PERFORMANCE ANALYSIS OF SCHEDULING AND ADMISSION CONTROL FOR MULTIUSER DOWNLINK SDMA

Diego Bartolomé
dbartolome@ieee.org
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
c/ Gran Capità 2-4, Planta 2, 08034 Barcelona (Spain)

Ana I. Pérez-Neira
anuska@gps.tsc.upc.es
Universitat Politècnica de Catalunya (UPC)
c/ Jordi Girona 1-3, Mòdul D5, 08034 Barcelona (Spain)

ABSTRACT

Multiple antennas are used here to enhance the scheduling task at a multi-antenna base station. After applying a zero forcing transmit beamforming, the scheduling shall distribute the resources among the users. Several criteria are presented and their performance in the high Signal to Noise Ratio (SNR) range is analyzed. The next issue is to take into account the SNR requirements from the users and perform the admission control accordingly. An algorithm that yields the optimum performance is proposed, together with a new criterion that falls between the optimization of the best global performance and the satisfaction of the individual needs. Simulation results are provided to show the performance of the techniques.

1. INTRODUCTION

We deal with the communication of a multi-antenna Base Station (BS) or Access Point (AP) with several single-antenna terminals. The spatial diversity is used here to improve the performance of the scheduling, i.e. the PHYsical layer (PHY) helps the Data Link Control (DLC) layer in this task. The goal of the scheduling is to assign a certain rate to the users, and to perform the admission control. At the PHY, the scheduler selects the users that are allowed to transmit, they are assigned a fraction of the scarce power, thus they are able to obtain a certain rate and Bit Error Rate (BER).

The solution of the scheduling at the PHY can be conceptually divided into two steps, namely the beamforming and the power allocation. Moreover, the admission control is performed if some minimum Signal to Noise Ratio (SNR) requirements shall be satisfied. As in [1], we assume dumb terminals, so that all the intelligence is located at the multi-antenna BS. Therefore, the BS applies a beamforming criterion that eliminates the inner-cell interference, so that the terminals receive their symbols multiplied by a channel gain and corrupted by additive Gaussian noise, which differs from the approach taken e.g. in [2]. With this, not only the computational needs at the terminals are relaxed and their battery life can be increased, but also the amount of signaling can be reduced.

The beamforming criterion that agrees with this viewpoint is Zero Forcing (ZF) [3]. Compared to optimum downlink beamforming, see e.g. [4], ZF provides not only a closed-form solution but also a good trade-off between performance and complexity. Moreover, the proposed strategy seems well suited for Spatial Division Multiple Access (SDMA), since the same idea holds for the multiple access in time (TDMA) or frequency (FDMA), where the resources granted for the users do not overlap.

This work is partially supported by the CTTC, by the European Union under IST-2003-507581, by the Spanish government under FIT-070000-2003-257, HF2001-0055, and TIC2002-04594-C02 (jointly financed by FEDER), and by the Catalan government thanks to grant 2003FI 00190.

After that, several criteria have been proposed for the distribution of the limited resource (power) among the users, see [5] and references therein. In this paper, we study their performance in the high SNR range, and identify useful equivalences and results. Since there is a trade-off between the individual satisfaction and the global optimization [6], the convenience of any technique has to be evaluated not only at the PHY level but also according to the needs of the DLC.

On the other hand, when Quality of Service (QoS) is required, the PHY scheduler shall decide which users are served according to their target SNR, i.e. the admission control mechanism. The interactions among the users are crucial, since users with correlated channels require more power to satisfy the same requirements. After identifying the constraints that shall be satisfied by each of the proposed schedulers, we outline an algorithm that yields the optimum performance. Between the best global performance and the fulfillment of the individual needs, we propose an alternative that aims at dividing proportionally among the users the loss with respect to their maximum achievable SNR. In any case, we deal with an instantaneous distribution of the resources, since if short term fairness is guaranteed, then long term fairness is also ensured.

In Section 2 we give an overview of the problem, present typical solutions and study their performance at high SNR. After that, we look into the admission control mechanism in Section 3, and then give some simulation results. Finally, conclusions are drawn.

