SPATIAL SCHEDULING ALGORITHMS FOR WIRELESS SYSTEMS

Diego Bartolomé, Antonio Pascual-Iserte, Ana I. Pérez-Neira *

{diego, tonip, anuska}@gps.tsc.upc.es

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

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

ABSTRACT

This paper addresses the problem of the spatial scheduling of users in a cell for simultaneous downlink transmission from a Base Station (BS) having multiple antennas under a perspective of joint Physical and Medium Access Control (PHY-MAC) design. First of all, we compute the transmit beamvectors for each group according to a Zero Forcing (ZF) criterion, which gives a simple closed-form solution. We show first that it is equivalent to the minimization of the maximum Bit Error Rate (BER). In this paper, the main contribution lies on the resolution of the NP-complete combinatorial problem that comes up as cost function if we want to minimize the total transmit power. The solution of the NP-complete problem is performed by the stochastic technique Simulated Annealing (SA). Additionally, we present two heuristic algorithms that may enable a real-time implementation of this scheduling approach.

1. INTRODUCTION

Multiple Element Antenna systems may provide a large increase in capacity for future wireless communications standards. However, in realistic scenarios and systems, the number of antennas might be limited by the high cost of the RF front-ends. Therefore, we focus on the use of a small set of antennas at the BS, or at the Access Point (AP) in Wireless Local Area Networks (WLANs).

In this paper, we address mainly the problem of clustering a set of users for simultaneous transmission in the downlink of a TDMA system. We assume that the BS has to allocate the K users in the cell in groups of Q, which is the number of transmit antennas. After solving this combinatorial problem, G groups are formed and the scheduler at the BS will choose how to allocate them in different time slots, but this is out of the scope of this paper.

To begin with, we have to choose a transmit beamforming design. Optimal downlink beamforming is addressed in [1] or in [2]. In the former, the transmit power is minimized while assuring a certain Signal to Interference and Noise Ratio (SINR) at the receivers. However, this kind of techniques are difficult to implement on a real system. As practical issues play an important role, we have selected the minimization of the transmit power subject to a ZF criterion. This criterion is suboptimal, but the key point is that it provides a simple closed-form solution and guarantees that there is no interference among users belonging to the same group.

*This work is partially supported by the Telecommunications Technological Center of Catalonia (CTTC), by the Spanish government FIT-070000-2000-649 (Medea+ A105 UniLan), TIC99-0849 and the Integrated Action HF2001-0055 and by the Catalan government thanks to grants 2001SGR 00268, 2003FI 00190 and 2001FI 00714.

As reported in [3], the performance of a Spatial Division Multiple Access (SDMA) system is better when each antenna at the BS is assigned to a user rather than using all the antennas for transmission to a single user. On the other hand, several papers in the literature have dealt with the problem of spatial scheduling, but they do not take into account the spatial signature of the users. Therefore, the overall efficiency of the system is reduced because co-linear spatial signatures may be allocated in the same group. In this paper, the spatial signatures of the users are taken into account in the optimal group assignment for the scheduling.

In [4], the authors maximize the capacity of a SDMA/TDMA system. As in our work, their problem also resides on the best combination of users that optimizes a certain criterion, in their case the maximization of capacity. They propose a solution based on graph-theory, which is a NP-complete problem, so they point out the need for efficient heuristic algorithms.

In our paper, instead of the maximization of the capacity, we have chosen the minimization of transmit power as a cost function. In fact, we vote for a realistic physical parameter that yields to a reduction in the power consumption and that can be easily used in higher layers. On the other hand, we are taking into consideration a spatio-temporal scheduler that could be developed in real time, so exhaustive search methods or the solution in [4] based on graph-theory have been disregarded. Firstly, we implement SA, which obtains good near-optimal solutions with reasonable computing time. In fact, simulations have shown to obtain the optimum when the number of users is low. As a main contribution, we propose simple heuristic solutions that may enable the real-time operation of the scheduler with only a slight power degradation.

This paper is organized as follows. First, in Section 2 we describe our problem and find the global cost function to be minimized, together with the above-mentioned equivalence. Then, Section 3 deals with the proposed solutions to the problem, i.e. SA and the heuristic algorithms. After that, we give some simulations results in Section 4 and the conclusions.

