

On User Scheduling for Maximum Throughput in K -user MISO Broadcast Channels

Martina Cardone[†], Daniela Tuninetti*, Raymond Knopp[†]

[†]Eurecom, Biot, 06410, France, Email: {cardone, knopp}@eurecom.fr

* University of Illinois at Chicago, Chicago, IL 60607, USA, Email: danielat@uic.edu

* Laboratoire de Signaux et Systèmes (L2S, UMR8506) CNRS-CentraleSupélec-Université Paris Sud, France

Abstract—This paper studies the sum-capacity of the Multiple Input Single Output (MISO) Gaussian broadcast channel where K single-antenna users are served by a base station with N antennas, with $N < K$. The generalized Degrees-of-Freedom (gDoF) for this system is derived as the solution of a Maximum Weighted Bipartite Matching (MWBM) problem, where, roughly speaking, each of the N transmit antennas is assigned to a different user. The MWBM problem inspires a user selection algorithm where a subset of N out of K users is served. The proposed algorithm runs in polynomial-time (rather than involving an exhaustive search among all possible subsets of size N out of K users) and extends the classical DoF analysis to more realistic wireless channel configurations where users can experience very different channel gains from the base station. Extensive numerical simulations, run in practically relevant Rayleigh fading environments for different numbers of users and of antennas, show that the throughput achieved by serving the set of N users selected by the MWBM-based algorithm is at most $N \log(K)$ bits away from an outer bound to the sum-capacity, where in principle all the K users are served. Comparisons with another widely used user scheduling algorithm are also provided.

I. INTRODUCTION

Using antenna arrays at the transmitters or receivers turned out to be a powerful technique to enhance the performance of wireless networks in terms of coverage, throughput, Degrees-of-Freedom (DoF), and robustness / diversity. In a system with K single-antenna users connected to a base station with N antennas, it is well known that at most $\min\{N, K\}$ DoF are possible [1], or in other words, at high SNR, the base station can only deliver $\min\{N, K\}$ independent streams to the users. Since in practical systems $N < K$ (there are fewer antennas at the base station than the number of users to serve), it becomes important to design appropriate user selection techniques to identify which N out of the K users should be served for achieving maximum DoF [2].

Motivated by the importance of user scheduling in the downlink of Multiple Input Single Output (MISO) systems in this work we seek to determine throughput / sum-rate optimal policies by also taking into account the fact that users experience sometimes very different channel qualities. We therefore propose to go beyond DoF (for which all channel gains are assumed as fixed and the transmit power is let go to infinity) and study the generalized DoF (gDoF) of a K -user MISO Broadcast Channel (BC). The gDoF is an asymptotic performance measure in the high-SNR regime and refines the

classical DoF metric since it captures the fact that, in wireless networks, the channel gains can differ by several orders of magnitude. The gDoF, first introduced in [3] for the 2-user interference channel, has shed light on different operating regimes depending on the strength of the interference compared to the intended signal; moreover, the gDoF analysis has inspired relatively simple achievable schemes that are to within a constant gap of known outer bounds. The gDoF analysis has been restricted to Single Input Single Output (SISO) networks with a very small number of nodes. This is because the gDoF requires the specification of one parameter per channel gain, therefore becoming quickly prohibitively complex for large networks – even more in case of multi-antenna nodes. In this work, we propose a polynomial time (rather than an exhaustive search) algorithm to determine the gDoF of a K -user MISO BC, which consists in solving a Maximum Weighted Bipartite Matching (MWBM) problem where, roughly speaking, each transmit antenna at the base station is assigned to a different user.

A. Contributions

In this work we focus on the gDoF of a K -user MISO BC that, to the best of our knowledge, has not been characterized before. We attack the problem by leveraging our recent result on the connection between the gDoF of a MIMO point-to-point channel and the MWBM problem [4, Theorem 2]. We analyze the sum-capacity of the K -user MISO BC (where the base station is equipped with N antennas and where N is strictly smaller than the total number K of users). We propose a new user scheduling algorithm inspired by the solution of the MWBM problem. One of the most appealing feature of the proposed strategy is its low-complexity since polynomial-time routines exist to solve the MWBM problem that do not require an exhaustive search among all possible subsets of size N out of K users. We finally evaluate, through extensive numerical simulations, the throughput performance attained by the proposed user selection algorithm in ergodic fading channels. Simulations are run in practically relevant Rayleigh fading environments, for different values of K and N , and show that the throughput achieved by serving the set of N users obtained as output of the MWBM-based algorithm is at most $N \log(K)$ bits away from an outer bound on the theoretical sum-capacity. The sum-rate achieved by the MWBM-based algorithm is also compared with the one attained by

the algorithm proposed in [5], which can be implemented in currently used LTE multi-user MIMO base stations.