2. SCHEDULING OVERVIEW

In the following, boldface capital (lowercase) letters refer to matrices (vectors). The conjugate transpose of $\mathbf{a}$ is $\mathbf{a}^H$ and the element at row ith and column jth of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. Unless indicated, the base-2 logarithm of a is $log(a)$, tr denotes the trace operation, and $diag(a_1, \ldots, a_K)$ refers to the $K \times K$ square matrix with diagonal elements given by $a_1, \ldots, a_K$. The $Q \times Q$ identity matrix is $\mathbf{I}_Q$, $a^+ = \max(a, 0)$, and $s.t.$ refers to $subject\ to$.

The Q-antenna BS communicates simultaneously with K single antenna terminals (users), which are gathered in the active set $\mathcal{K}$. At any time instant, the signal model for this SDMA system is

$$\mathbf{y} = \mathbf{HBs} + \mathbf{w} \in \mathbb{C}^{K\times 1}, \tag{1}$$

where the kth position of the vector $\mathbf{y}$ ($\mathbf{s}$) is the received (transmitted) signal for user k. The $K \times Q$ channel matrix $\mathbf{H}$ has i.i.d. complex Gaussian random entries with zero mean and unit variance. The beamforming matrix collects the K weight vectors for the users $\mathbf{B} = [\mathbf{b}_1\mathbf{b}_2 \ldots \mathbf{b}_K] \in \mathbb{C}^{Q\times K}$, and the noise is complex Gaussian, i.e. $\mathbf{w} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_Q\right)$.

In TDMA/TDD systems, estimates of the channel are obtained through an appropriate training sequence, thus the assumption of perfect channel knowledge might not be far from reality. As in [5], the ZF beamforming criterion is slightly modified in order to separate the effect of the equivalent channel gain α_k and the power allocation β_k. Therefore, the beamforming matrix becomes

$$\mathbf{B} = \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H\right)^{-1} \mathbf{D}_\alpha \mathbf{D}_\beta = \widetilde{\mathbf{B}}\mathbf{D}_\beta, \tag{2}$$

where $\mathbf{D}_\beta = diag\,(\beta_1, , \ldots, \beta_K)$ collects the power allocation factors, and $\mathbf{D}_\alpha = diag\,(\alpha_1, \ldots, \alpha_K)$ captures the effect of the equivalent channels $\alpha_k = 1/\sqrt{\left[(\mathbf{H}\mathbf{H}^H)^{-1}\right]_{k,k}}$ after the unitary beamforming $\widetilde{\mathbf{B}}$. With this model, the received signal for each user is only corrupted by noise and not by inner-cell interference, i.e.

$$y_k = \alpha_k \beta_k s_k + w_k, \forall k \in \mathcal{K}, \tag{3}$$

but the effect of the simultaneous service of several users is contained in α_k. Indeed, if the channels from the users are highly correlated, then α_k is low and more power is needed to fulfill the requirements of the active users. The SNR for user k is defined as

$$\gamma_k = \frac{|\alpha_k|^2 |\beta_k|^2}{\sigma^2}, \tag{4}$$

where we have assumed unitary mean energy symbols. Note that the maximum number of bits per symbol that can be transmitted reliably is $r_k = log\,(1 + \gamma_k)$. Moreover, since the scheduling methods presented next require an easy-differentiable BER expression, we use the approximation in [7], $\mathrm{BER}(\gamma) \approx c_1 e^{-c_2 \gamma}$, where c_1 and c_2 depend on the signal mapping.

2.1. Scheduling methods

Several options are proposed in the literature, either *equalizing* the performance of the users or optimizing the global performance of the cell at the expense of the users with worse channels. In this section, we review briefly the strategies in [5] so as to study their performance at high SNR. According to regulatory authorities, we specify a maximum instantaneous power (P_T).

2.1.1. Uniform Power Allocation (UPA)

Without any type of channel knowledge, the best option is the Uniform Power Allocation (UPA), i.e. the assignment of an equal proportion of the total available power among the K users. Assuming that the ratio $\frac{P_T}{\sigma^2}$ is denoted by γ^n, the SNR for the kth user is

$$\gamma_k^U = \frac{|\alpha_k|^2}{K} \frac{P_T}{\sigma^2} = \frac{|\alpha_k|^2}{K} \gamma^n. \tag{5}$$

