REAL-TIME SCHEDULING FOR WIRELESS MULTIUSER MISO SYSTEMS UNDER DIFFERENT FAIRNESS CRITERIA

Diego Bartolomé, Daniel P. Palomar, and Ana I. Pérez-Neira *

dbartolome@ieee.org, {daniel, anuska}@gps.tsc.upc.es

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
c/ Gran Capità 2-4, Edifici Nexus I, Planta 2
08034 Barcelona (Spain)

Universitat Politècnica de Catalunya (UPC)
c/ Jordi Girona 1-3, Mòdul D5, Campus Nord UPC
08034 Barcelona (Spain)

ABSTRACT

This paper deals with the spatio-temporal scheduling of a set of users for downlink transmission in a cell where the base station is provided with multiple antennas. Two main issues are addressed: how the users are distributed into space-time groups and how the available power is allocated among the users within a group. Since the former is an NP-complete combinatorial problem, we develop fast and low-complexity algorithms, which might be capable to fulfill the real-time requirements of a practical scheduling scheme. Regarding the power allocation, we consider different fairness criteria under the capacity point of view. To be precise, we compare the alternatives of equal rate and maximum sum rate with our proposed equal proportional rate solution.

1. INTRODUCTION

Multiple Element Antenna (MEA) systems may provide a large increase in capacity for future wireless communications standards. Here, we assume that the Base Station (BS) is provided with multiple antennas and the terminals have a single one. Particularly, the problem addressed in this paper consists of the scheduling of a large set of users in the downlink of a TDMA system. The solution to this problem can be separated into three different parts: i) the beamforming technique, ii) the power allocation schemes, and iii) the division of users into groups.

Optimal beamforming is solved in [1], but this kind of techniques may be difficult to implement in a practical system if there are real-time and low-complexity requirements. Therefore, a Zero Forcing (ZF) design criterion might provide an acceptable performance with much lower computational load [2]. Furthermore, [2] proves the benefits of assigning each antenna from the BS to a different user instead of assigning all the antennas to the same user.

Once a beamforming design is chosen, the total available power shall be distributed among the users. In the literature, two main techniques can be found for this difficult question: the maximum sum rate and the equal rate [4]. Besides, the definition of fairness is not clear. One could consider that fairness implies assigning the same rate to all users, while the so-called proportional fairness aims for providing a higher rate to the users closer to the BS. It is not clear either, which is the optimum rate sharing.

Finally, if a huge amount of users are active in the cell, the issue of clustering[1] them becomes an NP-hard combinatorial prob-

lem. In [3] the author solves this problem by a graph-theoretical technique, whereas in [4] we proposed the application of Simulated Annealing to this problem. However, this kind of techniques are too costly to be implemented in real-time, which is a mandatory requirement for a practical scheduling technique.

This paper addresses those three problems. First, in Section 2 we select ZF as the beamforming design criterion for the sake of simplicity. After that, we focus on the power allocation techniques, comparing the *traditional* equal rate and maximum sum rate approaches to our novel equal proportional rate scheme. This solution comes naturally with the definition of a fairness index that relates the actual rate with that a user would get if it were alone in the group (see Section 3 for further details and discussions). Finally, in Section 4 we investigate on simple techniques for the user clustering, which might be implemented in real-time and provide a better trade-off between performance and complexity than those in [3] and [4]. The simulations in Section 5 show the performance of the proposed schemes before the final conclusions.

2. PROBLEM STATEMENT

In the following, boldface capital (lowercase) letters refer to matrices (vectors). The conjugate transpose of $\mathbf{a}$ is given by $\mathbf{a}^H$ and the element at row ith and column jth of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. The natural logarithm of a is $log(a)$ and $(a)^+ = \max(0, a)$.

