

# Non-asymptotic performance bounds for downlink MU-MIMO scheduling

Georgios Zois<sup>‡</sup>, Antonios Michaloliakos<sup>†</sup>, Konstantinos Psounis<sup>†</sup>, Vasilis Vassalos<sup>‡</sup>, and Ioannis Mourtos<sup>\*</sup>

<sup>‡</sup>Department of Informatics, AUEB, Greece, {georzois,vassalos}@aueb.gr

<sup>†</sup>Electrical Engineering and Computer Science Departments, USC, USA, {michalol,kpsounis}@usc.edu

<sup>\*</sup>Department of Management Science and Technology, AUEB, Greece, mourtos@aueb.gr

**Abstract**—We consider the user selection downlink MU-MIMO scheduling problem in the practical case where there are more users than transmit antennas. First, we deduce a number of structural properties for the sum data rate maximization function under the reduced-complexity suboptimal approaches of zero-forcing dirty-paper (ZF-DP) and zero-forcing beamforming (ZF-BF) precoding. Next, we take advantage of the algorithmic literature proposed in the context of combinatorial auctions when bidders have subadditive valuations and propose a novel, fast, greedy approach with very low computational complexity. Then, we establish that both the proposed greedy algorithm and a previously proposed algorithm, which iteratively augments the scheduled user set, attain a  $M$ -approximation factor for both ZF-BF and ZF-DP precodings, where  $M$  is the number of antennas. To the best of our knowledge, this is the first time that non-asymptotic performance bounds are obtained for this problem. Finally, we compare via simulations the average sum rates achieved by the greedy algorithms with the capacity of the channel and show that they all achieve the full multiplexing gain and their performance is not far from the optimal.

**Index Terms**—MU-MIMO; greedy scheduling algorithms; approximation; user selection; subadditive functions

## I. INTRODUCTION

The ever increasing demand for wireless throughput has pushed research and the industry to explore the capacity limits of the wireless channel. Although frequency reuse, modulation and coding schemes have only provided minimal gains in the recent years, advances in the physical layer have made it possible to provide significant gains by the use of MIMO and MU-MIMO techniques [1], [2] as well as their distributed implementation (see [3], [4] and references therein). In these schemes, adding multiple antennas on the transmitters' and receivers' side, for sufficiently rich scattering environment, enables multiplicative gains equal to  $\min(M, N)$ , where  $M$  are the transmitter antennas and  $N$  the receive antennas compared to the throughput of the simple point-to-point case.

In this work we consider the user selection downlink MU-MIMO scheduling problem. Our analysis is applicable for both centralized and distributed scenarios, where the transmit-side antennas are not collocated. The theoretical optimal solution for such scenarios is provided by the so called Dirty Paper

Coding (DPC) [2], [5] which exhibits however a highly impractical implementation. Therefore, more practical, yet suboptimal approaches have been proposed based on linear and non-linear beamforming schemes such as Zero-Forcing Dirty Paper (ZF-DP) [2] and Zero-Forcing Beamforming (ZF-BF) [6]. When the number of downlink streams/users are larger than the number of transmit antennas, a user/antenna selection has to take place which identifies the optimal set of streams/users to be scheduled. This problem is well known to be computationally hard, and, motivated by this, to avoid an exhaustive search over all possible subsets of users, in [7] and [8] greedy policies have been presented and shown to have good performance in simulation compared to the optimal solution. In fact, the greedy ZF-DP has been shown to asymptotically achieve the capacity limit of the channel in [7]. These pioneering papers spurred a number of results in the user selection problem [9]–[11] and more recently [12]–[14].

As a major departure from previous contributions, in our work we are not just proposing a new heuristic for selecting a subset of transmit antennas to be served that showcases a good asymptotic and/or simulated behavior. After a systematic study and analysis of the characteristics of the objective functions, and leveraging the literature of subadditive set function maximization, which has been used in the context of combinatorial auctions, we provide deterministic (non-asymptotic) theoretical bounds for the existing and proposed scheduling algorithms for every channel realization. Our contribution is threefold: a) We show that both ZF-BF and ZF-DP schemes have subadditive utility functions, and prove an approximation bound of  $M$  for both ZF-DP and ZF-BF, b) we show that the ZF-BF function is non-submodular in the general case and thus the submodular function theory cannot be used to motivate tighter approximation bounds for the greedy policy, and c) we show that the monotonicity of the ZF-DP utility function allows to introduce a new heuristic, dubbed greedy subadditive, with the same approximation bound  $M$  and a significantly reduced computational overhead.

The rest of the paper is organized as follows. In Section II we outline the related work. In Section III we formulate the optimization problem and introduce our notation. In Section IV we establish the connection with problems in the context of set function maximization and prove a number of structural properties for the sum rate capacity objective. In Section V we introduce an algorithmic framework of greedy approaches

This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund (MIS: 380232), as well as by the United States National Science Foundation grant ECCS-1444060.

and prove the formal approximation bounds for the greedy algorithms. In Section VI we present simulation results comparing the ergodic sum rates of the presented algorithms with the channel capacity. Finally, in Section VII we discuss the practical implications of our results and propose ideas for future improvements.

## II. RELATED WORK

In the context of approximation bounds for user/antenna selection for MU-MIMO systems, most works have focused on statistical analysis of the asymptotic behavior of these schemes [7], [9], [10], [13] or on simulation results that show that a large portion of the channel capacity is achieved (see [8], [12] and references therein). Contrary to these results, our approach gives approximation bounds in a deterministic and not asymptotic sense, that is, for every channel realization we guarantee the maximum distance from the optimal solution.

More related to our work are the results in [15] and [16], where the submodularity and monotonicity of the utility function is exploited to provide deterministic guarantees for the performance of the greedy approximation algorithm. Both of these works focus on the point-to-point MIMO and uplink MU-MIMO cases and the capacity achieved by an optimal but impractical precoding scheme. In contrast, we target the MU-MIMO downlink and the practical ZF-BF and ZF-DP solutions whose sum rate functions are proven to not exhibit these properties.