2.1.2. Maximum Sum Rate (MSR)

If the transmitter has accurate estimates of the channel, a more elaborated strategy is the maximization of the sum rate of the cell:

$$\max_{|\beta_k|^2} \sum_{k \in \mathcal{K}} r_k \tag{6}$$

$$s.t. \sum_{k \in \mathcal{K}} |\beta_k|^2 \leq P_T, \tag{7}$$

which yields to a waterfilling. The power allocation is given by

$$|\beta_k|_M^2 = \left(\mu^{-1} - \frac{\sigma^2}{|\alpha_k|^2}\right)^+, \tag{8}$$

where μ is obtained imposing the constraint in (7). The SNR for the kth user γ_k^M is obtained by substituting these power factors in (4). In this case, the global performance is optimized regardless of some users, which might not even be allocated for transmission.

2.1.3. Equal Rate and BER (ERB)

Instead of seeking the best global performance, the BS might behave more fairly and try to serve equally all the K active users. The BS forces an equal SNR (also rate and BER) for all the users according to a max-min SNR criterion, which eventually degrades the cell performance. In fact, the BS uses more power for the weakest users. If we impose that the product $\alpha_k \beta_k$ is the same for all users, it is easy to see that the SNR for user k is obtained as

$$\gamma_k^E = \frac{\gamma^n}{tr\left[(\mathbf{H}\mathbf{H}^H)^{-1}\right]}, \forall k \in \mathcal{K}, \tag{9}$$

which yields the same service for all the users. This technique might also be expressed as a max-min rate or a min-max BER.

2.1.4. Minimum Sum BER (MSB)

Since the BER provides a direct link to the DLC because of the relation between the BER and the Packet Error Rate (PER), and thus the throughput, another possibility consists of Minimizing the Sum BER (MSB) of all the users in the cell, i.e.

$$\min_{|\beta_k|^2} \sum_{k \in \mathcal{K}} \mathrm{BER}_k \tag{10}$$

$$s.t. \sum_{k \in \mathcal{K}} |\beta_k|^2 \leq P_T, \tag{11}$$

which is solved using the BER approximation in [7]. This scheme is analogous to the MSR, therefore a waterfilling turns out again. The power allocation factors are given by

$$|\beta_k|_S^2 = \frac{\sigma^2}{c_2 |\alpha_k|^2} \left(log\left(c_1 c_2 \frac{|\alpha_k|^2}{\sigma^2}\right) - log\mu\right)^+, \tag{12}$$

where μ is obtained by imposing the constraint in (11), and the SNR of the kth user γ_k^S is again obtained by substituting these power allocation factors in (4).

2.2. High SNR Analysis

In this subsection, we give four important results concerning the performance of the previous methods in the high SNR range.

Result 1 If K and Q grow without bound, $K, Q \to \infty$, keeping their ratio $\zeta = Q/K > 1$ fixed, the power allocation for the MSR tends to be the same as for the UPA at high SNR.

First, we quote a result from [8]. If we let Q and K grow without bound, but their ratio $\zeta = Q/K$ remains fixed, then

$$\lim_{K,Q \to \infty} \mathbb{E}\left\{\frac{1}{|\alpha_k|^2}\right\} = \frac{1}{Q - K}. \tag{13}$$

Substituting (13) into (8) and applying the constraint in (7) to obtain μ, the power allocation factors for the MSR finally reduce to $\mathbb{E}\left\{|\beta_k|_M^2\right\} = \frac{P_T}{K}$. From another perspective, if the SNR is high, all the active users might be allocated for transmission and σ^2 is low. Then, the power allocation factors for the MSR are approximated as $|\beta_k|_M^2 \approx \mu^{-1}$, which reduce to $|\beta_k|_M^2 = \frac{P_T}{K}$ again by applying the constraint in (7).

Result 2 At high SNR, not only the MSR but also the UPA yield a higher sum rate than the ERB.