B. Related Work

The user scheduling problem is notoriously hard since often the space to be searched does not have “steepest descent” directions that would allow a fast (and possibly optimal) way to reach the desired stationary point. Actually, in general, this problem requires an exhaustive search over all subsets of users. This approach might be feasible in relatively small networks, but its complexity becomes prohibitive for large-scale networks. However, the fact that DoF-wise only $\min\{K, N\}$ independent streams can be delivered [1] suggests that the number of users effectively served is bounded by the number of antennas at the base station, since, in today’s cellular networks $N < K$. This fact is supported by the work in [6] where it was showed that the optimal sum-capacity transmission strategy schedules at most N^2 users. The prohibitive complexity of the brute-force best-set selection motivated the authors of [7] to design a suboptimal low-complexity algorithm for user selection, where at most N users are scheduled. In particular, at each step a user is selected only if this leads to a throughput increase with respect to the one attained in the previous step. It was numerically showed that this method, which employs zero-forcing beamforming, achieves a significant fraction of the sum-capacity. Another user selection algorithm was proposed in [5], where, at each step, the users are selected based on an orthogonality threshold. This approach, jointly with zero-forcing beamforming, was shown to achieve the sum-capacity when $K \rightarrow \infty$. In [8], it was demonstrated that, when the base station can transmit at large power, the sum-rate achieved by zero-forcing beamforming, where only N users are served, equals the sum-capacity. In [9], the authors proposed an algorithm that applies singular value decomposition to all channel matrices and selects only the eigenvectors whose corresponding singular values are above a threshold. Then, among these eigenvectors, only the N ones which are nearly orthogonal to each other are selected. In [9] it was also shown that the proposed algorithm achieves a significant fraction of the sum-capacity.

C. Paper Organization

The paper is organized as follows. Section II describes the static K -user MISO Gaussian BC and summarizes well-known sum-capacity results. Section III derives the gDoF in closed-form as the solution of a MWBM problem and presents a novel polynomial-time user selection algorithm where only the set of N users obtained as output of the MWBM-based algorithm is served. Section IV compares, through extensive numerical simulations, run in practically relevant Rayleigh fading models, the sum-rate achieved by the proposed MWBM-based user selection algorithm with the one proposed in [5]. Section IV also shows that the throughput achieved by the MWBM-based algorithm is at most $N \log(K)$ bits away from an outer bound on the theoretical sum-capacity where, in principle, all users are served. Section V concludes the paper.

D. Notation

Through the paper we use lower and upper case letters to indicate scalars, boldface lower case letters to denote vectors and boldface upper case letters to indicate matrices; $[n_1 : n_2]$ is the set of integers from n_1 to $n_2 \geq n_1$; $\mathbb{E}[\cdot]$ indicates the expected value; \mathbf{I}_N is the identity matrix of dimension N ; \mathbf{H}^\dagger is the Hermitian transpose of \mathbf{H} ; $|\mathbf{H}|$ is the determinant of \mathbf{H} ; $\text{Tr}(\mathbf{H})$ is the trace of \mathbf{H} ; $X \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ indicates that X is a proper-complex Gaussian random variable with mean μ and variance σ^2 .

II. SYSTEM MODEL

The static K -user MISO BC has input-output relationship

$$y_k = \mathbf{h}_k \mathbf{x} + z_k, \quad k \in [1 : K], \quad (1)$$

where the input $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is constrained to satisfy the average power constraint

$$\text{Tr}(\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]) \leq 1, \quad (2)$$

(a non-unitary power constraint can be incorporated into the channel gains), the vector $\mathbf{h}_k \in \mathbb{C}^{1 \times N}$ contains the channel gain coefficients from each transmit antenna at the base station to the k -th user and $z_k \sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

The transmitter has K messages; message $W_k \in [1 : 2^{LR_k}]$, $k \in [1 : K]$, has to be decoded from y_k^L , where L denotes the codeword length and R_k the transmission rate in bits per channel use (logarithms are in base 2). The transmitter sends $\mathbf{x}^L(W_1, \dots, W_K)$ and receiver k outputs the estimate $\widehat{W}_k(y_k^L)$. The capacity region is defined as the convex closure of all non-negative rates R_k , $k \in [1 : K]$, such that $\max_{k \in [1:K]} \mathbb{P}[\widehat{W}_k \neq W_k] \rightarrow 0$ as $L \rightarrow +\infty$. For the Gaussian BC the whole capacity region can be achieved by using Dirty-Paper-Coding (DPC) and time sharing [10]. Extension of the above setting to the ergodic fading case with full channel state information at all nodes is straightforward.

