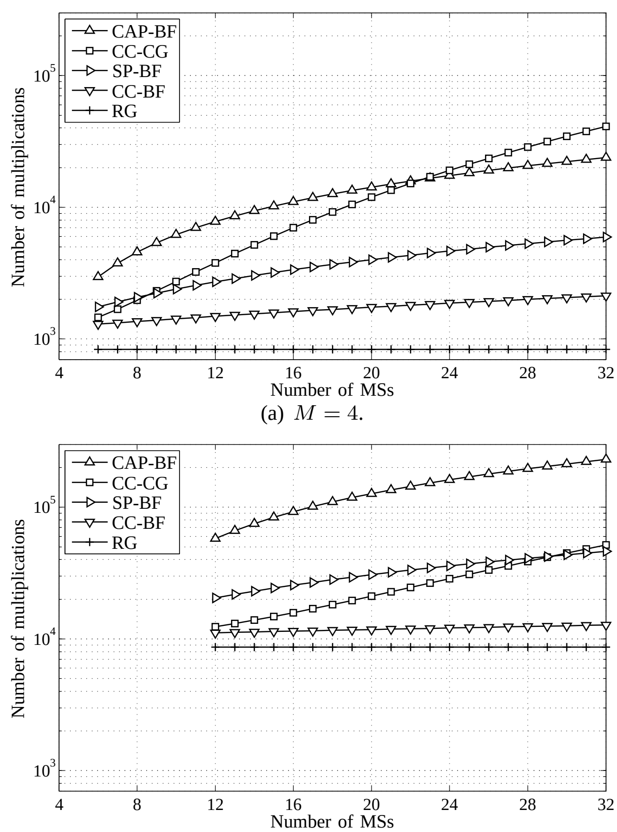
Because in practice the number K of MSs and the number M of antennas are limited to relatively small values, a useful estimation of the complexity order of the RA strategies might be difficult, especially because coefficients in the expressions in Table IV cannot be disregarded. In Fig. 5, the complexity of the RA strategies is shown for a varying number K of MSs and a fixed number M of transmit antennas. It is assumed that $G_t = M$, cf. Table III, and that the CG algorithm requires $I_{\text{CG}} = \frac{K}{2}$ iterations to converge.

As it can be noted, the proposed CC-BF strategy has lower complexity than the CAP-BF and SP-BF strategies in all the cases and is only slightly more complex than the RG strategy. By comparing Fig. 5(a) and Fig. 5(b), it can also be noted that the proposed CC-CG strategy has lower complexity than the CAP-BF and SP-BF strategies for large array sizes. Performing the best in terms of average sum rate, the CAP-BF strategy has a considerably higher complexity compared to the SP-BF and CC-BF strategies, which offer therefore a better performance-complexity trade-off.

In the following, multiple-resources are considered and the proposed resource-to-group algorithm will be used to improve the throughput fairness among the MSs in the system. Only the CAP-BF, SP-BF, and CC-BF strategies are considered in the sequel. The CAP-BF strategy approximates well the performance of the ES strategy, while the SP-BF and CC-BF strategies offer better trade-off between average sum rate and

Table IV
COMPLEXITY OF THE RA STRATEGIES.

RA strategy	Number of multiplications: SDMA algorithm	+	Precoding, power allocation, and SR algorithms	$\mathcal{O}(\cdot)$ for $G_t = M$
ES	$\left\{\sum_{G=2}^{G_t} \frac{K!}{G!(K-G)!}\left(\frac{G(G^2+7G(M+1)+9M+18)}{2}\right)\right\}$	+	$\{KM+6M+10\}$	Non-Polynomial
CAP-BF	$\left\{\sum_{G=2}^{G_t}(K-G+1)\left(\frac{G(G^2+7G(M+1)+9M+18}{2}\right)\right\}$	+	$\{KM+6M+10\}$	$\mathcal{O}(4KM^3)$
SP-BF	$\left\{\sum_{G=2}^{G_t}(K-G+1)\left(\frac{5M^2+5M+2}{2}\right)\right\}$	+	$\left\{KM+6M+10+\sum_{G=2}^{G_t}\frac{G^3+7G^2(M+1)+9GM+18G}{2}\right\}$	$\mathcal{O}\left(\frac{5KM^2}{2}\right)$
CC-BF	$\left\{\frac{G_t}{K}\frac{K^2(M+8)+K(M+2)+6}{2}\right\}$	+	$\left\{KM+6M+10+\sum_{G=2}^{G_t}\frac{G^3+7G^2(M+1)+9GM+18G}{2}\right\}$	$\mathcal{O}\left(\frac{KM^2}{2}\right)$
CC-CG	$\left\{I_{\text{CG}}(2K^2+2K)+\frac{K^2(M+8)+K(M+2)+6}{2}\right\}$	+	$\left\{KM+6M+10+\sum_{G=2}^{G_t}\frac{G^3+7G^2(M+1)+9GM+18G}{2}\right\}$	$\mathcal{O}\left(\frac{K^2(M+4I_{\text{CG}})}{2}\right)$
RG		+	$\left\{6M+10+\sum_{G=2}^{G_t}\frac{G^3+7G^2(M+1)+9GM+18G}{2}\right\}$	$\mathcal{O}(4M^3)$