In the high SNR range, the sum rate of the ERB R_T^E can be approximated as

$$R_T^E \approx Klog\left(\frac{\gamma^n}{tr\left[(\mathbf{H}\mathbf{H}^H)^{-1}\right]}\right) = Klog\left(\frac{\gamma^n}{\sum_{k\in\mathcal{K}}\frac{1}{|\alpha_k|^2}}\right) = \\ = \quad C + KlogH_\alpha, \qquad (14)$$

where H_α is the harmonic mean of the $|\alpha_k|^2$ and C is a constant,

$$\frac{1}{H_\alpha} = \frac{1}{K}\sum_{k\in\mathcal{K}}\frac{1}{|\alpha_k|^2}, \qquad C = Klog\left(\frac{\gamma^n}{K}\right). \qquad (15)$$

On the other hand, sum rate of the UPA[1] R_T^U is approximately

$$R_T^U \approx \sum_{k\in\mathcal{K}} log\left(\frac{\gamma^n|\alpha_k|^2}{K}\right) = C + log\prod_{k\in\mathcal{K}}|\alpha_k|^2 = C + KlogG_\alpha, \qquad (16)$$

in the high SNR range, where $G_\alpha = \left(\prod_{k\in\mathcal{K}}|\alpha_k|^2\right)^{1/K}$ is the geometric mean of the $|\alpha_k|^2$. Since the geometric mean is always greater or equal than the harmonic mean, i.e. $G_\alpha \geq H_\alpha$, the sum rate of the UPA (and obviously the MSR) in (16) is always greater or equal to the sum rate of the ERB in (14). Equality holds if $\alpha_j = \alpha_k, \forall j, k$, which occurs with probability 0 in this system.

Result 3 At high SNR, the MSB is equivalent to the ERB.

If all the users in the active set $\mathcal{K}$ are served, $log\mu$ in (12) is

$$log\mu = \frac{\sum_{j\in\mathcal{K}}\frac{\sigma^2}{c_2|\alpha_j|^2}log\left(c_1c_2\frac{|\alpha_j|^2}{\sigma^2}\right) - P_T}{\sum_{j\in\mathcal{K}}\frac{\sigma^2}{c_2|\alpha_j|^2}}, \qquad (17)$$

in which at high SNR, the first term in the numerator tends to zero. The power allocation factors in (12) can then be approximated as

$$|\beta_k|_S^2 \approx \frac{\sigma^2}{c_2|\alpha_k|^2}\left(log\left(c_1c_2\frac{|\alpha_k|^2}{\sigma^2}\right) + \frac{P_T}{\sum_{j\in\mathcal{K}}\frac{\sigma^2}{c_2|\alpha_j|^2}}\right). \qquad (18)$$

Since at a high SNR, the linear term is much higher than the logarithmic term, these power allocation factors reduce to

$$|\beta_k|_S^2 \approx \frac{\sigma^2}{c_2|\alpha_k|^2}\frac{P_T}{\sum_{j\in\mathcal{K}}\frac{\sigma^2}{c_2|\alpha_j|^2}} = \frac{P_T}{|\alpha_k|^2}\frac{1}{\sum_{j\in\mathcal{K}}\frac{1}{|\alpha_j|^2}}. \qquad (19)$$

Since $\sum_{j\in\mathcal{K}}\frac{1}{|\alpha_j|^2} = tr\left[\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\right]$, the SNR for the kth user is the same as for the ERB, that is,

$$\gamma_k = \frac{|\alpha_k|^2|\beta_k|_S^2}{\sigma^2} = \frac{P_T}{\sigma^2}\frac{1}{tr\left[(\mathbf{H}\mathbf{H}^H)^{-1}\right]} = \frac{\gamma^n}{tr\left[(\mathbf{H}\mathbf{H}^H)^{-1}\right]}. \qquad (20)$$

Result 4 The diversity order of the UPA is Q-K+1, and the diversity order of the ERB can be approximated as Q-K.

The equivalent channel $|\alpha_k|^2$ is distributed as a Chi-square random variable with $2(Q-K+1)$ degrees of freedom, that is $|\alpha_k|^2 \sim \chi^2_{2(Q-K+1)}$. Since its mean value is $2(Q-K+1)\times 1/2$, the diversity order of the UPA is $Q-K+1$. On the other hand, the diversity order of the ERB is obtained using **Result 1** as $Q-K$.

[1] As stated in **Result 1**, the MSR coincides with the UPA at high SNR. However, note that the MSR provides always the highest sum rate because it is explicitly designed for the purpose.