The focus of this paper is on the sum-capacity or throughput, which is given by the “Sato’s cooperative upper bound with least favorable noise correlation” [11]

$$C(\mathbf{H}) = \min_{\substack{0 \leq \mathbf{S}_z: \\ [\mathbf{S}_z]_{k,k} = 1, \forall k \in [1:K]}} \max_{\substack{0 \leq \mathbf{S}_x: \\ \text{Tr}(\mathbf{S}_x) \leq 1}} I(\mathbf{x}; \mathbf{H}\mathbf{x} + \mathbf{z}), \quad (3)$$

where $\mathbf{H} = [\mathbf{h}_1; \dots; \mathbf{h}_K] \in \mathbb{C}^{K \times N}$ is the overall channel matrix and $\mathbf{z} = [z_1; \dots; z_K]$ is the overall noise vector with covariance matrix $\mathbf{S}_z = \mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] \in \mathbb{C}^{K \times K}$. By exploiting the MAC-BC duality [12], [13], [14], the sum-capacity in (3) can be equally obtained by solving

$$C(\mathbf{H}) = \max_{\mathbf{D} \in \mathcal{D}} \log |\mathbf{I}_N + \mathbf{H}^\dagger \mathbf{D} \mathbf{H}|, \quad (4)$$

where \mathcal{D} is the set of the $K \times K$ non-negative diagonal matrices \mathbf{D} with $\text{Tr}(\mathbf{D}) \leq 1$.

It has been pointed out several times in the literature, (see for example [8, page 1697]), that in some regimes, such as when the base station can transmit at large power, serving only

$\min\{N, K\} = N$ destinations asymptotically attains the sum-capacity in (3). *In this work we seek to find a simple algorithm to select the best N users to serve and to show that the sum-rate achieved by the proposed user selection algorithm is close to (3).* The proposed user selection algorithm is based on the study of the gDoF of the system, which is defined as follows. For some $\text{SNR} > 0$ let the channel gains be parameterized as

$$|h_{k,n}|^2 = \text{SNR}^{\beta_{k,n}}, \quad \beta_{k,n} \geq 0, \quad (5)$$

for all $k \in [1 : K]$ and $n \in [1 : N]$. The gDoF, as a function of $\{\beta_{k,n}\}$, is defined as

$$d := \lim_{\text{SNR} \rightarrow \infty} \frac{C(\mathbf{H})}{\log(1 + \text{SNR})}, \quad (6)$$

where $C(\mathbf{H})$ is defined in (3).

III. MWBM-BASED USER SCHEDULING ALGORITHM

In order to determine the gDoF, we exploit our recent result in [4] that established a connection between the gDoF of a MIMO point-to-point channel and the MWBM problem. Based on this, we will present a MWBM-based user scheduling algorithm, which selects N users for transmission.

We first give some graph theoretical notions, which are crucial to understand our MWBM-based user selection algorithm. In graph theory, a weighted bipartite graph is a graph whose vertices can be partitioned into two sets in such a way that each edge, which has a non-negative weight, has exactly one endpoint in each set. For the K -user MISO BC with N transmit antennas, we define a weighted bipartite graph and its weight matrix \mathbf{B} as follows. The first set consists of K nodes (number of users), where K is the number of rows in \mathbf{B} , and the second set consists of N nodes (number of antennas), where N is the number of columns in \mathbf{B} . The element $[\mathbf{B}]_{k,n}$ is the weight of the edge between nodes $k \in [1 : K]$ and $n \in [1 : N]$, which is set to $[\mathbf{B}]_{k,n} = \beta_{k,n}$ defined in (5). A matching is defined as a set of edges without common vertices. The MWBM problem is a matching where the sum of the edge weights in the matching has the maximal value. The Hungarian algorithm is a polynomial time algorithm that efficiently solves the assignment problem in $O(\max\{K, N\}^3)$ [15].