(a) $M = 4$.

(b) $M = 8$.

Figure 5. Complexity of the RA strategies.

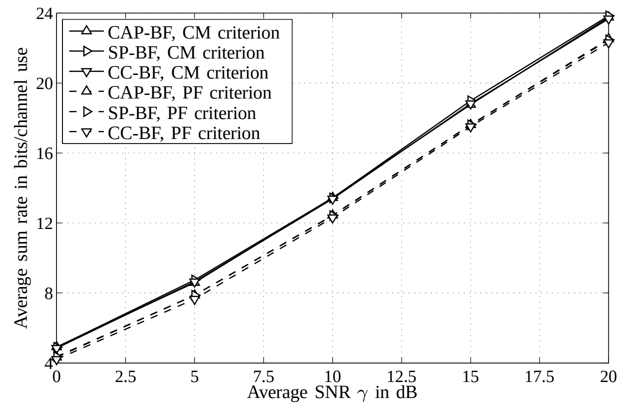


Figure 6. Average sum rate of the CAP-BF, SP-BF, and CC-BF strategies considering the CM and PF criteria for resource-to-group algorithm.

complexity. Because the proposed CC-CG strategy performs only as good as the proposed CC-BF strategy while being more complex, it is not considered in the sequel.

Initially, it is important to see how both the selection of B out of the L groups and the adoption of the PF criterion influence the average sum rate achieved by the RA strategies. In Fig. 6, the average sum rate achieved by the CAP-BF, SP-BF, and CC-BF strategies is shown as a function of the average SNR γ for the CM and PF priority criteria.

Comparing Fig. 2 and Fig. 6, it can be seen that the gap between the average sum rate achieved by the CAP-BF, SP-BF, and CC-BF strategies and the average sum rate obtained by the ES strategy is eliminated by allowing the suboptimal strategies to select among a larger number of candidate groups. Such an improvement comes at the expense of extra complexity since L candidate groups must be considered. Anyway, the CAP-BF, SP-BF, and CC-BF strategies remain substantially less complex than the ES strategy.

Comparing the PF and CM criteria in Fig. 6, only a reduction of about 10% is observed in the average sum rate achieved by the CAP-BF, SP-BF, and CC-BF strategies. The strategies also have the same performance in this case and the proposed CC-BF strategy offers again the best trade-off between performance and complexity.

In order to evaluate the impact of the adoption of the PF criterion on the throughput fairness among the MSs, Jain's Index of Fairness (JIF) $\mathcal{J}(\cdot)$ is employed [21]. For the average throughput of the MS, JIF is given by

$$\mathcal{J}(\bar{R}_k) = \frac{\left(\sum_{k=1}^{K}\bar{R}_k/R_k^c\right)^2}{K\sum_{k=1}^{K}\left(\bar{R}_k/R_k^c\right)^2}, \tag{27}$$

and assumes values between $\frac{1}{K}$ and 1 [21]. The higher the values JIF assumes, the more fair the throughput distribution among the MSs is. In particular, a value J of JIF can be interpreted as having $100\times J$% of the MSs being fairly served, i.e., perceiving the same throughput, and $100 \times (1 - J)$% of the MSs perceiving no throughput at all.

For the CAP-BF, SP-BF, and CC-BF strategies, Fig. 7 shows the average throughput fairness among the MSs after a varying number of frames. Both the CM and PF priority criteria are considered and an average SNR $\gamma = 10$ dB is assumed.