## III. PROBLEM FORMULATION

We consider a downlink MU-MIMO system consisting of  $M$  antennas that are deployed at the transmitter and are going to serve  $U$  single-antenna receivers (users). We define as  $\mathcal{U}$  the set of all users, where  $|\mathcal{U}| = U$ , and we assume that the total transmitted power is upper bounded by  $P_{\text{sum}}$ . The channel between the transmit antennas and user  $u \in \mathcal{U}$  is assumed to be flat-fading and is modeled by the vector  $\mathbf{h}_u$  whose elements,  $h_{u,m}$ , are the channel coefficients between antennas  $m$  ( $m = 1 \dots M$ ) and user  $u$ . By collecting all the user channel vectors in a matrix  $\mathbf{H} \in \mathbb{C}^{U \times M}$ , we can write the downlink received signal vector,  $\mathbf{y}$  as:  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$ , where  $\mathbf{x}$  is the transmitted signal vector and  $\mathbf{z}$  the noise vector assumed here to have i.i.d. circularly symmetric complex Gaussian zero mean, unit variance entries,  $z_u \sim \mathcal{CN}(0, 1)$ .

We focus on the typical case where there are more users available than antennas, i.e.,  $U > M$ , where *user selection* should be applied in order to approach the system's sum rate capacity. By user selection, at each scheduling slot, the goal is to select a subset of users that maximize a target utility function under a transmit power constraint. In a MU-MIMO system, the users are spatially separated and are not generally able to communicate with each other, so it is not possible to jointly decode all the users' observations. In this case, the successful use of the channel requires careful scheduling and *precoding* of the independent signals at the transmitter side in order to invert the channel matrix and control the multiuser interference.

We consider two well-known linear precoders, Zero-Forcing Dirty-Paper (ZF-DP) [2] and Zero-Forcing Beamforming (ZF-BF) [6]. As already mentioned, these two procedures lead to reduced-complexity, suboptimal yet high-performing solutions to the sum rate capacity maximization problem.

The use of ZF-DP precoding [2] is based on a QR-type decomposition of the channel matrix  $H = GQ$  obtained by applying Gram-Schmidt orthogonalization procedure to the rows of  $H$ ;  $G$  is a lower triangular matrix, and  $Q$  has orthonormal rows. By setting  $Q^*$  (the conjugate-transpose of  $Q$ ) as precoding matrix, it generates a set of  $M$  interference channels of the form  $x_u = g_{u,u}t_u + \sum_{j < u} g_{u,j}t_j + z_u$ ,  $u = 1, \dots, M$ , where  $t_u$  is the transmitted signal,  $\sum_{j < u} g_{u,j}t_j$  is the interference signal, and  $z_u$  the corresponding noise in channel  $u$ .

By applying DP coding on the transmitted signals, for each channel  $m = 1, 2, \dots, M$ , the interference caused by users  $j < m$  is nulled. Moreover, by the choice of  $Q^*$ , the interference caused by users  $j > m$  is also nulled, and thus all interferences are forced to zero. By using Theorem 1 in [2], it holds that the maximum sum rate capacity achieved by ZF-DP precoding is obtained as the solution to the following optimization problem:

$$\begin{aligned} \text{(MP1): maximize} \quad & \sum_{\substack{u \in S \\ S \subseteq \mathcal{U}: |S| \leq M}} [\log_2(\mu(S)d_u(S))]_+ \\ \text{subject to} \quad & \sum_{u \in S} \left[ \mu(S) - \frac{1}{d_u(S)} \right]_+ = P_{\text{sum}} \quad (1) \\ & d_u(S), \mu(S) \geq 0, \quad \forall S \subseteq \mathcal{U} : |S| = M, u \in S, \quad (2) \end{aligned}$$

where  $[\log_2 \mu(S)d_u(S)]_+ = \max\{0, \log_2(\mu d_u(S))\}$ ,  $d_u(S) = |g_{u,u}|^2$  and  $\mu(S)$  is the solution to the water-filling equation (1) for a subset  $S \subseteq \mathcal{U}$ . We refer to (MP1) as the ZF-DP MU-MIMO SCHEDULING problem. On the other hand, the use of ZF-BF precoding [6] consists of inverting the channel matrix  $H$  at the transmitter so as to create orthogonal channels between the transmitter and the receivers, while the receivers do cooperate with each other. This idea, in contrast to the more complex vector coding which is needed to implement DP, enables the individual encoding of users. Since in our setting we have  $\text{rank}(H) = M$ , in order to apply ZF-BF we need to select some  $1 \leq j \leq M$  and a set of channels  $\{h^1, h^2, \dots, h^j\}$  which produce the row-reduced channel matrix  $H(S) = [(h^1)^*, (h^2)^*, \dots, (h^j)^*]^*$ , such that the sum rate capacity is maximized. Similarly to ZF-DP precoding, it can be proved [2] that the maximum sum rate capacity is attained by solving the following optimization problem:

$$\begin{aligned} \text{(MP2): maximize} \quad & \sum_{\substack{u \in S \\ S \subseteq \mathcal{U}: |S| \leq M}} [\log_2(\mu(S)c_u(S))]_+ \\ \text{subject to} \quad & \sum_{u \in S} \left[ \mu(S) - \frac{1}{c_u(S)} \right]_+ = P_{\text{sum}} \quad (3) \\ & c_u(S), \mu(S) \geq 0, \quad \forall S \subseteq \mathcal{U} : |S| \leq M, u \in S \quad (4) \end{aligned}$$

where  $c_u(S)$  can be computed from the channel submatrix  $H(S)$ , i.e.,  $c_u(S) = \{[(H(S)H(S)^*)^{-1}]_{u,u}\}^{-1}$ , and  $\mu(S)$  as before, by using the water filling equation (3). The power  $p_u(S)$  allocated to each user  $u \in S$  is then given by  $p_u(S) = \left[ \mu(S) - \frac{1}{c_u(S)} \right]_+$ . We refer to (MP2) as the ZF-BF MU-MIMO SCHEDULING problem.

For notational convenience, we denote by  $R(S)$  the sum rate capacity over all users of a chosen subset  $S$ , in both ZF-DP and ZF-BF precodings.

#### IV. ANALYZING THE OBJECTIVE FUNCTION

Based on the formulations (MP1) and (MP2) of the ZF-DP and ZF-BF MU-MIMO SCHEDULING, a feasible solution to this problem corresponds to a subset of users  $S$ , of cardinality at most  $|S| \leq M$ , with total power allocated to them equal to  $P_{\text{sum}}$ , and objective value  $R(S)$ . It is important to note that during the selection process the power allocation is computed, for each candidate subset of users, in polynomial time using the water-filling equation. Thus, we may assume without loss of generality that the power allocation process is given as a “black box”, for a candidate subset of users  $S$ . Then, our problem can be formulated as  $\max\{R(S) \mid S \subseteq \mathcal{U} : |S| \leq M\}$ . Clearly, according to this formulation ZF-DP and ZF-BF MU-MIMO SCHEDULING can be described as a *set function maximization* problem. To proceed with the analysis of the sum rate capacity maximization objective, the following preliminaries are required.

