How Much Feedback is Multi-User Diversity Really Worth?

David Gesbert (1), Mohamed-Slim Alouini (2)

(1) Eurecom Institute, Sophia-Antipolis, France. Email: gesbert@eurecom.fr
(2) Dept. of ECE, University of Minnesota, USA. Email:alouini@ece.umn.edu

Abstract – Wireless scheduling algorithms can extract multi-user diversity (MUDiv) via prioritizing the users with best current channel conditions. One drawback of MUDiv is the required feedback carrying the instantaneous channel rates from from all active subscribers to the access point/base station. This paper shows that this feedback load is, for the most part, unjustified. To alleviate this problem, we propose a technique allowing to dramatically reduce the feedback (by up to 90%) needs while preserving the essential of the scheme performance. We provide a theoretical analysis of the feedback load as function of the system's ergodic and outage capacity for both the traditional MUDiv scheme and the new scheme.

1. INTRODUCTION

Multi-user diversity was introduced first by Knopp and Humblet [1], then recently extended by Tse [2, 3], as a means to provide diversity against channel fading in multi-user packet-switched wireless networks. Assuming a reasonably large number of users are actively transmitting/receiving packets in a given cell, and assuming these users experience independent time-varying fading conditions, the network scheduler, within this cell, extracts a diversity gain by granting access to the channel, at any given slot period, only to those users which are close to a peak in terms of transmission quality. The quality is typically expressed as the instantaneous capacity, function of the signal-to-noise ratio. In the case of stationary users, channel time selectivity can be artificially introduced through random antenna combining [4]. Interestingly, MUDiv is often seen as a competing approach to MIMO diversity (e.g. space time codes) with MUDiv having the advantage of riding on the users's SNR peaks rather than solely eliminating the SNR fades.

In order to maintain throughput fairness across the users, the scheduler may resort to one of several possible resource allocation techniques, including the Proportional Fair Scheduling (PFS) algorithm [2, 3].

In order to manage the priorities among the subscribers, the scheduler requires in theory the channel quality information of all users at all times. In non-reciprocal wireless links (in fact most systems today, e.g. FDD networks such as 3G-FDD) this information must be fed back regularly by all users and as often as the channel changes (up to 200HZ for vehicular applications) via dedicated or contention-based uplink channels. In either case, the spectrum resource that must be provisioned to carry this amount of feedback with an acceptable error and latency performance makes true MUDiv hardly practical when the number of simultaneously active users becomes high [5].

In this paper we revist the concept of MUDiv and ask ourselves the following questions:

1. How much feeback is the capacity gain given by MUDiv really worth?

2. Can we reduce the amount of feedback and still preserve the scheduler performance?

To address those points among others, we make the following two key points: i) The only information that the scheduler/link adaptation layer truly needs is the channel quality of the user *to be scheduled*, and ii) thanks to the multi-user diversity, the user to be scheduled is bound to have a "good enough" channel, thus giving the opportunity to greatly limit the range of feedback. This idea was independently mentioned in [6] but not exploited for the same purpose. To exploit the ideas above we propose:

- A new and simple scheduling technique, referred to as *Selective Multi-user Diversity (SMUD) scheduling* in which each user compares its channel quality to a threshold only those who fall above it are allowed to request access and feedback their achievable downlink transmission rate, the others remaining silent.

- The concept of *scheduling outage* whereby no user reaches the required threshold therefore leaving the scheduler with no channel information from any users.

Among other contributions, we also:

- optimize the threshold to attain a prescribed level of scheduling outage.

- analyze the network feedback load (average and standard deviation) as function of the chosen threshold.

- Finally, we study the system average capacity and outage capacity as function of the feedback load and show that a dramatic reduction in the feedback is possible while preserving most of the capacity performance.

The optimization and evaluation is done analytically and verified with our simulations over Rayleigh fading channels.

2. SYSTEM AND SIGNAL MODELS

We consider a single interference-free cell with K simultaneously active users served by one access point (AP). The scheduling is organized on a slot by slot basis, i.e. one and only one user may access the channel during any given slot. For the sake of exposition, we consider here a SISO system. However, a MIMO extension can be easily devised.