We consider the downlink of a communication system, where the BS is provided with Q transmit antennas, although the notation and solutions presented henceforth can be also applied to the uplink. Let $K \geq Q$ be the number of active users in the cell, each having a single antenna. The users shall be distributed into G groups. Each group is scheduled for transmission in a different time slot, whereas the K_g users in the gth group are served simultaneously by a Space Division Multiple Access (SDMA) scheme. As stated before, we have to specify first a bemforming design. Then, the received signal vector for group g at instant n is:

$$\mathbf{y}(n) = \mathbf{HBs}(n) + \mathbf{w}(n) \in \mathbb{C}^{K_g \times 1}, \qquad (1)$$

where the index g denoting the group has been omitted for simplicity. The kth position of vector $\mathbf{y}(n)$ ($\mathbf{s}(n)$) is the received (transmitted) signal for user k. $\mathbf{H}$ is the $K_g \times Q$ complex flat-fading channel matrix, the ith row of which contains the $1 \times Q$ vector of the channel gains for the ith user, i.e. $\mathbf{h}_i^H$, which is essentially determined by the Direction of Arrival (DOA) of the user and a scaling due to the path-loss and fast fading, which is the same for each antenna element[2]. Further, the components of the

---

[1]Clustering refers to the partition of the active users into groups.

[2]Note that in the more general case of independent fading for each ele-

noise vectors are i.i.d. gaussian random variables with zero mean and variance σ_w^2. Finally, the beamvectors for the K_g users are gathered in the matrix $\mathbf{B} = [\mathbf{b}_1 \mathbf{b}_2 \ldots \mathbf{b}_{K_g}] \in \mathbb{C}^{Q \times K_g}$.

First, we would like to separate the effect of the channel and that of power allocation. Therefore, given a set of groups, the problem is divided into two parts: first, we apply a unitary beamforming for each user in a group; second, the best-suited strategy for the power allocation within a group is chosen (Section 3).

In this paper, we have applied a ZF beamformer design for the downlink communication for simplicity. ZF implies that the K_g channels become parallel and orthogonal, thus there is no interference among the users belonging to the same group. Therefore, we can express the beamforming matrix as $\mathbf{B} = \widetilde{\mathbf{B}} \mathbf{D}_\beta$, where $\widetilde{\mathbf{B}}$ denotes the unitary beamforming and $\mathbf{D}_\beta$ is a diagonal matrix representing the allocated power for each user, i.e. $\mathbf{D}_\beta = diag(\beta_1, \beta_2, \ldots, \beta_{K_g})$. The effect of the channel is considered in the diagonal matrix $\mathbf{H}\widetilde{\mathbf{B}} = \mathbf{D}_\alpha = diag(\alpha_1, \alpha_2, \ldots, \alpha_{K_g})$. With all this, the received signal for user k can be expressed as:

$$y_k(n) = \alpha_k \beta_k s_k(n) + w_k(n), 1 \leq k \leq K_g, \qquad (2)$$

where, as stated, we have a set of parallel and orthogonal channels for each group. With these considerations, we have not yet described how the beamformer is computed. To obtain the beamvector, the ZF criterion requires that $\mathbf{H}\mathbf{b}_k = \mathbf{1}_k, \forall k$. Additionally, we apply the normalization factor $\alpha_k, \forall k$ to guarantee a unitary beamforming. Then, the beamvector is computed as[3]:

$$\widetilde{\mathbf{b}}_k = \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H \right)^{-1} \mathbf{1}_k \alpha_k, \qquad (3)$$

where $\alpha_k = 1/\sqrt{[(\mathbf{H}\mathbf{H}^H)^{-1}]_{k,k}}$. We have then two pending issues: the strategies for the power allocation (Section 3) and how the users are divided into groups (Section 4).

3. FAIRNESS AND POWER ALLOCATION

The criteria for the power allocation are determined by the fairness definition we consider, for which there are several points of view. A first possibility would consist of assigning the same rate to all users, which will appear as the best solution at a first sight. However, the users that are closer to the BS, i.e. those having a greater mean Signal to Interference and Noise Ratio (SINR), should be granted a better rate if we take into account the proportional fairness rule. On the the other hand, the second option can optimize the total performance of the group regardless of some users that might not even be allocated for transmission. The former is the equal rate scheme, whereas the latter is the maximum sum rate.