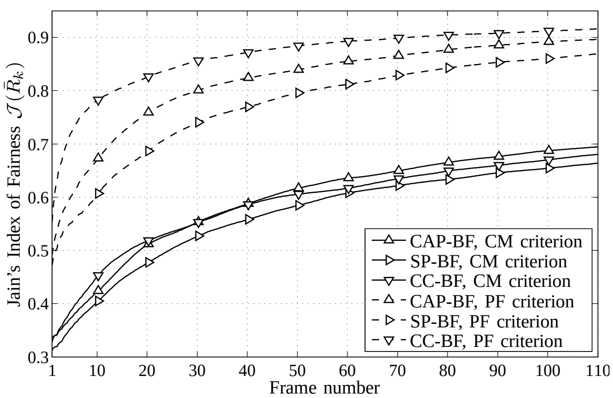


Figure 7. Jain's Index of Fairness for the CAP-BF, SP-BF, and CC-BF strategies considering the CM and PF criteria for resource-to-group algorithm. Average SNR $\gamma = 10$ dB.

It can be seen that the proposed resource-to-group algorithm with the PF criterion considerably improves the throughput fairness among MSs at the expense of only the small reductions of the average sum rate shown in Fig. 6.

Considering the PF criterion and the CC-BF strategy, it can be seen that about 90% of the MSs can be assumed as fairly served after about 60 ms (60 frames), while only about 60% of the MSs are fairly served considering the CM criterion. Because the SDMA groups built by the SP-BF strategy strongly favor MSs with high channel gains, this strategy shows slightly worse fairness figures than the CC-BF strategy. The CAP-BF strategy presents a slightly more fair throughput distribution than the SP-BF strategy. This occurs because the CAP-BF strategy is aware of the precoding and power allocation and can estimate the group capacity better than the SP-BF strategy, thus not favoring so much the MSs with high channel gain.

Considering the results in Fig. 6 and Fig. 7, it can be seen that the proposed CC-BF strategy considering the PF criterion also offers a good trade-off between average sum rate and throughput fairness.

X. Conclusions

In this work, several suboptimal RA strategies for the maximization of the sum rate of an MU MIMO-OFDMA system have been investigated. Two RA strategies have been proposed, namely the CC-CG and CC-BF strategies, which have been shown to achieve almost the same sum rate as the ES strategy. The complexity of the proposed strategies has been estimated and it has been shown that they provide better performance-complexity trade-offs than some existing RA strategies considered for benchmarking. An SR algorithm has been proposed, which provides considerable gains in terms of average sum rate and offers a good solution to determine the size of the SDMA groups considered in the system. Two simple resource assignment algorithms are considered. In particular, the proposed resource-to-group algorithm combined with a PF priority criterion has been shown to considerably improve the throughput fairness among MSs at the expense of only small reductions of the sum rate.

From the proposed model for suboptimal RA strategies, simple rules can be defined to implement efficient suboptimal RA strategies:

- Build groups using a low-complexity rather efficient SDMA algorithm that takes into account spatial compatibility.
- Adjust the size of the SDMA groups using an SR algorithm and taking into account precoding and power allocation.
- If the maximization of the sum rate is pursued, allocate resources sequentially according to the group capacity.
- If a good trade-off between fairness and sum rate is to be found, build various SDMA groups and assign the resources to a subset of the groups while taking into account group priorities.

As it has been seen, several RA strategies fit into the proposed model and from the results presented in this work it has been seen that efficient, low-complexity suboptimal RA strategies can be designed to maximize the sum rate of the system while providing a good degree of throughput fairness.

References

[1] 3GPP, "Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA)," 3rd. Generation Partnership Project, Tech. Rep. TR 25.814 V7.1.0, Sept. 2006. [Online]. Available: http://www.3gpp.org

[2] IEEE, "Air interface for fixed broadband wireless access systems," Institute of Electrical and Electronics Engineers, Tech. Rep. IEEE 802.16-2004, Oct. 2004.

[3] M. Doettling, M. Sternad, G. Klang, J. von Hafen, and M. Olsson, "Integration of spatial processing in the WINNER B3G air interface design," in *Proc. of the IEEE Vehic. Tech. Conf. (VTC)*, vol. 1, May. 2006, pp. 246–250.

[4] H. Liu and G. Li, *OFDM-based broadband wireless networks - design and optimization*, 1st ed. Wiley & Sons, 2005.