In the following we assume that the phases of the fading channel gains are such that all involved channel (sub)matrices are full rank almost surely given the parameterization in (5). Since the constraints that define the set \mathcal{D} in (4) imply $\mathbf{D} \preceq \mathbf{I}_K$, we have that the sum-capacity is upper bounded by

$$C(\mathbf{H}) \leq \log |\mathbf{I}_K + \mathbf{H}\mathbf{H}^\dagger| = \log |\mathbf{I}_N + \mathbf{H}^\dagger\mathbf{H}|. \quad (7)$$

By applying [4, Theorem 2] to the right-hand side (RHS) of (7) we immediately find that the gDoF in (6) is upper bounded by

$$d \leq \text{MWBM}(\mathbf{B}), \quad \mathbf{B} = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,N} \\ \vdots & \ddots & \vdots \\ \beta_{K,1} & \cdots & \beta_{K,N} \end{bmatrix}, \quad (8)$$

that amounts to solve a MWBM problem with weight matrix \mathbf{B} given by the SNR-exponents $\{\beta_{k,n}, k \in [1 : K], n \in [1 : N]\}$ introduced in (5).

To gain insights into the result in (8), we next consider the case $N = 2$ (the result can be straightforwardly extended to a general N). Without loss of generality, let the antennas and the users be numbered in such a way that

$$\begin{aligned} |h_{1,1}| &\geq \max_{k \in [1:K], n \in [1:N]} |h_{k,n}| \iff \\ \beta_{1,1} &\geq \max_{k \in [1:K], n \in [1:N]} \beta_{k,n}, \end{aligned} \quad (9)$$

i.e., the link from antenna 1 to user 1 is the strongest among all links to any user from any of the antennas; then, by using (9) in (8), it is easy to see that

$$d^{(N=2)} \leq \text{MWBM} \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \vdots & \vdots \\ \beta_{K,1} & \beta_{K,2} \end{bmatrix} \quad (10)$$

$$= \text{MWBM} \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{k^*,1} & \beta_{k^*,2} \end{bmatrix}, \quad (11)$$

$$k^* := \arg \max_{k \in [2:K]} \{\beta_{1,1} + \beta_{k,2}, \beta_{1,2} + \beta_{k,1}\}, \quad (12)$$

or in other words, destinations 1 and k^* form the best set of $N = 2$ users to serve in order to attain the gDoF upper bound in (10). Let $\mathbf{H}_\pi \in \mathbb{C}^{2 \times 2}$ be the channel matrix that contains the channel gains of user 1 and user k^* . Since the constraints that define the set \mathcal{D} in (4) allow $\mathbf{D} = \frac{1}{2}\mathbf{I}_2$ (by allocating equal power among users 1 and k^*), we have that the sum-capacity is lower bounded by

$$C(\mathbf{H}) \geq \log \left| \mathbf{I}_2 + \frac{1}{2}\mathbf{H}_\pi\mathbf{H}_\pi^\dagger \right| \geq \log |\mathbf{I}_2 + \mathbf{H}_\pi^\dagger\mathbf{H}_\pi| - 2 \log(2).$$

By applying [4, Theorem 2] to the RHS of the above equation we immediately find that the gDoF $d^{(N=2)}$ is lower bounded by (11). This implies that the gDoF $d^{(N=2)}$ is given by (11) (the upper and lower bounds coincide). Thus, from (12) it is easy to see that gDoF-wise just serving $N = 2$ users, among the possible K , is optimal. Moreover, it is simple to understand which $N = 2$ users have to be served: user 1, i.e., the user who has the strongest link from the source, always has to be served, and the ‘‘second best’’ user is the one defined in (12).

By extending the above reasoning to any N and K , it is straightforward to prove that the solution of the MWBM problem in (8), which outputs the $N = \min\{N, K\}$ users to be scheduled, represents the gDoF of the K -user MISO BC as long as the channel matrix is full rank. One appealing feature of the proposed algorithm is that it runs in polynomial time.

IV. NUMERICAL EVALUATIONS

In this section we numerically assess the performance of the MWBM-based algorithm for different values of N and K . We let π be the set of users which attains the optimal MWBM in (8), and $\mathbf{H}_\pi \in \mathbb{C}^{N \times N}$ be the matrix that contains the channel gains of the users indexed by π . Clearly, $C(\mathbf{H}_\pi)$ is achievable by using the optimal DPC scheme for the channel

\mathbf{H}_π . Here we aim to evaluate $C(\mathbf{H}_\pi)$ and compare it with the sum-capacity $C(\mathbf{H})$ and with the sum-rate achieved by the algorithm proposed in [5], which can be implemented in today's multi-user MIMO base stations.