2. PROBLEM STATEMENT

In the following, boldface capital (lowercase) letters refer to matrices (vectors). The operator $(\cdot)^*$ denotes conjugation, $(\cdot)^T$ transposition, and $(\cdot)^H = ((\cdot)^*)^T$. The element at row ith and column jth of matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{ij}$. We consider the downlink of a communications system, where the BS is provided with Q transmit antennas, although the notation and solutions presented henceforth can be directly applied to the uplink. Let $K \geq Q$ be the number of active users in the cell, each having a single antenna. The users are then distributed into $G = \lceil K/Q \rceil$ groups. Each group is scheduled for transmission in a different time slot, whereas the Q users in one group are served simultaneously by a SDMA scheme, i.e. a different beamformer for each user.

Motivated within the context of Orthogonal Frequency Division Multiplexing (OFDM) systems, such as WLANs[1], the channel is assumed flat fading. In those systems, the frequency-selective channel in the time domain is converted into a flat fading channel at each of the useful subcarriers, see [5] for details.

Our problem is the minimization of the total transmit power, subject to a ZF design criterion for each group. ZF implies that at group $g, 1 \leq g \leq G$, there is no interference among users and the same signal level is achieved for all of them. Let us first consider the signal model for this MISO link[2]. The received signal vector for the group at time instant n is $\mathbf{y}(n)$:

$$\mathbf{y}(n) = \sum_{k=1}^{Q} \mathbf{H}\mathbf{b}_k s_k(n) \in \mathbb{C}^{Q\times 1}, \quad (1)$$

where $\mathbf{b}_k$ is the transmit beamvector for the kth user and $\mathbf{H}$ is the $Q{\times}Q$ complex flat-fading channel matrix, the ith row ($1 \leq i \leq Q$) of which contains the $1 \times Q$ vector of the channel gains for the ith user, i.e. $\mathbf{h}_i$. At position $k, 1 \leq k \leq Q$, of vector $\mathbf{y}(n)$ we find the received signal for the user kth. The transmitted symbols have unitary mean energy $E\left\{|s_k(n)|^2\right\} = 1$.

As stated before, for each user we want to minimize the power of the beamvector according to a ZF criterion:

$$\min \quad \mathbf{b}_k^H \mathbf{b}_k \qquad s.t. \qquad \mathbf{H}\mathbf{b}_k = \mathbf{1}_k, \quad (2)$$

where the vector $\mathbf{1}_k$ is 1 at position kth and 0 elsewhere. It is straightforward to obtain the solution to this problem:

$$\mathbf{b}_k = \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1} \mathbf{1}_k \in \mathbb{C}^{Q\times 1}. \quad (3)$$

Applying an eigenvector decomposition to the hermitian matrix $\mathbf{H}\mathbf{H}^H = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$ [3], it is easy to verify that the power from the kth transmit beamvector reduces to:

$$\|\mathbf{b}_k\|^2 = \sum_{m=1}^{Q} \frac{1}{\lambda_m} \left|\mathbf{1}_k^H \mathbf{u}_m\right|^2, \quad (4)$$

where λ_m is the mth eigenvalue of matrix $\mathbf{H}\mathbf{H}^H$ and $\mathbf{u}_m$ is the normalized eigenvector associated with that eigenvalue (the mth column of matrix $\mathbf{U}$). If we sum up the powers of the beamvectors for the Q users in group, the total power can be expressed as:

$$\sum_{k=1}^{Q} \|\mathbf{b}_k\|^2 = \sum_{m=1}^{Q} \frac{1}{\lambda_m} = trace\left[\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\right]. \quad (5)$$

Inserting the index g denoting the group, the total transmit power P_t added up over all the groups is finally expressed as:

$$P_t = \sum_{g=1}^{G} trace\left[\left(\mathbf{H}(g)\mathbf{H}(g)^H\right)^{-1}\right]. \quad (6)$$

In order to minimize the global transmit power P_t in Eq. (6), we have to find the best distribution of the users in groups (among all alternatives). Indeed, this is a combinatorial problem. As it will be stated, the global transmit power may vary substantially depending on how the users are grouped together.

[1]The main current standards are Hiperlan/2 and IEEE 802.11a.

[2]Strictly speaking, we should also add the group index g. But for simplicity, this is omitted until Eq. (6).

[3]$\mathbf{\Sigma}$ is the diagonal matrix of the positive eigenvalues and the columns of the unitary matrix $\mathbf{U}$ contain the eigenvectors associated to them.

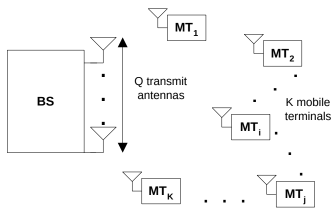

Fig. 1. System configuration: a BS with Q antennas aiming to transmit simultaneously to groups of Q users (mobile terminals).