[5] I. E. Telatar, "Capacity of multi-antenna gaussian channels," Lucent Technologies / Bell Labs Innovations, Technical Memorandum, Oct. 1995. [Online]. Available: http://mars.bell-labs.com/cm/ms/what/mars/papers/proof/

[6] K. B. Letaief and Y. J. Zhang, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 38–47, Aug. 2006.

[7] Q. H. Spencer and A. L. Swindlehurst, "Channel allocation in multi-user MIMO wireless communications systems," in *Proc. of the IEEE Internat. Conf. on Commun. (ICC)*, vol. 5, Jun. 2004, pp. 3035–3039.

[8] T. F. Maciel and A. Klein, "A low-complexity SDMA grouping strategy for the downlink of Multi-User MIMO systems," in *Proc. of the IEEE Personal, Indoor and Mob. Radio Commun. (PIMRC)*, Sept. 2006.

[9] ——, "A convex quadratic SDMA grouping algorithm based on spatial correlation," in *Proc. of the IEEE Internat. Conf. on Commun. (ICC)*, Jun. 2007.

[10] ——, "On the performance of SDMA with soft dropping and SINR balancing power control in the downlink of multi-user MIMO systems," in *Proc. of the IEEE Workshop on Smart Antennas*, Feb. 2007.

[11] ——, "Resource allocation strategies for SDMA/OFDMA systems," *Lecture Notes in Electrical Engineering (Advances in mobile and wireless communications - views of the 16th IST Mobile and Wireless Communications Summit)*, vol. 16, pp. 115–132, May 2008.

[12] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[13] M. Joham, "Optimization of linear and nonlinear transmit signal processing," Ph.D. dissertation, Munich University of Technology, Munich, Germany, Jun. 2004.

[14] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[15] P. Tejera, W. Utschick, G. Bauch, and J. A. Nossek, "Subchannel allocation in multiuser Multiple-Input-MultipleOutput systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4721–4733, Oct 2006.

[16] V. Stankovic, "Multi-user MIMO wireless communications," Ph.D. dissertation, Technische Universität Ilmenau, Ilmenau, Germany, Nov. 2006.

[17] A. Furuskär, "Radio resource sharing and bearer service allocation for multi-bearer service, multi-access wireless networks - methods to improve capacity," Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, May. 2003.

[18] G. Nemhauser and L. Wosley, *Integer and combinatorial optimization*. Wiley & Sons, 1999.

[19] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st ed. Cambridge Univ. Press, 2004.

[20] F. Shad, T. D. Todd, V. Kezys, and J. Litva, "Dynamic Slot Allocation (DSA) in indoor SDMA/TDMA using a smart antenna basestation," *IEEE/ACM Trans. Netw.*, vol. 9, no. 1, pp. 69–81, Feb. 2001.

[21] B. Calvo, "Fairness analysis of wireless beamforming schedulers," Ph.D. dissertation, Universitat Politénica de Catalunya, November 2004.

[22] J. Sethuraman, *Handbook of scheduling: algorithms, models, and performance analysis*. Chapman & Hall/CRC, 2004, ch. Convex quadratic relaxations in scheduling.

[23] Y. M. Tsang and R. S. Cheng, "Rate maximization in space-division multiple access (SDMA)/multi-input-single-output (MISO)/OFDM systems under QoS and power constraints," in *Proc. of the IEEE Internat. Conf. on Commun. (ICC)*, vol. 5, Jun. 2004, pp. 2970–2974.

[24] P. W. C. Chan and R. S. Cheng, "Capacity maximization for zero-forcing MIMO-OFDMA downlink systems with multiuser diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1880–1889, May. 2007.

[25] R. Stridh, M. Bengtsson, , and B. Ottersten, "System evaluation of optimal downlink beamforming with congestion control in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 743–751, Apr. 2006.

[26] M. Fuchs, G. Del Galdo, and M. Haardt, "Low-complexity space-time-frequency scheduling for MIMO systems with SDMA," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2775–2784, Sept. 2007.

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

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

[29] T. Yoo and A. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *Proc. of the IEEE Internat. Conf. on Commun. (ICC)*, vol. 1, May. 2005, pp. 542–546.

[30] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 107–116, Jan. 2005.

[31] E. Matskani, N. D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Joint multiuser downlink beamforming and admission control: a semidefinite relaxation approach," in *Proc. of the IEEE Internat. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, vol. 3, no. 585-588, Apr. 2007.