##### A. Definitions and Preliminaries

Let  $\mathcal{U}$  be a universe of  $U = |\mathcal{U}|$  elements and let  $f : 2^{\mathcal{U}} \rightarrow \mathbb{R}^+$  be a set function. We are interested to the standard SET FUNCTION MAXIMIZATION problem: We are given a set of items  $\mathcal{U}$ , and a positive integer  $M$  (cardinality constraint) and the goal is to find a subset  $S \subseteq \mathcal{U} : |S| \leq M$ , in order to maximize a set function  $f : 2^{\mathcal{U}} \rightarrow \mathbb{R}$ . More compactly, this combinatorial problem can be formulated as  $\max\{f(S) \mid S \subseteq \mathcal{U} : |S| \leq M\}$  and is clearly related to the ZF-DP and ZF-BF MU-MIMO SCHEDULING problem.

**Definition 1.** A set function  $f$  is said to be submodular if  $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$ , for every  $A, B \in \mathcal{U}$ .

**Definition 2.** A set function  $f$  is said to be monotone if  $f(A) \leq f(B)$ , for every  $A \subseteq B \subseteq \mathcal{U}$ .

An alternative and more practical definition of submodular functions is based on a natural diminishing returns property.

**Definition 3.** A set function  $f$  is said to be submodular if  $f(A \cup u) - f(A) \geq f(B \cup u) - f(B)$ , for every  $A \subseteq B \subseteq \mathcal{U}$  and  $u \in \mathcal{U} \setminus B$ .

It is well known [17], [18] that, when function  $f$  is monotone and submodular, the following sequential greedy algorithm obtains a tight approximation of  $(1 - 1/e)$  for the SET FUNCTION MAXIMIZATION problem.

Algorithm G-SUBMODULAR. Start with an empty set  $S$  and grow  $S$ , with respect to the cardinality constraint  $|S| \leq M$ ,

by iteratively adding the item  $u \in \mathcal{U}$  such that  $u = \arg \max_{u' \in \mathcal{U} \setminus (S \cup \{u'\})} f(S \cup \{u'\}) - f(S)$ .

The following class of *complement-free* set functions is a substantial generalization of submodular functions.

**Definition 4.** A set function  $f$  is said to be subadditive if  $f(A) + f(B) \geq f(A \cup B)$ , for every  $A, B \in \mathcal{U}$ .

Contrary to submodular maximization, for subadditive maximization only a few results are known. These results have been proposed in the context of combinatorial auctions (see e.g., [19]–[22]) where we are given a set of  $U$  (heterogenous indivisible) items that are sold to  $n$  competing buyers while the buyers value subsets of items, rather than individual single items, according to their valuation function. The main objective is to find an *allocation* of the items to buyers in order to maximize the social welfare [19], [23].

Note that a naive representation of a valuation function would require us to specify  $2^U$  values, one for each possible subset of items. However, in terms of algorithm design, the goal is to run in time polynomial in the input size (e.g., in  $U$  and  $n$ ). Therefore, the combinatorial auctions literature usually assumes that the valuation is represented by an oracle that can answer certain types of queries. From a computer science point of view, a very natural type of queries are called *value queries* i.e., the query specifies a subset of items and receives the value as the reply.

##### B. Properties of the objective function

Now, we are ready to deduce a number of structural properties for the sum rate capacity function of the ZF-DP and ZF-BF MU-MIMO SCHEDULING problem. The following key-proposition is based on an expression of  $d_u(S)$  and  $c_u(S)$  in [2, Section 3] and leads to some very useful properties of the objective function. Its proof is deferred to the Appendix.

**Proposition 1.** Consider the sum rate capacities of ZF-DP and ZF-BF MU-MIMO SCHEDULING, for any two  $S \subseteq S' \subseteq \mathcal{U} : |S|, |S'| \leq M$ . Then, for each user  $u \in S$ , it holds that

- i)  $d_u(S) = d_u(S')$ ,
- ii)  $c_u(S) \geq c_u(S')$ .

By using Proposition 1 together with some reasonable assumption on the power budget, the following lemma holds (see the Appendix for the proof).

**Lemma 1.** Consider the sum rate capacity  $R(S)$  in one of the ZF-BF or ZF-DP MU-MIMO SCHEDULING problems. Then, function  $R(S)$  is subadditive.

Since subadditivity does not negate the submodularity of a function, the following proposition is crucial in terms of algorithm analysis. Its proof can be found in the Appendix and it is based on the construction of an appropriate instance.

**Proposition 2.** There are instances of the ZF-BF MU-MIMO SCHEDULING problem, for which the sum rate capacity function  $R(S)$  is non-submodular.

It seems natural for some instances of ZF-BF MU-MIMO SCHEDULING that the objective function might be non-

monotone, i.e., when adding a new user  $u$  in a subset  $S$  it might be the case that  $p_u(S')$  ( $S' = S \cup \{u\}$ ) is small enough (or  $p_{u'}(S')$ , for some user  $u' \in S$ ) such that, by applying Proposition 1(ii), we yield that  $R(S') \leq R(S)$ . Contrary to ZF-BF MU-MIMO SCHEDULING, the sum rate capacity of ZF-DP MU-MIMO SCHEDULING is proved to be monotone non-decreasing. The next lemma, which is proved in the Appendix, outlines these properties of ZF-DP and ZF-BF MU-MIMO SCHEDULING.

**Lemma 2.** (i) *The sum rate capacity of ZF-DP MU-MIMO SCHEDULING is monotone non-decreasing.*  
(ii) *The sum rate capacity of ZF-BF MU-MIMO SCHEDULING is non-monotone.*

## V. AN ALGORITHMIC FRAMEWORK

Next, by taking advantage of the algorithmic framework that has been proposed for submodular and subadditive maximization, we present fast greedy algorithms with proven performance guarantee for the ZF-DP and ZF-BF MU-MIMO SCHEDULING problem.

Recall that our objective (for both ZF-DP and ZF-BF MU-MIMO SCHEDULING) falls into the broader class of subadditive functions, while also, for some instances, the submodularity condition of ZF-BF MU-MIMO SCHEDULING does not apply. Therefore, it seems natural to design algorithms that capture the nature of the objective function, while achieving the best possible performance guarantees.