2.1. Equivalent problem

We show here that the solution to Eq. (6) is equivalent to the minimization of the maximum BER of all users and groups. First, recall that an additional parameter α_k can be inserted in Eq. (2), which denotes the desired signal power at the receiver[4]. With this approach, the optimum solution is that minimizing the global BER of all the users in the cell, e.g. if the symbols are BPSK and there is no interference among users in the same group[5],

$$\min \quad \sum_{g=1}^{G}\sum_{k=1}^{Q} Q\left(\sqrt{2 \cdot SNR_{k,g}}\right), \quad (7)$$

where $SNR_{k,g} = \frac{|\alpha_{k,g}|^2}{\sigma^2}$, being σ^2 the noise power, is the SNR of user kth in group g. This is equivalent to the minimization of the 1-norm of the BERs for all users in all groups. As this problem is difficult to solve, we can choose the infinity-norm instead of the 1-norm. Therefore, Eq. (7) is now transformed into:

$$\min \quad \max_{g,k} Q\left(\sqrt{2 \cdot SNR_{k,g}}\right). \quad (8)$$

This problem is in fact the minimization of the maximum BER among all the users and groups. Intuitively, let us have an scenario where a user has a high SNR (thus very low BER) and the other users equally low SNR. Then, we can always increase the SNRs of *bad* users so that their BER is reduced. In the end, all the users will have the same SNR (and BER), therefore $|\alpha_{k,g}|^2 = \alpha, \forall k, g$, thus the ZF solution if we set $\alpha = 1$. Finally, note that the ZF approach does not guarantee the minimization of the mean BER.

3. PROPOSED METHODS

Several options have been evaluated so as to solve the combinatorial problem of Eq. (6). First, exhaustive search in the possibility space or the graph-based solution in [4], are prohibitively expensive in terms of complexity when the number of users increases. Instead, we have applied the stochastic method SA, which may yield to the optimum solution with a reasonable computing time. Moreover, we have developed some low complex heuristic algorithms that achieve a low degradation with respect to SA.

3.1. Simulated annealing

SA is an iterative algorithm which is capable of finding the global optimum of a function even if the problem is not convex. SA has analogies with the physical process of annealing, i.e. cooling of a

[4]The $\mathbf{1}_k$ vector will have now α_k at position kth.

[5]This condition is satisfi ed by the ZF solution.

Table 1. Simulated Annealing

1. Choose an initial solution $\mathbf{G}$. This $G \times Q$ matrix contains a user index at each position g, k, i.e. each row contains the Q users belonging to the same group.
2. Compute the global transmitted power $P(\mathbf{G})$ as in Eq. (6).
3. Initially, $P_{min} \leftarrow P(\mathbf{G})$ and $\mathbf{G}_{min} \leftarrow \mathbf{G}$.
4. Generate a new solution $\mathbf{G}'$ exchanging two users' positions in different groups.
5. As in Eq. (6), compute the power $P(\mathbf{G}')$ and the difference with the previous solution $\Delta(P) = P(\mathbf{G}') - P(\mathbf{G})$.
6. Accept $\mathbf{G}'$ with probability $\min\left(1, e^{\frac{-\Delta(P)}{T}}\right)$. If $\mathbf{G}'$ is accepted, $\mathbf{G} \leftarrow \mathbf{G}'$.
7. If $P(\mathbf{G}') \leq P_{min}$, then $P_{min} \leftarrow P(\mathbf{G}')$ and $\mathbf{G}_{min} \leftarrow \mathbf{G}'$.
8. If stopping condition not satisfied, decrease T and repeat 4.
9. Return the solution: $\mathbf{G}_{min}$ and P_{min}.

system [6]. The key parameter is the temperature T, which helps avoiding local minima, as it is explained below. The application of this algorithm to our problem is summarized in Table 1. Now, let us briefly describe the technique.

Given a current solution, a new one is proposed by exchanging two users belonging to different groups. If it is *better* than the previous one, it is accepted as the current solution, else, it is accepted with a certain probability. This mechanism is called *hill-climbing*, and avoids finding a local minimum. The parameter that controls the acceptance probability is the temperature T. The higher it is, the higher the acceptance probability. Therefore, T shall be lowered gradually, so that asymptotically only *better* solutions are accepted and a minimum is achieved. *Better* in our case means having less global transmitted power, see Eq. (6).

Initially, the temperature shall be high enough in order to accept most of the proposed solutions, in our case we increase T until the acceptance ratio is 95%. In this work, we run $N_{it} = 50$ iterations per value of temperature. After that, we run the algorithm until the acceptance ratio is lower than 0.05 for 5 times (stopping condition). After each run, we lower the temperature with an exponential profile $T \leftarrow \beta T, \beta = 0.9$. We have chosen a low β in order to speed up the algorithm.