2.1. Signal-to-Noise Ratio

Each user k, $k = 1..K$ is experiencing a fading link to/from the AP, with signal-to-noise ratio during slot s given by $\gamma_k(s)$ where $\gamma_k(s)$ is the SNR of a Rayleigh fading channel. The average SNR for user k is given by: $\bar{\gamma}_k = E(\gamma_k(s))$,

2.2. Transmission Rate

For a SISO system, the Shannon capacity achievable by user k over time slot s, if that user was selected for transmission, is simply:

$$C(k, s) = \log_2(1 + \gamma_k(s)). \tag{1}$$

In what follows we follow the work of [3] in using the capacity in (1) as a measure of the transmission rate used by the scheduler to select the transmission format to that user. This is clearly an approximation over the more practical case where the system actually selects one out of a finite set of discrete modulation/coding modes for transmission [5]. However the algorithms presented here also extend to the discrete modulation case.

3. MUDIV AND SCHEDULING ALGORITHMS

3.1. Proportional Fair Scheduling

In a multi-user diversity framework, the scheduler grants access to the channel to the user which experiences the best relative conditions in terms of rate at any given slot s. We assume all K users are active and have non empty queues of packets to deliver. To maintain fairness over a given finite time horizon or in the case of unequal average SNR conditions among users (due e.g. to distance to the AP), the scheduler must exploit a metric that takes into account the accumulated throughput up to slot s. An example of this is given by the PFS technique below [2, 3]:

At slot s, we schedule user $k_*(s)$ with maximum normalized capacity, i.e. such that

$$k_*(s) = \operatorname{argmax}_{k=1..K}\{\frac{C(k, s)}{R(k, s)}\}, \tag{2}$$

where $R(k, s)$ is the actual transmission throughput of user k over the link up to slot s. The throughputs are updated on a per slot basis according to:

$$R(k, s + 1) = R(k, s)(1 - 1/t_c) \quad k \neq k_* \tag{3}$$
$$R(k_*, s + 1) = R(k_*, s) + C(k_*, s)/t_c, \tag{4}$$

where t_c is a time constant adjusted to maintain fairness over a pre-determined time horizon, i.e. The larger t_c, the longer the horizon, the less stringent the fairness constraint.

3.2. Max SNR Scheduling

Importantly, when all users experience the same SNR distribution, a simpler scheduler giving access to the user such that:

$$k_*(s) = \operatorname{argmax}_{k=1..K}\{C(k, s)\} = \operatorname{argmax}_{k=1..K}\{\gamma_k(s)\} \tag{5}$$

will, by symmetry, also maintain fairness over a "long enough" horizon. In that case, we have:

Lemma: *The PFS and max SNR algorithms are equivalent for large t_c.*

Proof: Under ergodicity of the channel, $R(k, s)$ is equal to the average throughput of user k over a window t_c. If all users have the same SNR statistics, then:

$$R(k, s) = R(l, s) \; \forall k, l \; \text{ for } t_c \to \infty$$

Under this condition, eqns. (2) and (5) are identical.

In what follows we exploit this result and use the max SNR strategy in our derivations. Simulations illustrate robustness with respect to this approximation for finite t_c.

4. SELECTIVE MULTI-USER DIVERSITY

4.1. SNR thresholding

Because the user to be scheduled for transmission is the one with best relative channel conditions it is unlikely that, for a reasonable total number of users K, a user with bad relative signal-to-noise ratio will be selected by the scheduler. Therefore, the feedback resource provisioned for a such a user is wasted bandwidth. In fact, the intuition is that only those users with good enough conditions have a decent chances of being selected and should feedback their channel quality to the AP. To exploit this idea, we propose to consider a subset of users, coined *feedback users* by only considering those for which the channel quality is greater than a prescribed threshold, resulting in what is referred to here as Selective Multi-User Diversity (SMUD). Importantly, in the SMUD framework, users *decide locally* whether they should attempt to access the channel and send feedback to the AP or not. In the negative case, they remain silent for that slot.

In the particular case, considered from now on, where all users have the same average SNR $\bar{\gamma}_k = \bar{\gamma}$ for all k, we may for instance define the threshold in terms of the instantaneous SNR γ_{th}. :

At slot s, user k will feedback its channel quality to the AP if and only if

$$\text{Feedback condition: } \gamma_k(s) \geq \gamma_{th}.$$

Note that in this case, the thresholding is applied to the SNR. Clearly it could also be applied to other quality metrics such as the capacity, or normalized quality metrics when not all users have same average SNR, such as the normalized capacity. However the philosophy and key trade-offs offered by the approach remain the same there.