### A. Two greedy algorithmic approaches

Before presenting our results, it is reasonable to assume that  $R(S)$  is normalized ( $R(\emptyset) = 0$ ). Moreover, according to Lemma 2(i) and Lemma 2(ii) we know that  $R(S)$  is monotone non-decreasing for ZF-DP MU-MIMO SCHEDULING and non-monotone for ZF-BF MU-MIMO SCHEDULING respectively.

The first approach is a novel algorithm which is inspired by the setting proposed in [19], [23] for the problem of allocating items to buyers in combinatorial auctions in the presence of value oracles; note that the use of value oracles for our problem is quite natural, since, for each chosen subset  $S$ , the value of the sum rate capacity  $R(S)$  can be computed analytically in polynomial time. The algorithm, called G-SUBADDITIVE, is executed as follows:

---

#### G-SUBADDITIVE

- 1: Partition  $\mathcal{U}$ , uniformly at random, into  $k = \lceil \frac{U}{M} \rceil$  disjoint subsets. Let  $\{S_1, S_2, \dots, S_k\}$  be these subsets.
  - 2: **for** each subset  $S_j$ ,  $j = 1, 2, \dots, k$  **do**
  - 3:   Let the power budget be equal to  $P_{\text{sum}}$  and compute the value of  $R(S_j)$ .
  - 4: Compute  $R(S) = \max_{j=1,2,\dots,k} R(S_j)$ .
- 

By combining Lemma 2(i) with Lemma 1 we are able to prove that:

**Theorem 1.** *Algorithm G-SUBADDITIVE is a  $M$ -approximation for the ZF-DP MU-MIMO SCHEDULING problem.*

*Proof.* Consider an optimal solution  $(S^*, R(S^*))$  to the ZF-DP MU-MIMO SCHEDULING problem and let  $(S^{ALG}, R(S^{ALG}))$  be a solution of Algorithm G-SUBADDITIVE. Note also that by Lemma 2(i), the sum rate capacity function is monotone non-decreasing, so we can assume that  $|S^*| = M$ .

By Lemma 1 the sum rate capacity is a subadditive function, thus it must hold that

$$R(S^*) \leq \sum_{u \in S^*} R(\{u\}).$$

Then, there must exist a user  $v \in S^*$  such that  $R(\{v\}) \geq \frac{1}{M}R(S^*)$ .

Let  $S_j$  be the subset produced by Algorithm G-SUBADDITIVE such that  $v \in S_j$ . By the monotonicity of the sum rate capacity function it holds that

$$R(S_j) \geq R(\{v\}) \geq \frac{1}{M}R(S^*). \quad (5)$$

Now, since Algorithm G-SUBADDITIVE chooses the subset of the partition with the maximum sum rate we have that

$$R(S^{ALG}) \geq R(S_j). \quad (6)$$

Combining the inequality (6) with (5) we yield that  $R(S^{ALG}) \geq \frac{1}{M}R(S^*)$ , and the theorem follows.  $\square$

Unfortunately, Theorem 1 does not apply in the case of ZF-BF MU-MIMO SCHEDULING, since the sum rate capacity function is non-monotone. A well-studied problem that is closely-related to ZF-DP and ZF-BF MU-MIMO SCHEDULING is the BUDGET-LIMITED PROCUREMENT AUCTION problem [21], where we are given a set of  $U$  items and a *single buyer*, each item  $i$  is associated with a cost  $c_i \in \mathbb{R}^+$ , and the buyer has a budget  $B \in \mathbb{R}^+$  and a valuation function over all subsets of resources. The goal for the buyer is to maximize his value with respect to the budget  $B$ . More precisely, our problem can be expressed as a special case of the latter problem, where  $c_i = 1$ ,  $\forall i = 1, 2, \dots, n$  (thus, budget  $B$  becomes a cardinality constraint equal to  $M$ ) for subadditive valuations in the presence of value oracles. Therefore, in the case of ZF-BF MU-MIMO SCHEDULING, where the sum rate capacity is also subadditive and is not submodular, a negative result (in terms of communication complexity) proposed in Theorem 5.4 of [21, Section 5] for the above special case of BUDGET-LIMITED PROCUREMENT AUCTION, can be directly applied.

**Theorem 2.** *Any algorithm that attains an approximation factor better than  $U^{\frac{1}{2}-\epsilon}$ , for constant  $\epsilon > 0$ , for the ZF-BF MU-MIMO SCHEDULING problem, requires exponentially many value queries, even when the valuation function is fractionally subadditive.*

A valuation function  $v$  is said to be fractionally subadditive if there is a set of additive (linear) valuations  $\{a_1, a_2, \dots, a_l\}$

such that for every  $S \subseteq \mathcal{U}$ ,  $v(S) = \max_{r \in [l]} a_r(S)$ . Note that the class of fractionally-subadditive valuation functions is known to be strictly contained in the class of subadditive valuations and to strictly contain all submodular valuations [19].

A different algorithmic approach is a greedy algorithm called G-ZFS, that has been proposed by Dimic and Sidiropoulos [7, Section III-B] for the ZF-BF MU-MIMO SCHEDULING problem. The idea of G-ZFS is along the line of the earlier described G-SUBMODULAR algorithm [17] for maximizing a submodular function under a cardinality constraint, while also taking into account the non-monotonicity of the objective.

Specifically, the algorithm G-ZFS proceeds as follows: Iteratively add a user  $u$  that is not yet been included in (an initially empty) set  $S$ . In each iteration, choose to add in  $S$  the user  $u$  with the maximum marginal contribution i.e., for which the sum rate capacity  $R(S \cup \{u\})$  is maximized. Break when there is a user  $u$  such that, the sum rate capacity  $R(S \cup \{u\})$ , where  $|S| < M$ , is less or equal than the sum rate capacity  $R(S)$ .

Note that in each iteration, for computing the sum rate capacity of a candidate user set  $S$ , we solve the water-filling equation for allocating power  $P_{\text{sum}}$  in the users of  $S$ .

A crucial fact during the execution of G-ZFS is the use of ‘break’ command. Actually, this is due to the fact that the sum capacity rate of ZF-BF MU-MIMO SCHEDULING is non-monotone function and thus, an optimal subset might contain strictly less users than the cardinality constraint  $M$ . More interestingly, as it is observed in [7] the optimum number of active users decreases as the power budget ( $P_{\text{sum}}$ ) decreases.