We consider the case of Rayleigh fading, where $h_{k,n} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_k^2)$, $k \in [1 : K]$ and $n \in [1 : N]$, is assumed to be constant over the whole slot (block-fading model), i.e.,

$$h_{k,n} = \sigma_k g_{k,n}, \quad (13)$$

where $g_{k,n} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. We define

$$\sigma_k^2 = \mathbb{E}[|h_{k,n}|^2] = \frac{c}{d_k^\alpha}, \quad (14)$$

where c is a constant that depends on the model parameters (e.g., base station's transmit power), d_k is the distance of the k -th user from the base station and $\alpha \geq 2$ is the path loss exponent. We assume a short-term average power constraint on the inputs, i.e., the power constraint in (2) is enforced in each fading block. With this model, we start by considering a *dynamic scheduling* that depends on $|h_{k,n}|^2$ (note that our proposed algorithm does not make use of phase information), which later on will be compared to a *static scheduling* based on $\mathbb{E}[|h_{k,n}|^2]$ only. We set

$$\beta_{k,n} = 10 \log_{10} \left(|h_{k,n}|^2 \right). \quad (15)$$

Note that multiplying the weight matrix \mathbf{B} in (8) by a constant and/or adding a constant to each matrix entry does not change the nature of the matching in the MWBM problem. We assume that the K users are independently uniformly distributed on an annulus with minimum radius equal to r_{\min} and maximum radius equal to r_{\max} . Moreover, we consider that the model parameters are such that the average SNR at the cell edges is $\text{SNR}(r_{\max})$, that is, the average SNR at distance d_k is given by

$$\begin{aligned} \text{SNR}(d_k) &:= \mathbb{E} \left[\sum_{n=1}^N |h_{k,n}|^2 \middle| d_k \right] = \frac{Nc}{d_k^\alpha} \\ &= \text{SNR}(r_{\max}) \left(\frac{d_k}{r_{\max}} \right)^{-\alpha}. \end{aligned}$$

Let (X, Y) be the coordinate of the random position of a user; then the cumulative distribution function (cdf) of its position is

$$\begin{aligned} F_{\frac{d}{r_{\max}}}(x) &= \mathbb{P} \left[\sqrt{X^2 + Y^2} \leq x r_{\max} \right] \\ &= \frac{\left[\min(1, x^2) - \frac{r_{\min}^2}{r_{\max}^2} \right]^+}{1 - \frac{r_{\min}^2}{r_{\max}^2}} \text{ for } x \geq 0, \end{aligned} \quad (16)$$

and hence the probability density function (pdf) is

$$f_{\frac{d}{r_{\max}}}(x) = \frac{2x}{1 - \frac{r_{\min}^2}{r_{\max}^2}} \text{ for } x \in \left[\frac{r_{\min}}{r_{\max}}, 1 \right]. \quad (17)$$

Fig. 1 shows the cdf of the throughput for different values of N and K , with fixed $\alpha = 3$, $\text{SNR}(r_{\min}) = 40$ dB, and $\text{SNR}(r_{\max}) = 0$ dB. The cdf was estimated with MATLAB

command `ecdf` with a confidence level of 0.05 (default value) whose input was generated by considering $N_{\text{iter},1} = 100$ different user positions (i.e., for each $k \in [1 : K]$ we consider $N_{\text{iter},1} = 100$ different values of d_k in (14)), for each of which we considered $N_{\text{iter},2} = K \cdot 10^3$ different realizations of $g_{k,n}$ in (13), $k \in [1 : K]$, $n \in [1 : N]$. In Fig. 1, the average throughput $\mathbb{E}[C(\mathbf{H}_\pi)]$ achieved by our MWBM-based algorithm is also reported for all values of K and N .

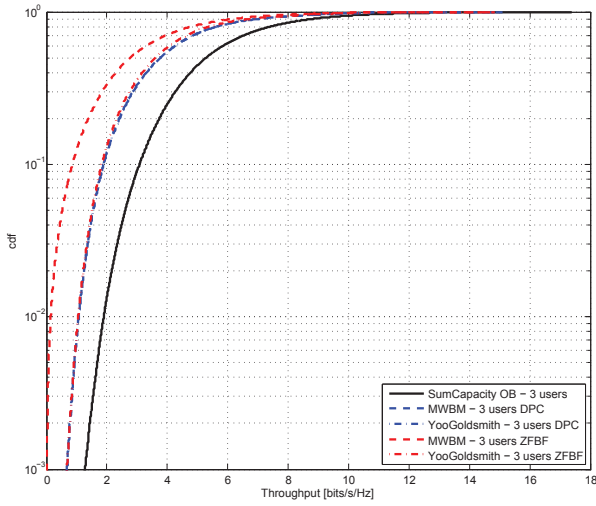
From Fig. 1, we observe that the throughput performance of our MWBM-based algorithm is very close to the one of [5] when DPC is used in both cases (blue dashed lines versus dash-dotted lines). Differently from [5], our scheduling algorithm does not use the knowledge of the channel phases. This means that, in a practical scenario, less information has to be fed back to the base station for the purpose of scheduling users. Once our MWBM-based algorithm has selected the N users to serve, only the channel phases of the N selected users need to be fed back to the base station in order to implement the DPC strategy. In other words, given a fixed amount of bits on the feedback link, the base station can get a more accurate representation of the phases of the N selected users, as opposed to [5] that requires phases from all the K users.