Table 1. Spatial Admission Control Procedure

1. Set $\mathcal{K} = \{1, \ldots, K\}$.
2. Build matrix $\mathbf{H}$ for the users in $\mathcal{K}$, and compute $|\alpha_k|^2 = 1/\left[\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\right]_{k,k}, \forall k \in \mathcal{K}$.
3. According to scheduling mechanism, if the condition in (21) or (22) is satisfied, go to step 5.
4. Otherwise, select the worst user $k^s : \min_k|\alpha_k|^2$, which is eliminated from the active set, $\mathcal{K} = \mathcal{K} - k^s$. Go to step 2.
5. Perform the UPA or the ERB with the users in $\mathcal{K}$.

3. ADMISSION CONTROL

We have seen that the MSR tends to the UPA, and that the MSB tends to the ERB. Therefore, we focus now on the UPA and the ERB, and assume a high SNR. The main goal of the PHY scheduler is to reduce the amount of information that shall be processed by the traffic scheduler at the DLC. Particularly, the PHY scheduler performs the admission control. Due to the interactions in this SDMA system, a crucial point is which subset of users is served. This shall be decided taking into account the BER or rate requirements, which can be mapped into a target SNR γ^T. First, we provide the solvability conditions for the UPA and the ERB with SNR constraints, and then propose an admission control strategy with an intermediate solution both in global performance and in fairness.

3.1. Solvability for the UPA and the ERB with SNR constraints

The UPA requires each user to satisfy a certain constraint. Looking at the SNR of the UPA in (5), if the problem with SNR constraints is feasible, then the equivalent channel gains shall be

$$|\alpha_k|^2 \geq \frac{K\gamma^T}{\gamma^n}, \forall k \in \mathcal{K}. \qquad (21)$$

On the other hand, the constraint for the ERB is a function of the whole system, and does not depend on the individual users, see (9). Particularly with SNR constraints, a feasible solution exists if

$$tr\left[\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\right] \leq \frac{\gamma^n}{\gamma^T}. \qquad (22)$$

With these conditions, we summarize in Table 1 an algorithm that obtains the most efficient set of users. Very briefly, it tries to serve all the users in the active set $\mathcal{K}$. While a solution in agreement with the requirements cannot be obtained, the worst user is taken out from $\mathcal{K}$. At each iteration, the most penalizing user is *eliminated*. When the problem is feasible, the scheduling strategy (the ERB or the UPA here) is applied for the active set $\mathcal{K}$ (step 5).

3.2. Equal Proportional SNR (EPS)

Here, we propose a new algorithm for the scheduling and the admission control. It is based on the fact that the users might agree to loose the same proportion δ_k of their maximum achievable SNR, γ_k^a, which is obtained as if they were served alone in the cell, i.e. $\gamma_k^a = \gamma^n\|\mathbf{h}_k\|^2$. Note that the channel $\mathbf{h}_k$ is the kth row of the channel matrix $\mathbf{H}$. In fact, δ_k can be seen as the price paid for the collective satisfaction and could be computed according to the traffic requirements. It is obtained as

$$\delta_k = \frac{\gamma_k}{\gamma_k^a}, \forall k \in \mathcal{K}. \qquad (23)$$

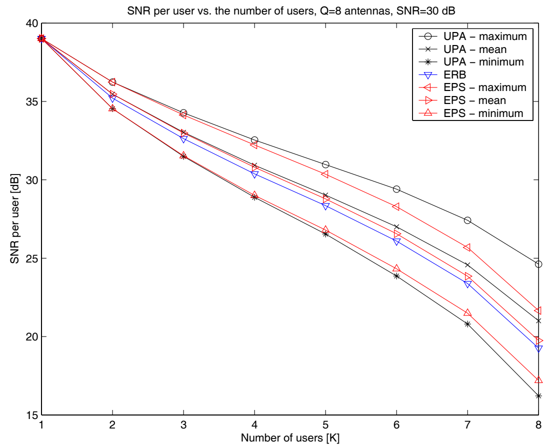

Fig. 1. Maximum, mean, and minimum output SNR as a function of the number of users for the scheduling mechanisms. For the ERB, the maximum, mean, and minimum coincide.

If all the users are homogeneous and allow the same loss in proportion to their maximum SNR, i.e. $\delta_k = \delta, \forall k \in \mathcal{K}$, the cost function of this problem is expressed as

$$\max \delta \tag{24}$$

$$s.t. \sum_{k \in \mathcal{K}} |\beta_k|^2 \leq P_T, \tag{25}$$

which has a closed-form solution for δ,

$$\delta^{-1} = \sum_{k \in \mathcal{K}} \frac{\|\mathbf{h}_k\|^2}{|\alpha_k|^2}, \tag{26}$$

and yields a SNR for the kth user given by $\gamma_k^P = \delta \gamma_k^a$. With BER constraints, the EPS can also be implemented according to Table 1. The difference is that the solvability condition is now given by $\gamma_k^P \geq \gamma^T, \forall k \in \mathcal{K}$. The suitability of this proposal is shown next.