For the performance guarantee of G-ZFS, we have:

**Theorem 3.** *Algorithm G-ZFS is a  $M$ -approximation for the ZF-BF MU-MIMO SCHEDULING problem.*

*Proof.* Consider an optimal solution  $(S^*, R(S^*))$  to the ZF-BF MU-MIMO SCHEDULING problem and let  $(S^{ALG}, R(S^{ALG}))$  be a solution of Algorithm G-ZFS. By subadditivity (Lemma 1) of the sum rate capacity function we have

$$R(S^*) \leq \sum_{u \in S^*} R(\{u\}),$$

and thus there must be a user  $v \in S^*$  such that

$$R(\{v\}) \geq \frac{1}{M} R(S^*). \quad (7)$$

Now, consider a user - let  $u_{\text{max}}$  - which attains the maximum sum rate capacity (when all the power budget is allocated to it). Since  $v$  is chosen arbitrarily, it must hold that

$$R(\{u_{\text{max}}\}) \geq R(\{v\}). \quad (8)$$

Since Algorithm G-ZFS chooses first (to include in  $S^{ALG}$ ) the user  $u_{\text{max}}$ , we have that

$$R(S^{ALG}) \geq R(\{u_{\text{max}}\}). \quad (9)$$

Combining the inequality (9) with (8) and (7) we yield that  $R(S^{ALG}) \geq \frac{1}{M} R(S^*)$ , and the theorem follows.  $\square$

A greedy approach for the ZF-DP MU-MIMO SCHEDULING problem has been previously proposed by Tu and

Blum [8]. It is executed similarly to G-ZFS, however, instead of adding the user that maximizes the sum rate capacity, in each iteration, it evaluates the user with the highest 2-norm projection. For consistency, here we present and analyze Algorithm G-ZF-DP which is executed in the same way as G-ZFS, while using the ZF-DP nulling scheme which has been described in Section III. Moreover, due to Lemma 2(i), the ‘break’ step is removed, and the algorithm runs until exactly  $M$  users have been selected. Similarly to the proof of Theorem 3, we can show the following:

**Theorem 4.** *Algorithm G-ZF-DP is a  $M$ -approximation for the ZF-DP MU-MIMO SCHEDULING problem.*

**Remark:** A note on *time complexity* is in order. For algorithms G-ZFS and G-ZF-DP the overall complexity is  $O(M^3U)$  (see in [7, Section III-C]). For Algorithm G-SUBADDITIVE, the time for computing the sum rate capacity of a subset  $S_j$  is  $O(M^2)$ , while there are  $\lceil \frac{U}{M} \rceil$  subsets to be checked. The overall running time of Algorithm G-SUBADDITIVE is  $O(MU)$ . Obviously, Algorithm G-SUBADDITIVE under both ZF-DP and ZF-BF schemes, is executed significantly faster than G-ZFS and G-ZF-DP respectively.

## VI. SIMULATION RESULTS

In this section we present simulation results about the performance of the various described schemes. As a performance metric we compute the ergodic sum rate through Monte Carlo simulation over 100s of realizations of the random channel matrices that follow the Rayleigh fading model with  $h_{u,m} \sim \mathcal{CN}(0, 1)$ .

The schemes under comparison are the G-ZFS algorithm, the G-ZF-DP algorithm and their subadditive counterparts. For comparison purposes we also compute the channel capacity upper bound achieved by DPC. In Figure 1 we see that all schemes achieve the so called *degrees of freedom* gain, that is the slope of their ergodic sum rate against the SNR is the same as that of the channel capacity bound. Moreover, all four greedy policies respect the theoretic bounds we provided and in fact achieve up to 90% of the channel capacity in practice for the ZF-DP based schemes. Even the G-SUBADDITIVE applied to ZF-BF, although no theoretic guarantee could be provided (the function is non-monotone, see Lemma 2(ii)), still achieves more than 70% of the channel capacity in most cases with a complexity significantly reduced compared to G-ZFS.

It is interesting to note that, although both the greedy and greedy subadditive algorithms give the same theoretical guarantees, their performance is not guaranteed to be the same. In fact, we see that the greedy user selection will outperform the subadditive algorithms since a significantly greater amount of work is put into carefully selecting the users. We hasten to point out however, that in terms of time complexity, which is important in practical user selection scenarios with fast varying channels, the subadditive algorithms can be up to  $O(M^2)$  times faster. In fact, this gives us the ability, to tradeoff part of the greedy subadditive algorithms time efficiency for higher performance by creating multiple random user set partitions

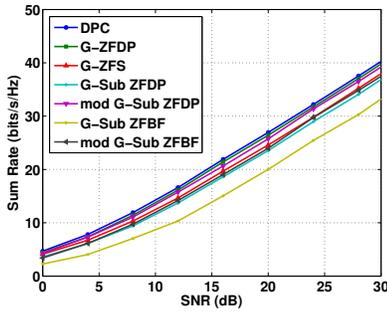


Fig. 1: Sum Rate vs SNR for  $U = 12$  users,  $M = 4$  antennas.

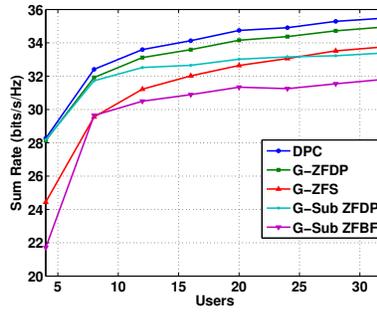


Fig. 2: Sum Rate vs Number of Users for  $M = 4$  antennas and 25dB SNR.

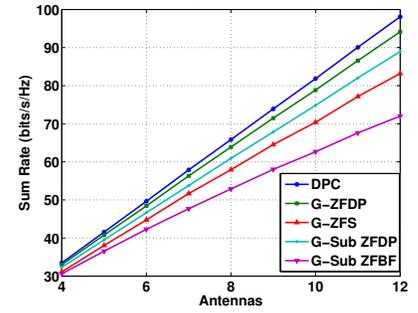


Fig. 3: Sum Rate vs Number of Antennas for  $U = 3xM$  users and 25dB SNR.

and choosing the best. These schemes are depicted in Figure 1 as modified (mod) G-SUBADDITIVE algorithms for a constant number of random partitionings of the user set, making sure that time complexity remains significantly below the G-ZFS and G-ZF-DP algorithms.

Finally, a note on the scaling of the algorithms is in order. As can be seen in Figures 2 and 3, the proposed G-SUBADDITIVE schemes exhibit the same scaling properties as their greedy versions. Specifically, for a fixed number of antennas, as the number of users increases, higher multi-user diversity allows for an increase in the sum rate of all algorithms [7], [9]. Moreover, for a fixed user/antenna ratio, we see in Figure 3, that as the number of antennas increases, the G-SUBADDITIVE low complexity scheme still manages to unlock most of the multiplexing gains of the system with a loss of 30% at worst case which is still well above the guaranteed  $1/M$  approximation bound.