[32] WINNER Project, "Final report on identified RI key technologies, system concept, and their assessment," IST, Tech. Rep. D2.10, Dec. 2005. [Online]. Available: https://www.ist-winner.org/DeliverableDocuments/D2.10.pdf

[33] J. Salo, G. Del Galdo, J. Salmi, P. Kyösti, L. Hentilä, M. Milojevic, D. Laselva, P. Zetterberg, and C. Schneider, "Matlab implementation of the WINNER phase I channel model ver1.5," WINNER Project, Tech. Rep., Dec. 2005. [Online]. Available: https://www.ist-winner.org/

[34] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[35] V. Stankovic, M. Haardt, and M. Fuchs, "Combination of block diagonalization and THP transmit filtering for downlink beamforming in multi-user MIMO systems," in *European Conference on Wireless Technology*, 2004, pp. 145–148.

[36] T. Yoo and A. Goldsmith, "Capaciy and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May. 2006.

[37] A. Kühne and A. Klein, "Adaptive subcarrier allocation with imperfect channel knowledge versus diversity techniques in a multi-user OFDM-system," in *Proc. of the IEEE Personal, Indoor and Mob. Radio Commun. (PIMRC)*, Sept. 2007.

[38] ——, "An analytical consideration of imperfect CQI feedback on the performance of a multi-user OFDM-system," in *Proc. of the Internat. OFDM Workshop*, Aug. 2007.

[39] T. F. Maciel and A. Klein, "A performance-complexity analysis of four suboptimal SDMA algorithms," in *Proc. of the IEEE Workshop on Smart Antennas*, Feb. 2008.

[40] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 687–695, Feb. 2005.

[41] H. Zhang, "Service disciplines for guaranteed performance service in packet-switching networks," *Proc. IEEE*, vol. 83, no. 10, pp. 1374–1396, Oct. 1995.

[42] P. Svedman, S. K. Wilson, L. J. C. Jr., and B. Ottersten, "Opportunistic beamforming and scheduling for OFDMA systems," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 941–952, May 2007.

[43] F. Bourgeois and J.-C. Lassalle, "An extension of the Munkres algorithm for the assignment problem to rectangular matrices," *Communications of the ACM*, vol. 14, no. 12, pp. 802–804, Dec. 1971.

[44] H. Rohling, R. Grünheid, and A. Tassoudji, "Beamforming and scheduling in a cellular OFDM system," in *Proc. of the IEEE Workshop on Smart Antennas*, Feb. 2008.

Tarcisio F. Maciel (S'01-M'09) received the B.Sc. and M.Sc. degrees from the Federal University of Ceará, Fortaleza, Brazil, in 2002 and 2004, respectively, and the Dr.-Ing. degree from the Technische Universität Darmstadt, Germany, in 2008, all in electrical engineering.

From 2001 to 2004 he was a researcher at the Wireless Telecom Research Group - GTEL, Fortaleza, Brazil. From 2005 to 2008 he was a research assistant at the Communications Engineering Lab of the Technische Universität Darmstadt, Germany. Since 2008 he has been with GTEL as a senior researcher. Since 2009, he is a professor at the Computer Engineering Course at the Federal University of Ceará. His research interests include multi-user and multi-antenna communications, and radio resource management.

Anja Klein (S'93-M'96) received the diploma and Dr.-Ing. (Ph.D.) degrees in electrical engineering from University of Kaiserslautern, Germany, in 1991 and 1996, respectively. From 1991 to 1996, she was a member of the staff of the Research Group for RF Communications at University of Kaiserslautern. In 1996, she joined Siemens AG, Mobile Networks Division, Munich and Berlin. She was active in the standardization of third generation mobile radio in ETSI and in 3GPP, for instance leading the TDD group in RAN1 of 3GPP. She was vice president, heading a development department and a systems engineering department.

In May 2004, she joined Darmstadt University of Technology, Germany, as full professor, heading the Communications Engineering Lab. Her main research interests are in mobile radio, including multiple access, single and multi carrier transmission schemes, multi antenna systems, radio resource management, network planning and dimensioning, relaying and cooperative communication.

Dr. Klein has published over 150 refereed papers and has contributed to five books. She is inventor and co-inventor of more than 45 patents in the field of mobile radio. In 1999, she was inventor of the year of Siemens AG. Dr. Klein is a member of IEEE and of the Verband Deutscher Elektrotechniker - Informationstechnische Gesellschaft (VDE-ITG).