4. SIMULATIONS

We take $\gamma^n = 30$ dB and a BS provided with $Q = 8$ antennas. First, we simulate the SNR performance of the UPA, the ERB, and the EPS without SNR constraints as a function of the number of users in the cell. We evaluate in Fig. 1 how the SNR degrades as the number of users increases. In order to show the differences among users, we plot in Fig. 1 not only the mean values, but also the mean value from the maximum and minimum SNR per user at each trial. It is shown that the mean SNR is maximized by the UPA, whereas the ERB provides a lower mean SNR, which coincides with its maximum and minimum SNR. As seen in (9), the ERB produces no dispersion among the SNR from the users. Finally, the EPS reduces the asymmetries among users of the UPA and provides a higher mean SNR than the ERB.

Besides, we simulate the admission control, provided that the users shall fulfill a SNR requirement. In Fig 2, we plot the mean number of served users as a function of the SNR requirement, which is the same for all the users. It is stated that the UPA gives service to the lowest number of users so as to improve the global performance by not serving the poorer users. On the other hand, the ERB serves the highest number of users, but the global performance is penalized, as it has been seen before. Finally, our EPS strategy provides an intermediate solution between them.

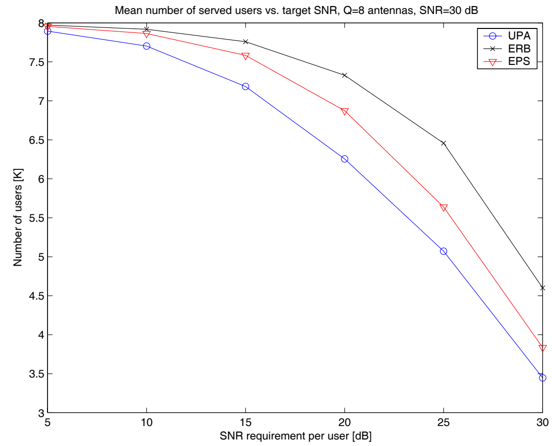

Fig. 2. Number of served users as a function of the SNR requirements for the admission control mechanisms.

5. CONCLUSIONS

In this paper, we have studied several options for the scheduling at a multi-antenna BS, either concentrating on the optimization of the global performance or on the fulfillment of the individual needs. The criteria have been analyzed in the high SNR range, providing useful insights into their performance. Furthermore, admission control mechanisms for the two limiting schedulers have been presented, and a new intermediate equal proportional SNR solution is developed. It yields the best balance of the trade-off between the global performance and the individual needs.

6. REFERENCES

[1] P. Viswanath, D.N.C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[2] I. Koutsopoulos and L. Tassiulas, "Adaptive Resource Allocation in SDMA-based Wireless Broadband Networks with OFDM Signaling," *in Proc. of IEEE INFOCOM*, June 2002.

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

[4] F. Rashid-Farrokhi, K.J. Ray Liu, and L. Tassiulas, "Transmit Beamforming and Power Control for Cellular Wireless Systems," *IEEE J. Select. Areas Commun.*, October 1998.

[5] D. Bartolome and A.I. Perez-Neira, "Spatial Scheduling in Multiuser Wireless Systems: From Power Allocation to Admission Control," *submitted to IEEE Trans. on Wireless Commun.*, October 2003.

[6] S. Borst and P. Whiting, "The Use of Diversity Antennas in High-Speed Wireless Systems: Capacity Gains, Fairness Issues, Multi-User Scheduling," *Bell Labs Tech. Mem.*, 2001.

[7] S.T. Chung and A.J. Goldsmith, "Degrees of Freedom in Adaptive Modulation: A Unified View," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, September 2001.

[8] B. Hochwald and S.Wishwanath, "Space-Time Multiple Access: Linear Growth in the Sum Rate," *in Proc. of the Allerton Conf. on Commun., Control and Computing*, October 2002.