In this paper, we would like to have an intermediate solution: we wish to have a mean rate slightly lower than what the maximum sum rate achieves, but increasing the fairness. For this purpose, we need to specify first a fairness criterion. In agreement with the proportional fairness, a well-suited index seems to be the one that links the actual rate obtained by a certain user C_k with the rate it

would achieve if it were alone in the group C_k^{max}:

$$C_k = log \left(1 + \frac{|\alpha_k|^2 |\beta_k|^2}{\sigma_w^2} \right), \qquad (4)$$

$$C_k^{max} = log \left(1 + \frac{|\alpha_k|^2 P_T}{\sigma_w^2} \right), \qquad (5)$$

where $\text{SINR}_k = \frac{|\alpha_k|^2 |\beta_k|^2}{\sigma_w^2}$ and $\text{SINR}_k^{max} = \frac{|\alpha_k|^2 P_T}{\sigma_w^2}$[4] can be defined. The proposed Fairness Index (FI) is then the following:

$$\text{FI}_k = \frac{C_k}{C_k^{max}}, \qquad (6)$$

which is clearly in the range $0 \leq \text{FI}_k \leq 1$. The upper bound occurs if all the available power is allocated to user k, thus getting the capacity C_k^{max}, while the lower bound reflects a situation of no scheduling or null power. Based on this FI, we propose our novel equal proportional rate in Subsection 3.3.

Finally, note that we are evaluating the *instantaneous* fairness. We can further increase the capacity by taking into account the temporal variation of the channel as in [5].

3.1. Equal Rate Scheme

In this case, we want the same SINR (rate) for all users, regardless of their position or channel quality. Therefore, we have to impose that the product $\alpha_k \beta_k$ is the same for all users and equal to $\bar{\beta}$. This is equivalent to the maximization of the minimum capacity for all users within a group. Recalling the definition of the capacity C_k in (4), the problem is

$$\max_{|\beta_k|^2} \min_k C_k \qquad (7)$$

$$s.t. \sum_{k=1}^{K_g} |\beta_k|^2 \leq P_T, \qquad (8)$$

where the total output power is limited by P_T due to regulatory issues. Applying the restriction on the total power, $\bar{\beta}$ reduces to

$$\bar{\beta} = \sqrt{\frac{P_T}{tr\left((\mathbf{H}\mathbf{H}^H)^{-1}\right)}}, \qquad (9)$$

where tr denotes the trace operation. With all this, the capacity is the same for all users in the group, $C_k = C, \forall k$, and it is given by

$$C = log \left(1 + \frac{1}{\sigma_w^2} \frac{P_T}{tr\left((\mathbf{H}\mathbf{H}^H)^{-1}\right)} \right). \qquad (10)$$

3.2. Maximum Sum Rate

In this case, instead of guaranteeing the same SINR for all users, we would like to optimize the global performance of the cell, without considering fairness. It could happen that some users in a group might not be able to transmit, allowing other users to have a higher SINR and thus a higher capacity. Then, we maximize the total capacity subject to an instantaneous power constraint, i.e.

$$\max_{|\beta_k|^2} \sum_{k=1}^{K_g} C_k \qquad (11)$$

$$s.t. \sum_{k=1}^{K_g} |\beta_k|^2 \leq P_T. \qquad (12)$$

---

[3]The pseudo-inverse reduces to the inverse if we allocate Q users to Q antennas. And strictly speaking, it is ZF up to a scalar for each substream.

ment, the algorithms would certainly work. However, the DOA would no longer mean a physical angle at the BS.

[4]Note that in this case $|\alpha_k|^2 = \|\mathbf{h}_k\|^2$.