From Fig. 1 we also observe that, if Zero-Forcing Beam-Forming (ZFBF) is used instead of DPC, our algorithm does not perform as well as the one in [5] (red dashed lines versus dash-dotted lines). This is because ZFBF is most effective when the selected users have nearly orthogonal channel vectors. Hence, when ZFBF is used, it becomes essential to schedule those users whose channel gains are as orthogonal as possible. Thus, the knowledge of the channel phases becomes critical. Our MWBM-based algorithm, which is based on a ‘‘coarse’’ approximation of the channel gains (since only the magnitude of the channel gains is considered while the phases are neglected), does not capture this aspect. This appears to cost in performance at low-SNR if ZFBF is employed. Indeed we expect our MWBM-based algorithm to be nearly optimal at high-SNR, where the phases become negligible; in the simulated scenario the average SNR, averaged over the random positions of the users, is

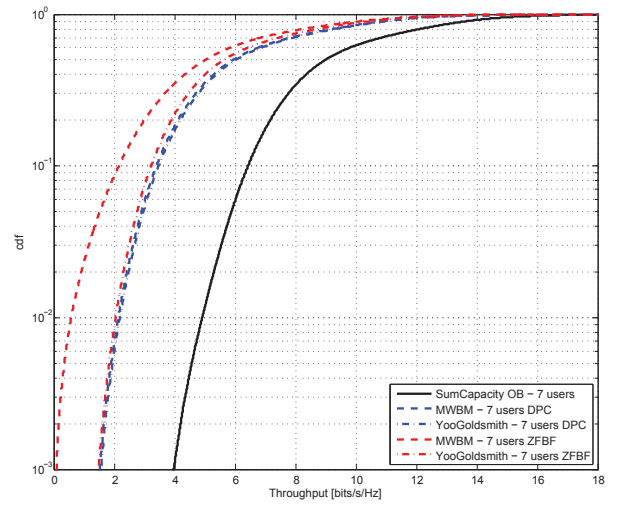
$$\begin{aligned} \mathbb{E}[\text{SNR}(d)] &= \text{SNR}(r_{\max}) \int_{\frac{r_{\min}}{r_{\max}}}^1 \frac{2x^{1-\alpha}}{1 - \frac{r_{\min}^2}{r_{\max}^2}} dx \\ &= \text{SNR}(r_{\max}) \frac{2 \left(1 - \left(\frac{r_{\min}}{r_{\max}} \right)^{2-\alpha} \right)}{(2-\alpha) \left(1 - \frac{r_{\min}^2}{r_{\max}^2} \right)} = 16.1481 \text{ dB}, \end{aligned}$$

which is far from being in the high-SNR regime, thus explaining the better performance of [5] if ZFBF is used.

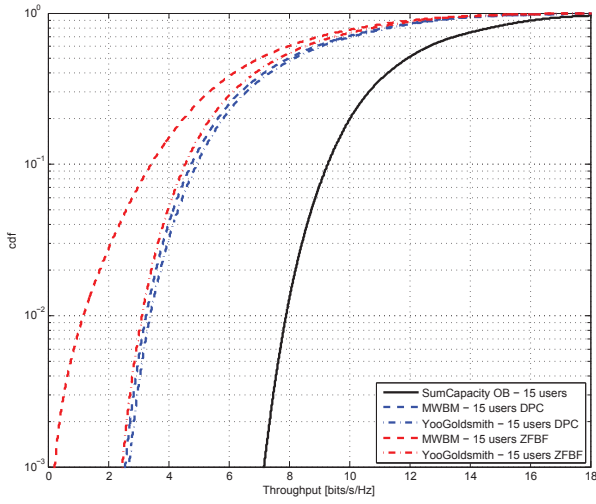
Fig. 1(d) shows that the throughput increases when the number of users increases for a fixed value of N . This is due to *multiuser diversity*: as K increases for a fixed N , the base station has a larger pool of users to choose from and it is therefore more likely to find a subset of users with ‘good’ channels thereby attaining a larger throughput. Fig. 1(d) also shows the throughput performance of our MWBM-based al-