## VII. DISCUSSION AND FUTURE WORK

A crucial open question is whether ZF-DP MU-MIMO SCHEDULING is submodular or not. Note that in the first case, since the sum rate capacity is monotone non-decreasing, Algorithm G-ZF-DP would obtain a tight approximation of  $(1 - 1/e)$ , following the same analysis as in Algorithm G-SUBMODULAR [17], [18]. However, if the problem is not submodular then we would still like to tighten our approximation ratios either by improving our analysis or by proposing other efficient algorithms.

Similarly, for the ZF-BF MU-MIMO SCHEDULING problem we would also like to improve the performance guarantee of G-ZFS despite proving, by a counterexample, that this is not a submodular problem. Note that although we have found a few instances for which the ZF-BF MU-MIMO SCHEDULING problem is not submodular, the performance of the G-ZFS algorithm under high enough multi-user diversity leads us to the conclusion that for the subset of users on which the algorithm operates the channel is such that conditions for the submodularity of the sum rate function are met. In our future work we plan to explore this property to get tighter approximation bounds. Finally, an interesting direction for overcoming the non-monotonicity and non-approximability of the sum rate capacity in ZF-BF MU-MIMO SCHEDULING

is to investigate the use of demand oracles, instead of value oracles, in combinatorial auctions. To this direction, we could build on ideas of Badanidiyuru et al. [22], which design constant approximations for the (monotone and non-monotone) subadditive maximization problem in the presence of demand oracles.

## REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [2] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691 – 1706, Jul. 2003.
- [3] H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire, "Achieving high data rates in a distributed MIMO system," in *Proceedings of the 18th annual international conference on Mobile computing and networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 41–52.
- [4] H. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire, "Airsync: Enabling distributed multiuser mimo with full spatial multiplexing," *Networking, IEEE/ACM Transactions on*, vol. 21, no. 6, pp. 1681–1695, Dec 2013.
- [5] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [6] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [7] G. Dimic and N. 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, Oct. 2005.
- [8] Z. Tu and R. Blum, "Multiuser diversity for a dirty paper approach," *Communications Letters, IEEE*, vol. 7, no. 8, pp. 370–372, Aug 2003.
- [9] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528 – 541, Mar. 2006.
- [10] J. Wang, D. J. Love, and M. D. Zoltowski, "User selection with zero-forcing beamforming achieves the asymptotically optimal sum rate," *IEEE Transactions on Signal Processing*, vol. 56, no. 8-1, pp. 3713–3726, 2008.
- [11] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1468 – 1477, Sep. 2007.
- [12] W. Shen, K. C. Lin, M. Chen, and K. Tan, "SIEVE: scalable user grouping for large MU-MIMO systems," in *2015 IEEE Conference on Computer Communications, INFOCOM 2015, Kowloon, Hong Kong, April 26 - May 1, 2015*, 2015, pp. 1975–1983. [Online]. Available: <http://dx.doi.org/10.1109/INFOCOM.2015.7218581>
- [13] S. Ozyurt and M. Torlak, "Unified performance analysis of orthogonal transmit beamforming methods with user selection," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1026–1037, 2013.

- [14] A. Michaloliakos, R. Rogalin, H. V. Balan, K. Psounis, and G. Caire, "Efficient MAC for distributed multiuser MIMO systems," in *10th Annual Conference on Wireless On-demand Network Systems and Services, WONS 2013, Banff, AB, Canada, March 18-20, 2013*, 2013, pp. 52–59.
- [15] R. Vaze and H. Ganapathy, "Sub-modularity and antenna selection in mimo systems," *Communications Letters, IEEE*, vol. 16, no. 9, pp. 1446–1449, September 2012.
- [16] N. Prasad, H. Zhang, H. Zhu, and S. Rangarajan, "Multi-user mimo scheduling in the fourth generation cellular uplink," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 9, pp. 4272–4285, September 2013.
- [17] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Mathematics of Operations Research*, vol. 3, no. 3, pp. 177–188, 1978.
- [18] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.
- [19] S. Dobzinski, N. Nisan, and M. Schapira, "Approximation algorithms for combinatorial auctions with complement-free bidders," *Math. Oper. Res.*, vol. 35, no. 1, pp. 1–13, 2010.
- [20] V. S. Mirrokni, M. Schapira, and J. Vondrák, "Tight information-theoretic lower bounds for welfare maximization in combinatorial auctions," in *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008), Chicago, IL, USA, June 8-12, 2008*, 2008, pp. 70–77.
- [21] Y. Singer, "Budget feasible mechanisms," in *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, 2010, pp. 765–774.
- [22] A. Badanidiyuru, S. Dobzinski, and S. Oren, "Optimization with demand oracles," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, ser. EC '12. New York, NY, USA: ACM, 2012, pp. 110–127.
- [23] S. Dobzinski and N. Nisan, "Limitations of vcg-based mechanisms," *Combinatorica*, vol. 31, no. 4, pp. 379–396, 2011.
- [24] A. W. Marshall, I. Olkin, and B. Arnold, *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

## Appendix

### Proof of Proposition 1

For the shake of simplicity, we consider the case of subsets  $S \subseteq S'$  that differ only in a single user, i.e.,  $S = \{u_1, u_2, \dots, u_k\}$  and  $S' = \{u_1, u_2, \dots, u_k, u_{k+1}\}$ , where  $1 \leq k < M$ ; the generalization can be easily deduced by induction on the cardinality of  $S$ .

For statement (i) we have that [2]

$$d_{u_i}(S) = h^i P(S_i)^\perp (h^i)^* = |h^i P(S_i)^\perp|^2, \quad (10)$$

where  $h^i, i = 1, \dots, k$  are the ordered rows of  $H(S)$ ,  $P(S_{i-1})^\perp$  is the orthogonal projector on the orthogonal complement of  $\mathcal{F}_{S_{i-1}} = \text{span}\{h^j : u_j \in S_{i-1}\}$ , for  $i = 2, \dots, k$  and the last equality holds due to the fact that  $P(S_i)^\perp$  is idempotent.

Similarly, for the subset  $S'$ , equation (10) gives that  $d_{u_i}(S') = |h^i P(S_{i-1})^\perp|^2$ , where  $P(S_{i-1})^\perp$  is again the orthogonal projector on the orthogonal complement of  $\mathcal{F}_{S_{i-1}}$ , and the proof follows.