Table 1: Initialization procedure for user clustering

1. Compute the cost of all possible combinations of two users $J_{i,j}, \forall i \neq j$.

2. Sort the cost values in descending order in the vector $\mathbf{c}^s$, keeping the information of the indexes of users i and j. Set l to 0.

3. Increase l and select that position l in $\mathbf{c}^s$. If users i, j corresponding to index l have not yet been assigned, separate them in different groups $g(i) \neq g(j)$. Note that $g(i)$ is the group where we put user i and that only one user shall be assigned per group.

4. If the number of assigned users is not G, go to step 3.

Taking derivatives of the Lagrangian of the problem with respect to the power allocation factors $|\beta_k|^2$ and to the Lagrange multiplier μ, we obtain the following water-filling algorithm:

$$|\beta_k|^2 = \left(\mu^{-1} - \frac{\sigma_w^2}{|\alpha_k|^2} \right)^+ , \qquad (13)$$

where μ is chosen to satisfy the power constraint in (12) with equality. For implementation issues, we refer the reader to [6]. With this approach, the capacity for the kth user is:

$$C_k = \left(log \left(\frac{\mu^{-1}|\alpha_k|^2}{\sigma_w^2} \right) \right)^+ . \qquad (14)$$

3.3. Equal Proportional Rate

In this case, for each group we want to *equalize* and maximize the FI, i.e. the users in a group shall have the same proportion of the rate they would get if they were alone in the group, thus yielding to a proportionally fair solution. This can be expressed as

$$\max \delta \qquad (15)$$

$$s.t. \ FI_k = \delta, \forall k, \qquad (16)$$

$$\sum_{k=1}^{K_g} |\beta_k|^2 \leq P_T. \qquad (17)$$

We expect to have a lower capacity than in the sum rate, but we shall overcome the problem of assigning a null power to certain users. The solution to this problem can be easily found, but it is left out for space limitations.

4. USER CLUSTERING

As stated before, user clustering refers to the separation of the users into groups. The optimal solution would imply exhaustive search or techniques based in graph-theory [3], which are NP-hard combinatorial problems. On the other hand, Simulated Annealing can solve the problem in much less time [4], but it may neither fulfill the real-time limitations. Suboptimal approaches where developed in [4]. Since the aim was assuring the same rate for all users (equal rate scheme), the proposed solution had to compute the minimization of the $tr\left(\left(\mathbf{H}\mathbf{H}^H \right)^{-1} \right)$, which is tantamount to the maximization of the capacity as it is seen in (10).

Since the approach in [4] could still be unfeasible for a practical system, we propose here even simpler techniques. Intuitively, the users that are closer in terms of DOA should be distributed into separate groups, since they would require an increase in power.

Table 2: Iterative User Clustering

1. Build matrix $\mathbf{D}$ as in (19).

2. Store the cost values, i.e. the ratio between the maximum and minimum values for each row in matrix $\mathbf{D}$, in $\mathbf{c}$ and keep the information about the user and about the group having the minimum cost for that user.

3. Sort $\mathbf{c}$ in descending order, $\mathbf{c}^s$ is the sorted vector. Set position index l to 0. Note that $g(l)$ is the group having minimum cost for the user at position l.

4. Increase l and select that position in the sorted vector $\mathbf{c}^s$. If the group $g(l)$ has not been filled in current iteration, assign to group $g(l)$ the user corresponding to index l.

5. Go to step 4 until l points to the last element of $\mathbf{c}^s$ or if the G groups have been filled by a different user at this iteration.

6. Go to step 1 until every user is assigned to a group.

Table 3: Simple User Clustering

1. Sort the users in descending order of $\|\mathbf{h}\|^2$ in $\mathbf{c}$ and set $k = 0$.

2. Increase index k, and select that position in $\mathbf{c}$. For that user, build the vector $\mathbf{d}_k$ for the G' available groups $[d_k]_{g'} = \max_l J_{k,l}$.