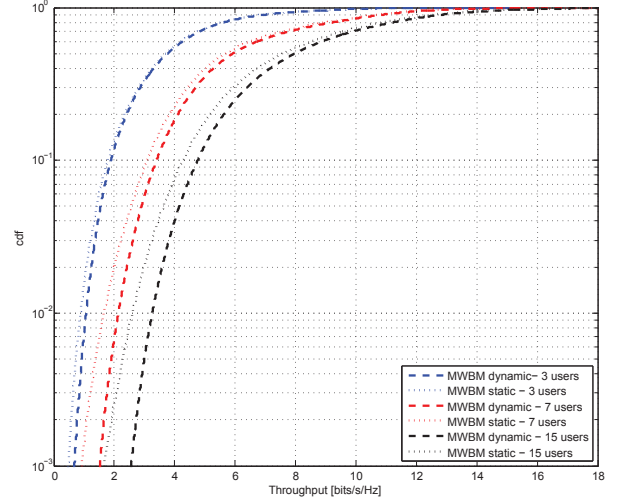
(a) $K = 3, N = 2, \mathbb{E}[C(\mathbf{H}_\pi)] = 4.1840$ bits/s/Hz.



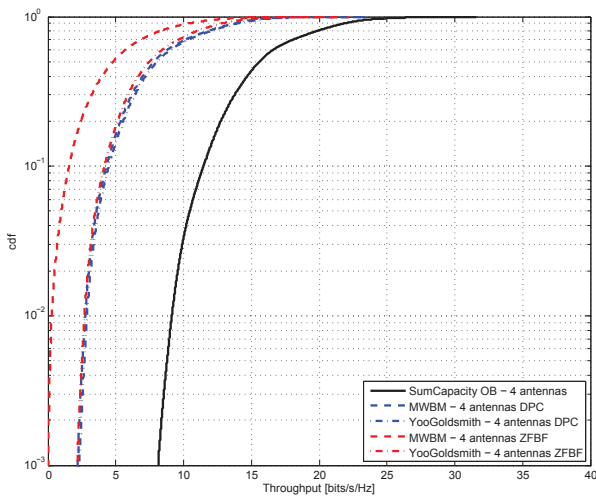
(b) $K = 7, N = 2, \mathbb{E}[C(\mathbf{H}_\pi)] = 6.6346$ bits/s/Hz.



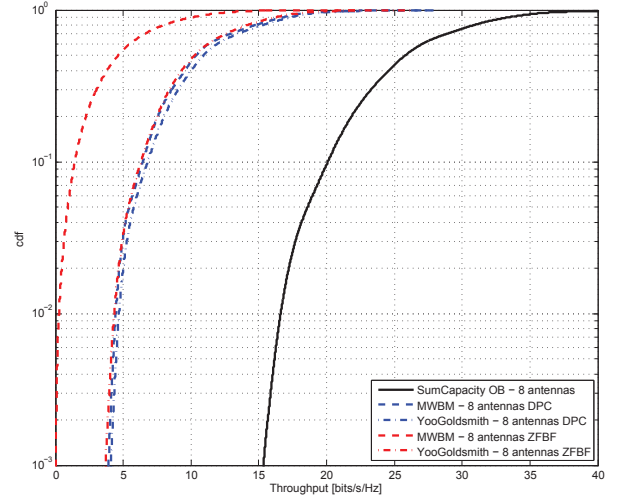
(c) $K = 15, N = 2, \mathbb{E}[C(\mathbf{H}_\pi)] = 8.4949$ bits/s/Hz.



(d) $K \in \{3, 7, 15\}, N = 2.$



(e) $K = 10, N = 4, \mathbb{E}[C(\mathbf{H}_\pi)] = 8.5260$ bits/s/Hz.



(f) $K = 10, N = 8, \mathbb{E}[C(\mathbf{H}_\pi)] = 11.1568$ bits/s/Hz.

Fig. 1: Cumulative density function (cdf) of the throughput with $N \in \{2, 4, 8\}$ and $K \in \{3, 7, 15\}$.

gorithm when a *static scheduling* is performed, i.e., a schedule which is based only on the fading expected value. We observe that the *dynamic scheduling* (dashed lines) outperforms the *static scheduling* (dotted lines), since the former is adapted to each instantaneous channel realization; this aspect is critical especially at low throughput.

Figs. 1(e)-1(f) show that the throughput increases when the number of antennas increases for a fixed K . This is due to the *multiplexing gain*: for a fixed K , as the number of transmit antennas increases (always considering $N < K$), more users can be served leading to a throughput's boost.