For statement (ii) we have that [2]

$$c_{u_i}(S) = |h^i P(S_{-i})^\perp|^2, \quad (11)$$

where  $S_{-i} = S \setminus \{u_i\}$  and  $P(S_{-i})^\perp$  is the orthogonal projector on the orthogonal complement of  $\mathcal{F}_{S_{-i}}$ .

When subset  $S'$  is considered, by (11) we yield,  $c_{u_i}(S') = |h^i P(S'_{-i})^\perp|^2$ , where  $P(S'_{-i})^\perp$  is the orthogonal

complement of  $\mathcal{F}_{S'_{-i}}$ . By definition of  $\mathcal{F}_{S_{-i}}$  we have that  $\mathcal{F}_{S_{-i}} \subset \mathcal{F}_{S'_{-i}}$  and thus  $c_{u_i}(S) \geq c_{u_i}(S')$ .

### Proof of Lemma 1

We will prove the lemma for the ZF-BF MU-MIMO SCHEDULING problem. A similar proof can be also applied for ZF-DP MU-MIMO SCHEDULING.

In terms of our analysis we will need the following notation and result from the theory of majorization [24].

**Definition 5.** Let  $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$  and  $y = (y_1, y_2, \dots, y_m) \in \mathbb{R}^m$  and  $x_i \geq x_{i-1}, y_i \geq y_{i-1}$ , for  $1 < i \leq m$ . Then,  $x$  is said to be weakly majorized by  $y$  from below if  $x \prec_w y$ , i.e.,

$$x \prec_w y \quad \text{if} \quad \sum_{i=1}^k x_i \leq \sum_{i=1}^k y_i \quad \forall k = 1, 2, \dots, m.$$

The following theorem is equivalent to Theorem C.1.b in [24, Chapter 3] and is a key ingredient of our proof.

**Theorem 5.** Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a function which is convex and decreasing. Let also  $x, y \in \mathbb{R}^m$ , with their components in decreasing order. Then,  $\sum_{i=1}^m g(\cdot)$  is schur-convex and decreasing on  $\mathbb{R}^m$ . Consequently,  $x \prec_w y$  implies that

$$\sum_{i=1}^m g(x_i) \leq \sum_{i=1}^m g(y_i)$$

Recall that, by (MP2) formulation, the sum rate capacity of any subset  $T$  is  $R(T) = \sum_{u \in T} [\log_2 \mu(T) c_u(T)]_+$  where  $\mu(T)$  is the solution to the water-filling equation  $\sum_{u \in T} \left[ \mu(T) - \frac{1}{c_u(T)} \right]_+ = P_{\text{sum}}$ . According to our assumption for high channel signal-to-noise ratio we have that the power of each user in  $T$  will be strictly positive. Then, by using equation  $p_u(T) = c_u(T) \left[ \mu(T) - \frac{1}{c_u(T)} \right]_+$  from Section III, it must be the case where  $\mu(T) \cdot c_u(T) > 1$ , for each  $u \in T$ , and thus the sum rate capacity is written as  $R(T) = \sum_{u \in T} \log_2 \mu(T) c_u(T)$ . Moreover, the solution to the water-filling equation is equal to  $\mu(T) = \frac{1}{|T|} (P_{\text{sum}} + \sum_{u \in T} \frac{1}{c_u(T)})$ .

Now, consider two subsets  $S_1, S_2 \subseteq \mathcal{U}$  and let  $S = S_1 \cup S_2$ . Let also  $R(S_1), R(S_2)$  and  $R(S)$  be the sum rate capacity and let  $\mu(S_1), \mu(S_2), \mu(S)$  be the solution to the water-filling equation for each of  $S_1, S_2$  and  $S$  respectively. Then, by using Definition 4 we must prove that

$$R(S_1) + R(S_2) \geq R(S).$$

Since the sum of logarithms of the same base is equal to the logarithm of the product and the logarithm function is monotone non-decreasing, we have to prove that

$$\left( \prod_{u \in S_1} \mu(S_1) c_u(S_1) \right) \cdot \left( \prod_{u \in S_2} \mu(S_2) c_u(S_2) \right) \geq \left( \prod_{u \in S} \mu(S) c_u(S) \right),$$

which is equivalent (since all ingredients are positive) to

$$\mu(S_1)^{|S_1|} \mu(S_2)^{|S_2|} \frac{\prod_{u \in S_1} c_u(S_1) \prod_{u \in S_2} c_u(S_2)}{\prod_{u \in S} c_u(S)} \geq \mu(S)^{|S|}.$$

Now consider the index  $x = \arg \min_{i=1,2} \{\mu(S_i)\}$  and let  $S'_x = \{u \in S_x | u \in S_1 \cap S_2\}$ . Then, by applying Proposition 1(ii), we should prove that

$$\mu(S_1)^{|S_1|} \mu(S_2)^{|S_2|} \prod_{u \in S_1 \cap S_2} c_u(S'_x) \geq \mu(S)^{|S|}$$

or equivalently,

$$\mu(S_1)^{|S_1|} \mu(S_2)^{|S_2|} \geq \frac{1}{\prod_{u \in S_1 \cap S_2} c_u(S'_x)} \mu(S)^{|S|}. \quad (12)$$

The following fact is an extended form of the classic inequality of arithmetic and geometric means, and is very useful for our analysis.

**Fact 1.** *It must hold that*

$$\frac{1}{\prod_{u \in S_1 \cap S_2} c_u(S'_x)} \leq \left( \frac{P + \sum_{u \in S_1 \cap S_2} \frac{1}{c_u(S'_x)}}{|S'_x|} \right)^{|S'_x|}. \quad (13)$$

Note that  $\frac{1}{|S'_x|} (P + \sum_{u \in S_1 \cap S_2} \frac{1}{c_u(S'_x)}) = \mu(S'_x)$ . Thus, combining Fact 1 and (12) it suffices to prove that  $\mu(S_2)^{|S_2|} \mu(S_1)^{|S_1|} \geq \mu(S)^{|S|} \mu(S'_x)^{|S'_x|}$ .