3. Assign the user corresponding to index k to the group having minimum d_k.

4. Go to step 2 until all users are assigned to a group.

Therefore, we use the scalar product as a cost function for the clustering techniques, which has a meaning of angle. For each pair of users we define the *grouping* cost by the following expression

$$J_{i,j} = \frac{\mathbf{h}_i^H \mathbf{h}_j}{\|\mathbf{h}_i\| \|\mathbf{h}_j\|}. \qquad (18)$$

Then, the initialization procedure of the two algorithms is described in Table 1. Essentially, the users closer in terms of angle, those having a greater cost in (18), are separated into different groups. After that, G users are assigned to G groups. Then, at each iteration of the algorithm we need to compute the cost of assigning the user kth to the G' available groups[5], i.e.

$$[\mathbf{D}]_{k,g'} = \max_l J_{k,l}, \qquad (19)$$

which is the worst case regarding the user clustering, i.e. the cost is determined by the user closer to the desired user in terms of angle.

Table 2 describes the first method, which tries to compute iteratively the best group for the users. The main idea behind is to allocate first the users having a greater dispersion in being assigned to the groups, i.e. the users that would penalize the system if they are not assigned to the best possible group.

On the other hand, a simpler algorithm has also been evaluated: we sort the users in descending order of the module of $\mathbf{h}_i$, and start assigning the users to the groups. Now, the cost for each group is defined by the maximum scalar product, and we assign the user to the group where it is lowest, i.e. to the best worst case (min-max). This is summarized in Table 3.

5. SIMULATIONS

In the simulations we have considered a BS with $Q = 4$ transmit antennas and five groups $G = 5$ in a fully-loaded system, thus

---

[5]We consider that group g' has K_g' users at that iteration, $1 \leq l \leq K_g'$.
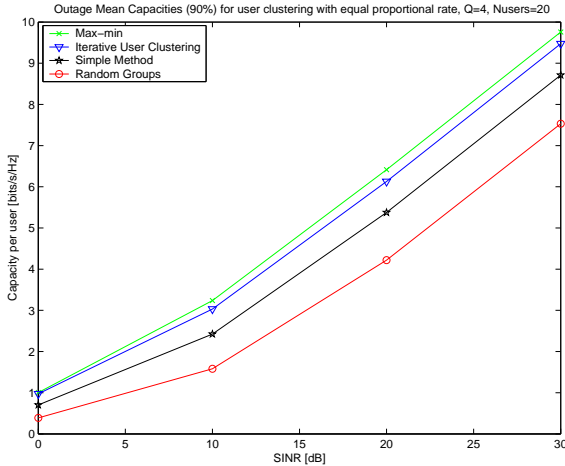
Figure 1: Mean capacity for the user clustering techniques with equal proportional rate power allocation.

the total number of users is $K = 20$. The channel is flat fading and consists of an equal path-loss and fast fading for all antennas and a DOA, which is distributed uniformly between $-60°$ and $60°$ (sectored antennas at the BS). The SINR[6] goes from 0 to 30 dB.

First, simulations have been conducted to evaluate the user clustering schemes. In Figure 1, we plot the outage capacity for the users vs. the SINR. We see the performance gain it is obtained if the users are distributed into groups in an intelligent manner rather than with a random scheduling. The best trade-off between performance and complexity is obtained by our simple iterative method, face off the max-min technique developed in [4]. Therefore, the figures are given for this strategy in the following. Note also that the simple user clustering provides an intermediate performance between the random scheduling and the iterative method with much less complexity than the latter.

Figure 2 evaluates the power allocation strategies in terms of maximum, mean, and minimum outage capacities for the users. As expected, the mean capacity is maximized by the sum rate approach, but the low minimum capacity indicates that some users might be penalized. The equal rate technique has the same maximum, minimum, and mean values, coming from the design of the problem. The intermediate performance is achieved by the equal proportional rate scheme, which increases the rate of the *poor* users paying the price of some loss in terms of mean capacity.