4.2. Scheduling outage

Let $P(s)$ be the number of feedback users at slot s, defined by $P(s) = \text{card}\{k, \text{ such that } \gamma_k(s) \geq \gamma_{th}\}$, where card is the cardinal operator. When $P(s) > 0$, the scheduler performs the selection as in (2) or (5) but this time within the set of feedback users only. The throughput is updated normally as in (3),(4).

In the opposite case, $P(s) = 0$ and no user feeds back any information to the AP, in which case we declare a *scheduling outage*. In the event of a scheduling outage, the scheduler may revert to i) a conventional 'blind' fair selection mode (e.g. round robin, random pick etc..) or ii) assume the previous best user remains optimal, given some finite coherence time for the channel. In this paper we limit ourselves to the worst-case scenario i), while the opportunistic strategy described in ii) is considered in a separate publication [7]. In option i), The AP polls the selected user for its achievable rate, allowing the link adaptaion algorithm to continue, however the blind scheduling mode deprives the system from any multi-user diversity in *that particular slot alone*.

5. PERFORMANCE ANALYSIS OF SELECTIVE MU DIVERSITY

In what follows we proceed with some analytical derivations of key performance metrics in view of the proper choice of the threshold parameter.

Importantly, we do not deal here with the actual multi-access access design for the feedback channel. We assume that the feedback channel is sized properly so that the feedback collision probability (in case a contention channel is used for the feedback of rates) is small enough to be ignored. We justify this assumption later in 5.3.2.

5.1. System Capacity

In this subsection, we first characterize the statistics of the post-scheduling SNR in terms of its cumulative distributive function (CDF) and probability density function (PDF). We then use these results to obtain the system outage capacity then the average system capacity.

5.1.1. CDF and PDF of SNR

For large values of the time constant and for the case $\bar{\gamma}_k = \bar{\gamma}$, the PFS becomes equivalent to scheduling the user with maximum SNR $\gamma_k(s)$. Hence we approximate $\gamma_{k_*}(s)$ by:

$$\gamma_{k_*}(s) = \max\{\gamma_1(s), \gamma_2(s), .., \gamma_K(s)\} \text{ if } P(s) > 0$$
$$\gamma_{k_*}(s) = \text{rand}\{\gamma_1(s), \gamma_2(s), .., \gamma_K(s)\} \text{ if } P(s) = 0$$

where rand is a random pick.

For simplicity of notation, let γ^* denote the SNR post scheduling. The CDF of γ^*, denoted by $P_{\gamma^*}(\gamma)$ is by definition the probability that γ^* falls below γ. Two cases have to be considered. First when $\gamma \leq \gamma_{th}$ then $P_{\gamma^*}(\gamma)$ is equal to the probability that all users SNRs are below γ_{th} and that the randomly picked user SNR is below γ. Mathematically speaking, we can then write assuming that the users experience independent identically distributed (i.i.d) fading

$$P_{\gamma^*}(\gamma) = (P_\gamma(\gamma_{th}))^{K-1} \; P_\gamma(\gamma), \qquad \gamma \leq \gamma_{th}, \qquad (6)$$

where $P_\gamma(\gamma)$ is the CDF of the users' SNR and is for example given by $P_\gamma(\gamma) = 1 - e^{-\gamma/\bar{\gamma}}$ in the Rayleigh fading case. For the second case (i.e., $\gamma > \gamma_{th}$), $P_{\gamma^*}(\gamma)$ is given by

$$P_{\gamma^*}(\gamma) = \sum_{k=0}^{K} \left(\begin{array}{c} K \\ k \end{array} \right) \; (P_\gamma(\gamma_{th}))^{K-k} \; (P_\gamma(\gamma) - P_\gamma(\gamma_{th}))^k . \tag{7}$$

Taking the derivative of $P_{\gamma^*}(\gamma)$ in (6) and (7) with respect to γ, we obtain the PDF of γ^*, $p_{\gamma^*}(\gamma)$ as

$$p_{\gamma^*}(\gamma) = (P_\gamma(\gamma_{th}))^{K-1} \; p_\gamma(\gamma), \qquad \gamma \leq \gamma_{th}$$
$$p_{\gamma^*}(\gamma) = \sum_{k=1}^{K} \left(\begin{array}{c} K \\ k \end{array} \right) \; (P_\gamma(\gamma_{th}))^{K-k} \; k \; p_\gamma(\gamma)$$
$$\times (P_\gamma(\gamma) - P_\gamma(\gamma_{th}))^{k-1} , \qquad \gamma > \gamma_{th}, \quad (8)$$

where $p_\gamma(\gamma)$ is the PDF of the SNR, given by $p_\gamma(\gamma) = \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}}$ in the Rayleigh fading case.