Table 2. Initialization procedure for heuristic algorithms

1. Compute the cost of all possible combinations of two users. The cost of clustering together users i and j is:

$$[\mathbf{C}]_{ij} = trace\left[\left(\mathbf{H}_{ij}\mathbf{H}_{ij}^{H}\right)^{-1}\right], \qquad (9)$$

   where the matrix $\mathbf{H}_{ij}$ is defined as in Eq. (1), but in this case only for users i and j.
2. Sort the cost values of the matrix $\mathbf{C}$ in descending order in the vector $\mathbf{c}^s$, keeping the information of the index i, j in the matrix $\mathbf{C}$. Set l to 0.
3. Increase l. Select the lth position in vector $\mathbf{c}^s$. If the users i, j corresponding to the 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, repeat step 3.

Table 3. Heuristic algorithms

1. Build matrix $\mathbf{D}$ as in Eq. (10).
2. For the selected criterion, either *MAX-MIN* or *MAX-RATIO*, store the cost values in the vector $\mathbf{c}$ and keep the information about the user and also about the group that has the minimum cost for that user, i.e. the minimum of each row of $\mathbf{D}$.
3. Sort $\mathbf{c}$ is 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. Repeat 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. Repeat step 1 until every user is assigned to a group.

3.2. Heuristic techniques

In this section, we propose much simpler methods than exhaustive search or SA, which yield to suboptimal performances but reduce drastically the computational load. Basically, we propose two approaches: *MAX-MIN* and *MAX-RATIO*. Essentially, they separate the users that are close in terms of angle of arrival, the spatial signatures of which are very similar or highly co-linear, see Table 4. If we grouped these users together, the required power would increase substantially.

Their initialization procedure is described in Table 2. After that, we have the G users that are most difficult to separate associated to G different groups. We can then proceed in assigning remaining users to groups. Therefore, at each iteration of the algorithm we need to build the $K' \times G'$ matrix $\mathbf{D}$, where K' is the number of remaining users and G' the number of uncompleted groups. At the k'th row and g'th column we put the cost of assigning user index k' to group g', i.e.

$$[\mathbf{D}]_{k'g'} = trace\left[\left(\mathbf{H}(g')\mathbf{H}(g')^{H}\right)^{-1}\right], \qquad (10)$$

where $\mathbf{H}(g')$ is the $Q' \times Q$ channel matrix of group g', the last row of which corresponds to the channel of user k'. Note that $Q' - 1$ is the number of users belonging to group g' at that iteration.

The basic procedure for both algorithms is depicted in Table 3. They differ in the way the cost values are computed (step 2). Whereas the *MAX-MIN* approach selects the minimum cost of assigning a user to a group, i.e. the minimum of each row of $\mathbf{D}$, *MAX-RATIO* calculates the ratio between the maximum and minimum cost for each user, i.e. the maximum and minimum values at every row of matrix $\mathbf{D}$. We shall remark that at each iteration we assign at most G users, i.e. we do not assign two users to the same group nor fill groups that are not optimal for the users.

Now, let us briefly explain the ideas behind the selected criteria. The *MAX-MIN* approach selects the best-suited group for each user and it assigns then first those who are most difficult to cluster, i.e. those having a maximum value of the minimum cost. In terms of performance, it does not take into account that a user may not be assigned to its best-suited group, e.g. because it is already full. Then, a big degradation may occur.

Instead, the selection criterion shall be related with the dispersion of the cost values. We shall first allocate in the groups the users that may provoke a high penalty in terms of power if they are not assigned to the best group. This is the behavior of the

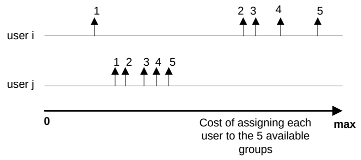

Fig. 2. Example of two users that have to be assigned to groups. *MAX-MIN* first allocates user j whereas *MAX-RATIO* selects user i. If user i is not assigned to group 1 a high degradation occurs, while if user j is not in group 1, the loss in performance is negligible. *MAX-RATIO* allocates first the users incurring in a big loss if they are not in the best group, thus achieving a better performance.

Group	**User angles** [degrees]		
1	-41.17	-1.11	29.82
2	-41.10	-10.15	33.78
3	12.92	-28.58	-57.43
4	14.85	-50.99	-20.66

Table 4. Spatial distribution of users. Single run, $G = 4, Q = 3$.