Finally, we remark that in Fig. 1 the black curves represent the sum-capacity outer bound in (7) and not the exact sum-capacity in (4). Numerically, we notice that the gap between the black curves (outer bounds to the sum-capacity) and the achievable throughputs grows with K and N . In particular, the gap is always smaller than $N \log(K)$ bits since $\mathbf{D} = \mathbf{I}_K$ gives the sum-capacity upper bound in (7) while using $\mathbf{D} = \frac{1}{K} \mathbf{I}_K$, i.e., the power is equally split among the K users, gives a sum-capacity lower bound. The difference between these upper and lower bounds is upper bounded by

$$\begin{aligned} \text{GAP} &\leq \log |\mathbf{I}_N + \mathbf{H}^\dagger \mathbf{I}_K \mathbf{H}| - \log \left| \mathbf{I}_N + \mathbf{H}^\dagger \frac{1}{K} \mathbf{I}_K \mathbf{H} \right| \\ &\leq \log |\mathbf{I}_N + \mathbf{H}^\dagger \mathbf{I}_K \mathbf{H}| - \log \left(|\mathbf{I}_N + \mathbf{H}^\dagger \mathbf{I}_K \mathbf{H}| \left(\frac{1}{K} \right)^N \right) \\ &= N \log(K). \end{aligned}$$

In general, it is not possible to prove that the achievable throughput $C(\mathbf{H}_\pi)$, with our MWBM-based algorithm, is a constant number of bits away from the sum-capacity $C(\mathbf{H})$ (or an upper bound on it) uniformly over all channel gains. In particular, these cases occur when the channel matrix \mathbf{H}_π is ill-conditioned, a feature which is not captured by the gDoF analysis. However, as also observed numerically, this kind of scenario is unlikely to occur and $C(\mathbf{H}_\pi)$ appears to be at most $N \log(K)$ bits apart from the sum-capacity outer bound in (7). Moreover $C(\mathbf{H}_\pi)$ is asymptotically optimal at high-SNR and gives the optimal gDoF.

V. CONCLUSION

In this work we studied the K -user MISO Gaussian broadcast channel where the number of antennas N at the base station is strictly smaller than K , thus motivating the question of what is the best set of users to schedule for transmission in order to attain a throughput that is not that far from the optimal sum-capacity. We proposed a MWBM-based user scheduling algorithm that selects N users for transmission and is asymptotically optimal at high-SNR. The proposed algorithm, which runs in polynomial time, solves a MWBM problem, where, roughly speaking, each transmit antenna at the base station is assigned to a different user.

For the Rayleigh fading channel, we numerically evaluated the throughput cumulative density function with the proposed user scheduling algorithm and compared it with that of a state-of-the-art user scheduling algorithm. We numerically showed

that the sum-rate achieved by our proposed algorithm is at most $N \log(K)$ bits away from an outer bound on the sum-capacity where, in principle, all the K users are served.

Future work will consider the study of different fading models and the comparison with others scheduling algorithms known in the literature.

ACKNOWLEDGMENT

Eurecom's research is partially supported by its industrial partners: BMW Group Research & Technology, IABG, Monaco Telecom, Orange, SAP, SFR, ST Microelectronics, Swisscom and Symantec. The research carried out at Eurecom has also received funding from the EU Celtic+ Framework Program Project SHARING and from a 2014 Qualcomm Innovation Fellowship. The work of Dr. D. Tuninetti was partially funded by NSF under award number 1218635; the contents of this article are solely the responsibility of the author and do not necessarily represent the official views of the NSF.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–598, Nov. 1999.
- [2] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.
- [3] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
- [4] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim, "Gaussian half-duplex relay networks: improved constant gap and connections with the assignment problem," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3559–3575, June 2014.
- [5] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [6] W. Yu and W. Rhee, "Degrees of freedom in wireless multiuser spatial multiplex systems with multiple antennas," *IEEE Transactions on Communications*, vol. 54, no. 10, pp. 1747–1753, Oct. 2006.
- [7] G. Dimic and N. Sidropoulos, "Low-complexity downlink beamforming for maximum sum capacity," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2004*, vol. 4, May 2004, pp. iv–701–iv–704 vol.4.
- [8] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
- [9] A. Bayesteh and A. Khandani, "On the user selection for MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1086–1107, Mar. 2008.
- [10] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.
- [11] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 374–377, May 1978.
- [12] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.
- [13] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
- [14] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.
- [15] H. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.