5.1.2. Outage Capacity

The outage capacity O_c corresponds to the probability that system capacity falls below a predetermined threshold C_T. We can then write

$$O_c = \text{Prob}[\log_2(1 + \gamma*) \leq C_T] = \text{Prob}[\gamma* \leq 2^{C_T} - 1] \quad (9)$$

which is the CDF of $\gamma*$ evaluated at $2^{C_T} - 1$. Hence we have

$$O_c = P_{\gamma^*}\left(2^{C_T} - 1\right), \tag{10}$$

5.1.3. Average Capacity

The system average capacity is given by

$$E(C_*) = E(\log_2(1 + \gamma*)) = \int_0^\infty \log_2(1 + \gamma*) \; p_{\gamma^*}(\gamma*) \; d\gamma*,$$

where $p_{\gamma^*}(\cdot)$ was obtained in (8). For the Rayleigh fading case, it can be shown using integration by part, the binomial expansion, and equations [8, Eqs. (3.352.1) and (3.352.2)], that the average system capacity is expressible in terms of first order exponential integral functions $E1(x) = \int_1^\infty e^{-xt}/t \; dt$ as [7]:

$$E(C_*) = \log_2(e) \left(1 - e^{-\gamma_{th}/\bar{\gamma}} \right)^{K-1}$$
$$\left[e^{1/\bar{\gamma}} \left(E_1 \left(\frac{1}{\bar{\gamma}} \right) - E_1 \left(\frac{1 + \gamma_{th}}{\bar{\gamma}} \right) \right) - e^{-\gamma_{th}/\bar{\gamma}} \; \ln(1 + \gamma_{th}) \right]$$
$$+ \log_2(e) \sum_{k=1}^{K} \left(\begin{array}{c} K \\ k \end{array} \right) k \left(1 - e^{-\gamma_{th}/\bar{\gamma}} \right)^{K-k}$$
$$\sum_{n=0}^{k-1} \left(\begin{array}{c} k-1 \\ n \end{array} \right) (-1)^n \; \frac{e^{(-k+1+n)\gamma_{th}/\bar{\gamma}}}{n+1}$$
$$\left[e^{-(n+1)\gamma_{th}/\bar{\gamma}} \; \ln(1 + \gamma_{th}) + e^{\frac{n+1}{\bar{\gamma}}} \; E_1 \left(\frac{(n+1)(1 + \gamma_{th})}{\bar{\gamma}} \right) \right].$$
$$(11)$$

5.2. Scheduling Outage Probability

Here we give an expression for $P_o = \text{Prob}(P(s) = 0)$. This event corresponds to the probability that all users fail to exceed the predetermined threshold γ_{th}, i.e.,

$$P_o = \text{Prob}(\gamma_k(s) < \gamma_{th}, \text{ for all } k = 1 \cdots K).$$

Assuming again that all users experience i.i.d. Rayleigh fading with the same average SNR $\bar{\gamma}$ then we have

$$P_o = (P_\gamma(\gamma_{th}))^K = \left(1 - e^{-\gamma_{th}/\bar{\gamma}}\right)^K. \tag{12}$$

5.3. Feedback Load

We are interested in quantifying the reduction in the feedback load obtained by adopting the SMUD scheme instead of the classical full feedback MUDiv algorithm. While the load is fixed equal to the number of users K with the latter scheme, it can range anywhere from zero to K at each slot time with the new scheme. As such we look in this subsection into the statistics of the feedack load with SMUD.

One interesting result is obtained, showing that the feedback load actually converges to a fixed deterministic (lower) value in the large K region, helping the provisioning of the feedback channel in practice.