Finally, in Figure 3 we evaluate the FI for the three proposed techniques. In that case, the FI is optimized by the equal proportional rate approach. Further, we see that the equal rate technique is not fair, since we are wasting some power in poor users and the dispersion from the maximum and minimum fairness index is high. Interestingly, the maximum sum rate solution tends to the equal proportional rate when the power is increased.

6. CONCLUSIONS

In this paper, we have studied two main problems, namely the spatio-temporal clustering of a set of users for simultaneous transmission and the power allocation within a given group. We have proposed simple clustering techniques based on the scalar product, providing a good performance compared to other existing approaches. Regarding the power allocation, an evaluation of the

---

[6] In this case, the SINR refers to the ratio $\frac{P_T}{\sigma_w^2}$.
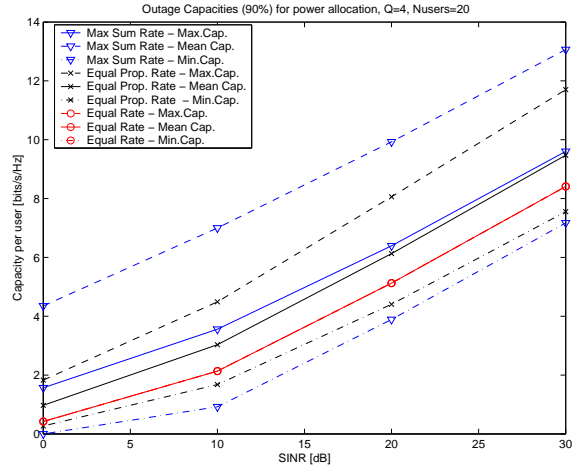

Figure 2: Outage maximum (dashed), mean (solid), and minimum (dash-dotted) capacities for the proposed power allocation strategies. The groups are formed by the iterative user clustering.
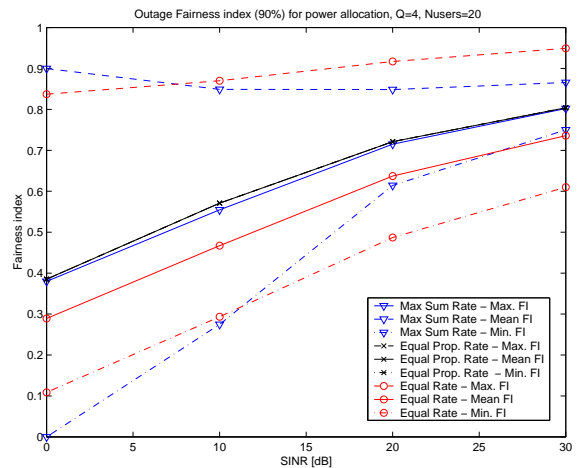

Figure 3: Outage maximum (dashed), mean (solid), and minimum (dash-dotted) fairness index for the power allocation strategies. The groups are formed by the iterative user clustering.

traditional equal rate and maximum sum rate techniques is performed. The best trade-off is obtained, though, by our novel equal proportional rate scheme.

7. REFERENCES

[1] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, Lal Chand Godara, Ed. CRC Press, 2001.

[2] R.W. Heath and M. Airy and A.J. Paulraj, "Multiuser Diversity for MIMO Wireless Systems with Linear Receivers," in *Asilomar*, 2001.

[3] R. Zhang, "Scheduling for Maximum Capacity in SDMA/TDMA Networks," in *ICASSP*, 2002.

[4] D. Bartolomé, A. Pascual-Iserte and A. I. Pérez-Neira, "Spatial Scheduling Algorithms for Wireless Systems," *in Proceedings of ICASSP'03*, April 2003.

[5] D.N.C. Tse, "Multiuser Diversity in Wireless Networks," *Wireless Communications Seminar, Standford University*, April 2001.

[6] D. P. Palomar and M. A. Lagunas, "Joint Transmit-Receive Space-Time Equalization in Spatially Correlated MIMO channels: a Beamforming Approach," *to appear in IEEE Journal on Selected Areas in Comm.*, June 2003.