MAX-RATIO proposal: instead of selecting the users according to the minimum costs, we sort them by means of the ratio between the maximum and the minimum costs. Note that other approaches might also be possible as long as they consider the variation of the cost values. For an illustrative example of the behavior of the heuristic algorithms, see Fig. 2.

4. SIMULATIONS

$N_r = 10000$ runs of the simulations have been conducted for a number of transmit antennas ranging from $Q = 2$ to $Q = 4$, and a number of groups[6] from $G = 2$ to $G = 7$. The channel is flat fading, and static users are distributed uniformly between $-60°$ and $60°$ (sectored antennas at the BS). No noise is included in the simulations. The aim is to see what is the performance in terms of power for the techniques that have been previously described.

We have also simulated the random scheduler, i.e. the users are arbitrarily allocated to the groups. We see in Figs. 3 and 4 that this solution is not convenient at all as the output power[7] is at least one magnitude order worse than the *MAX-RATIO* solution. We also see that the *MAX-MIN* approach yields to bad performances due to the fact explained in the example (see Fig. 2): it does not take into account all the possible options for a user, i.e. it may not be assigned to the best-suited group, so this worsens the performance because of a high penalty in terms of power.

As expected, SA has the best performance. Note that the total power when G is low is greater than in the high range. This is due to the fact that, if we have very few users and they are close together, we cannot cope with their separation. On the other hand, when we increase G, we also increase the degrees of freedom of the system. This allows a better separation of the groups. In turn, the *MAX-RATIO* technique approaches the minimum power of SA. The degradation is mainly between 3 dB and 6 dB. The key advantage is that the the computational load is reduced drastically paying the price of a slight increase in the global transmit power.

Finally, we show in Table 4 the spatial distribution of users given by a single run of the simulation with the *MAX-RATIO* solution. It is seen that the users that are closer (in terms of angle of arrival) are separated in different groups. If we grouped them together, the increase in power would be remarkable.

[6]For simplicity, we assume K to be a multiple of Q.

[7]In those figures, we show the outage power for a certain threshold x. That means that the $x\%$ of the obtained powers are below the plotted value.

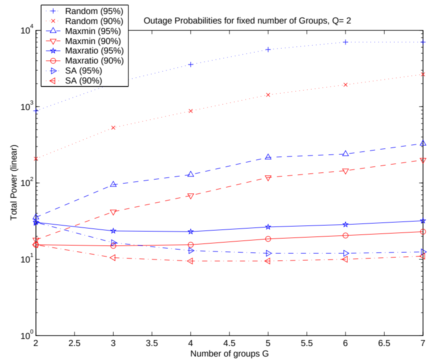

Fig. 3. Outage Powers (90% and 95%) for Q=2.

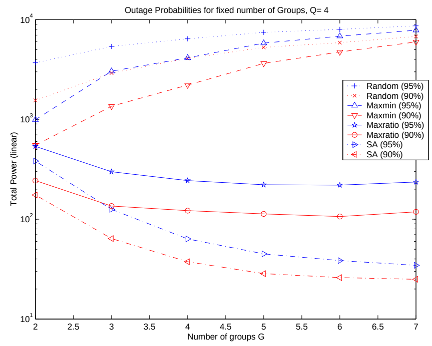

Fig. 4. Outage Powers (90% and 95%) for Q=4.

5. CONCLUSIONS

There are two main contributions of this paper. First, we deal with the spatial scheduling a set of users in groups for simultaneous service and propose a new joint MAC-PHY spatial channel allocation. Second, we focus on the techniques for the resolution of this NP-complete problem. Therefore, we apply the stochastic technique SA and propose suboptimal solutions that are shown to have a low degradation with respect to the SA performance with much lower computational load. As this is a first insight in the joint Physical and Medium Access Control (MAC-PHY) design, the next step is to take fairness issues [7] into consideration.

6. REFERENCES

[1] F. Rashid-Farrokhi et al., "Transmit Beamforming and Power Control for Cellular Wireless Systems," *IEEE JSAC*, Oct. 1998.

[2] R. Stridh and M. Bengtsson and B. Ottersten, "System Evaluation of Optimal Downlink Beamforming in Wireless Communication," in *VTC Fall*, 2001.

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

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

[5] L.J. Cimini, "Analysis and Simulation of a Digital Mobile Radio Channel using Orthogonal Frequency Division Multiplexing," *IEEE Transactions on Communications*, July 1985.

[6] P.J.M. van Laarhoven and E.H.L. Aarts, *Simulated Annealing: Theory and Applications*, Kluwer Academic Publishers, 1987.

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