5.3.1. Normalized Average Feedback Load

The normalized average feedback load $\bar{F}$ is defined as the ratio of the average load per time slot by the total number of users K. Mathematically this can be simply written as

$$\bar{F} = \frac{E(P(s))}{K}. \tag{13}$$

The conditional probability that k out of K users are preselected during a particular time slot is equal to the conditional probability that the SNRs of these k users equal or exceed the threshold γ_{th} (and of course the remaining $K-k$ users SNRs do not), i.e. $(1 - P_\gamma(\gamma_{th}))^k (P_\gamma(\gamma_{th}))^{K-k}$. Thus, for i.i.d. fading among users, the probability that $P(s) = k$ during time slot s is equal to

$$\text{Prob}(P(s) = k) = \binom{K}{k} (1 - P_\gamma(\gamma_{th}))^k (P_\gamma(\gamma_{th}))^{K-k}. \tag{14}$$

Therefore the normalized average load is given in these conditions by

$$\bar{F} = \frac{1}{K} \sum_{k=0}^{K} k \binom{K}{k} (1 - P_\gamma(\gamma_{th}))^k (P_\gamma(\gamma_{th}))^{K-k}., \tag{15}$$

which, for Rayleigh fading, simplifies to

$$\bar{F} = 1 - P_\gamma(\gamma_{th}) = e^{-\gamma_{th}/\bar{\gamma}} \tag{16}$$

in the Rayleigh fading case. We can also easily express the feedback load in terms of the scheduling outage and the number of users (for large K):

$$\bar{F} \approx \frac{-\log P_o}{K} \tag{17}$$

5.3.2. Feedback Load Variation and collisions

The fluctuation in the feedback load from one slot to another can be quantified by looking at the coefficient of variation in the number of users $P(s)$ feeding back their SNR, V_F and defined as

$$V_F = \frac{\text{Var}(P(s))}{(E(P(s)))^2}, \tag{18}$$

where $\text{Var}(\cdot)$ is the variance operator. Following similar steps than the ones developed in the previous subsection, it can be shown that [7]

$$\text{Var}(P(s)) = K (1 - P_\gamma(\gamma_{th})) P_\gamma(\gamma_{th}), \tag{19}$$

which, for Rayleigh, results into this coefficient of variation:

$$V_F = \frac{P_\gamma(\gamma_{th})}{K (1 - P_\gamma(\gamma_{th}))} = \frac{1}{K} \left(e^{\gamma_{th}/\bar{\gamma}} - 1\right), \tag{20}$$

This means that in a scenario where the threshold is fixed in advance (function of the average SNR) the feedback load tends to stabilize around the mean as the number of users increases. This is important from a provisioning point of view because this means the resource devoted to the feedback of channel rates can be sized in advance so the probability of collision in the feedback channel reaches a (small) desired value. In principle the probability of collision can be made as small as needed for the no-collision aproximation made in the analysis above effectively holds.

5.4. Threshold Choice

Various strategies are possible toward optimizing the threshold γ_{th}, including choosing it to reach a predetermined scheduling outage probability P_o. For instance in i.i.d. fading environment, inverting (12) leads to

$$\gamma_{th} = -\bar{\gamma} \ln\left(1 - P_o^{1/K}\right).$$

If the size of the feeback channel pipe is limited, it may be also of interest to choose the threshold γ_{th} in order to meet a certain normalized average feedback load specification. In i.i.d Rayleigh fading this can be obtained from (16) as $\gamma_{th} = -\bar{\gamma} \ln(\bar{F})$. In the next section, we show among others how the capacity is affected by both the threshold and the feeback load.

6. SIMULATIONS AND CONCLUSIONS

We simulate the performance of the SMUD scheme and compare it with a full feedback scheme, both using the PFS algorithm with $t_c = 500$ (slots). We compare with the analytical results which are obtained under the approximately equivalent max SNR scheduler. We use the same average SNR of 5dB for all users. Fig. 1 shows the average system capacity versus number of users for the SMUD scheme with γ_{th} given successively by (from top down): 0dB, 3dB, 6dB, 9dB, 12dB, 15dB, 18dB. We give the performance of the full

feedback MUDiv scheme for comparison. Relatively little is lost in performance below 9dB threshold. The analytical results are shown in dashed curves, showing the quality of the prediction.

Fig. 2 shows the feedback load ($\bar{F}$) as function the threshold γ_{th}. At 9dB theshold, the load is less than 10% of what it is with the original PFS algorithm.

Fig. 3 shows the scheduling outage probability vs. γ_{th}, for different values of K. With 28 active users and $\gamma_{th} = 9dB$ the SMUD scheme is very close to the full feedback scheme yet incurs only 10% of the feedback load, for an outage probability of less than 10%.