Since  $|S'_x| = |S_1 \cap S_2|$  and  $|S| = |S_1 \cup S_2|$ , it must hold that  $|S_1| + |S_2| = |S| + |S'_x|$ . Moreover, for convenience, let us denote by

$$\tilde{\mu}(T) = \left( P + \sum_{u \in T} \frac{1}{c_u(T)} \right),$$

the part of  $\mu(T)$  without the term  $1/|T|$ , for any  $T \subseteq \mathcal{U}$ . Then, we should prove that

$$\frac{|S|^{|S|} |S_1 \cap S_2|^{|S_1 \cap S_2|}}{|S_1|^{|S_1|} |S_2|^{|S_2|}} \tilde{\mu}(S_2)^{|S_2|} \tilde{\mu}(S_1)^{|S_1|} \geq \tilde{\mu}(S)^{|S|} \tilde{\mu}(S'_x)^{|S'_x|}.$$

Now, it is not difficult to show that the term  $\frac{|S|^{|S|} |S_1 \cap S_2|^{|S_1 \cap S_2|}}{|S_1|^{|S_1|} |S_2|^{|S_2|}}$  is lower bounded by one, and thus it suffices to prove that

$$\tilde{\mu}(S_2)^{|S_2|} \tilde{\mu}(S_1)^{|S_1|} \geq \tilde{\mu}(S)^{|S|} \tilde{\mu}(S'_x)^{|S'_x|}. \quad (14)$$

Using Proposition 1 ii), it must hold that  $\tilde{\mu}(\cdot)$  is a monotone non-decreasing set function, i.e.,

$$\tilde{\mu}(T') \leq \tilde{\mu}(T) \quad \forall T' \subseteq T \subseteq \mathcal{U}, \quad (15)$$

By assuming, without loss of generality, that  $\tilde{\mu}(S_1) \leq \tilde{\mu}(S_2)$  and by using (15), it must hold that

$$\tilde{\mu}(S'_x) \leq \tilde{\mu}(S_1) \leq \tilde{\mu}(S_2) \leq \tilde{\mu}(S). \quad (16)$$

For notational convenience, let us define the subsets

$$\tilde{S}_k = \begin{cases} S_2 & k = 1, \dots, |S_1| \\ S_1 & k = |S_1| + 1, \dots, |S_1| + |S_2| \end{cases},$$

$$\bar{S}_k = \begin{cases} S & k = 1, \dots, |S| \\ S'_x & k = |S| + 1, \dots, |S_1| + |S_2|. \end{cases}$$

The following is a key-fact and can be proved by appropriate use of Proposition 1(ii) and (16).

**Fact 2.** *For each  $k = 1, 2, \dots, |S_1| + |S_2|$ , it holds that*

$$\sum_{i=1}^k \tilde{\mu}(\tilde{S}_i) \leq \sum_{i=1}^k \tilde{\mu}(\bar{S}_i) \quad (17)$$

Now consider the convex and decreasing function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , where  $g(\cdot) = -\log(\cdot)$ , and two orders  $x = (\tilde{\mu}(S_2), \dots, \tilde{\mu}(S_2), \tilde{\mu}(S_1), \dots, \tilde{\mu}(S_1)) \in \mathbb{R}_+^{|S_1|+|S_2|}$ ,  $y = (\tilde{\mu}(S), \dots, \tilde{\mu}(S), \tilde{\mu}(S'_x), \dots, \tilde{\mu}(S'_x)) \in \mathbb{R}_+^{|S_1|+|S_2|}$ , where their components are in decreasing order; note that there are  $|S_2|$  copies of  $\tilde{\mu}(S_2)$  and  $|S_1|$  copies of  $\tilde{\mu}(S_1)$  in  $x$ , while in  $y$  there are  $|S|$  copies of  $\tilde{\mu}(S)$  and  $|S'_x|$  copies of  $\tilde{\mu}(S'_x)$  respectively. Then, by Fact 2 it must hold that  $x \prec_w y$  and by Theorem 5, the Lemma follows.

**Proof of Proposition 2**

We will prove this using a counterexample. Without loss of generality we choose a setting with  $U = 4$  users and  $M = 4$  transmit antennas and available transmit power of  $P_{\text{sum}} = 30$  dB. Assume that an instance of the channel matrix  $H$  whose elements are drawn from a  $\mathcal{CN}(0, 1)$  is the following:

$$H = \begin{bmatrix} 1 - 0.5i & -0.5 - 0.6i & 0.1 + 1.1i & 0 - 0.2i \\ -0.7 + 0.4i & 0.5 - 0.1i & -0.2 - 0.3i & 1.8 + 0.9i \\ -0.6 + 0.6i & 0.3 + 0.1i & 0.2 - 0.2i & -0.2 - 1.5i \\ -0.7 + 0.5i & -1.4 + 0.4i & 0.4 + 0.5i & -0.7 - 0.6i \end{bmatrix}$$

Computing the optimal sum rates for the sets of users  $S_1 = \{1, 2\}$ ,  $S_2 = \{1, 2, 3\}$ ,  $S_3 = \{1, 2, 4\}$  and  $S_4 = \{1, 2, 3, 4\}$  by solving the water-filling power allocation problem (3) we find  $R(S_1) = 14.57$ ,  $R(S_2) = 12.46$ ,  $R(S_3) = 19.07$  and  $R(S_4) = 17.02$ . Since  $S_1 \subset S_3$  and  $R(S_1 \cup \{3\}) - R(S_1) < R(S_3 \cup \{3\}) - R(S_3)$ , the function is not submodular due to Definition 3.

**Proof of Lemma 2**

(i). Consider the sum rate capacities of ZF-DP MU-MIMO SCHEDULING,  $R(S_1)$ ,  $R(S_2)$ , computed for two subsets  $S_1 \subseteq S_2$ . As logarithm function is monotone non-decreasing, it suffices to prove that

$$\mu(S_1)^{|S_1|} \prod_{u \in S_1} d_u(S_1) \leq \mu(S_2)^{|S_2|} \prod_{u \in S_2} d_u(S_2)$$

which, by using Proposition 1(i), is equivalent to

$$\mu(S_1)^{|S_1|} \leq \mu(S_2)^{|S_2|} \prod_{u \in S_2 \setminus S_1} d_u(S_2). \quad (18)$$

Now, by using the classic inequality of arithmetic and geometric means, it must hold that

$$\mu(S_1)^{|S_1|} \mu(S_2 \setminus S_1)^{|S_2| - |S_1|} \leq \mu(S_2)^{|S_2|} \quad (19)$$

By substituting (19) in (18) it suffices to show that  $\mu(S_2 \setminus S_1)^{|S_2| - |S_1|} \prod_{u \in S_2 \setminus S_1} d_u(S_2) \geq 1$ , which is true as the values of  $\mu(\cdot)$  and  $d_u(\cdot)$  are positive.

(ii). This is the case for the instance used in Proposition 2.