To summarize, in Fig. 4, we show the system capacity in relation to the required feedback load, for various number of users. This confirms that, for K above 25 or so, a feeback load greater than 10% results in very little aditional gain in terms of capacity and is therefore unnecessary.
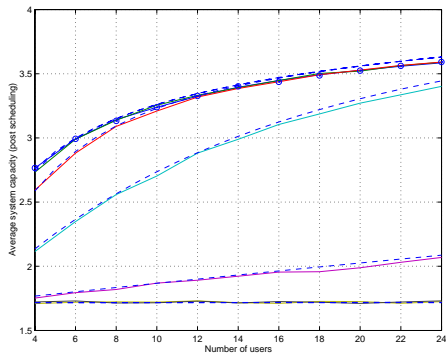

Figure 1: Average system capacity vs. number of users K for the SMUD scheme, for varions thresholds $\gamma_{th} = 0dB, 3dB, 6dB, 9dB, 12dB, 15dB, 18dB$ (from top down). 5dB average SNR. Dotted is the full feedback PFS algorithm. Dashed is the theoretical result.
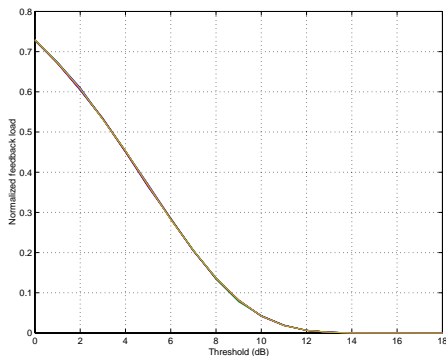

Figure 2: Average feedback load vs. threshold γ_{th} for the SMUD scheme. 5dB average SNR. Load for the full feedback PFS algorithm is 1. Dashed (superposed) is the theoretical result.
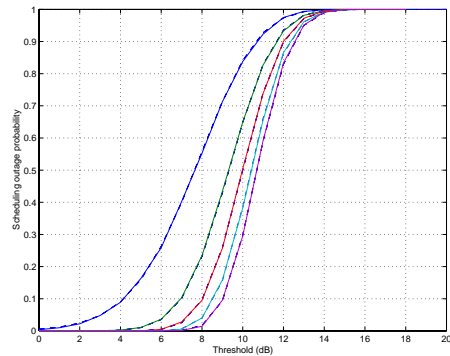

Figure 3: Scheduling outage vs. threshold γ_{th} for the SMUD scheme, for $K = 4, 10, 16, 22, 28$ users (left to right). 5dB average SNR. Dashed (superposed) is the theoretical result.
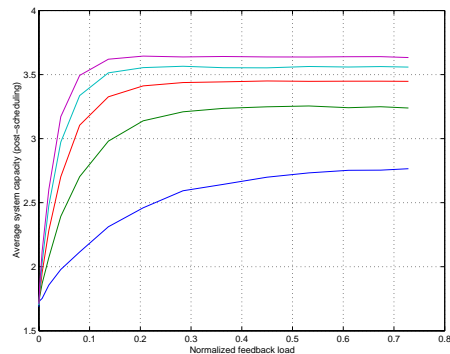

Figure 4: Average system capacity vs. required feedback load for the SMUD scheme, for $K = 4, 10, 16, 22, 28$ (bottom to top). 5dB average SNR.

[2] J. M. Holtzman, "CDMA forward link waterfilling power control," in *IEEE VTC Spring*, pp. 1663–1667, 2000.

[3] D. Tse, "Multiuser diversity and proportional fair scheduling," *In preparation*, 2002.

[4] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, June 2002.

[5] S. Catreux, D. Gesbert, V. Erceg, and R. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Communications Magazine*, June 2002.

[6] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channel with partial side information," *Submitted to IEEE Trans. on Information Theory*, 2003.

[7] D. Gesbert and M.-S. Alouini, "How much feedback is Multi-User Diversity really worth?," *IEEE Transactions on Wireless Communications, in preparation*, 2003.

[8] I. S. Gradshteyn and I. M. Ryzhik, eds., *Table of Integrals, Series, and Products*. 1994.

REFERENCES

[1] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications ," in *Proc. of Int. Conf. on Communications*, June 1995.