

# Adaptive Feedback Rate Control in MIMO Broadcast Systems with User Scheduling

Randa Zakhour  
Mobile Communications Department  
Institut Eurecom  
06560 Sophia Antipolis, France  
zakhour@eurecom.fr

David Gesbert  
Mobile Communications Department  
Institut Eurecom  
06560 Sophia Antipolis, France  
gesbert@eurecom.fr

**Abstract**— In [1], an adaptive scheme was introduced in view of optimizing the overall spectral efficiency of a multiuser MIMO wireless broadcast channel where the channel state information at the transmitting base station (CSIT), to be used for user scheduling and beamforming, is acquired over a limited-rate feedback channel. In this scheme, the feedback rate is no longer constant per scheduling period but rather optimized as a function of the time-dependent channel quality seen at the user side. The present paper further refines this idea and elaborates on some of the associated practical concerns.

## I. INTRODUCTION

CSIT plays an essential role in a MIMO broadcast channel, particularly when receivers have fewer antennas than the transmitter, i.e. are incapable of eliminating or even significantly reducing the interference due to the signals destined for others. CSIT makes it possible to achieve the multiplexing gain made possible by the multiple antennas at the transmitter, by adapting the transmission to the selected receivers' channels. Further rate enhancements may also be achieved by appropriately selecting spatially compatible users to serve, thus benefiting from multi-user diversity (MUD) [2].

However, acquiring said CSIT consumes system resource (especially in an FDD system). These two facts, potential rate gains on the one hand, and the cost of resource to make that possible on the other, have stimulated a large body of research on the partial CSIT, limited channel state feedback, case (see [3] and references therein).

Among strategies for MIMO broadcast transmission based on limited CSIT, one may distinguish between two large categories of schemes, namely (i) opportunistic schemes and (ii) non-opportunistic channel quantization-based schemes. In a subset of the former (e.g. [5], [6]), randomly designed beams are launched from the base station, then the users reply with a simple SINR feedback to allow the assignment of the best set of users to the pre-launched beams, while in another (e.g. [12], [13], [14]) *only an appropriate subset* of the users feed back their channel and a subgroup of these users is selected for transmission. In the second category, the beamforming matrix is designed after receiving CSIT from all users in the form of their quantized channel information (see among others [8], [9], [10]). In this line of work, the channel quantization is based on a codebook with *fixed* number of feedback bits,

though this number may possibly be adapted to macroscopic system parameters such as number of users and average SNR. Recently, several ideas have emerged which suggest that the system could gain significantly from the *adaptivity* of the allocated feedback load. This is under the framework that a flexible design of the feedback channel could be adopted in forthcoming wireless standards, one in which the feedback load can be made time-varying, fulfilling an *average* feedback load constraint rather than a fixed load constraint.

In an example of such proposals, [11] noted that scheduling and beamforming require very different accuracy levels in CSIT description and recommended that the feedback rate should be split optimally across two feedback stages, one for scheduling and another one for beamforming, where the number of CSIT bits is different for the two stages (yet the sum remains constant). Note that as a variant of this idea applied to opportunistic schemes, [15] investigated having a low-rate first stage, complemented with a second stage to enhance the performance of random beamforming [6] for cases when the number of users is relatively low.

Recently, in [1], yet another *adaptive feedback* approach is adopted, where the key idea is that, if each user were subject to an average feedback rate constraint, rather than a peak constraint, then the resource allocated for feedback at each scheduling period could be optimized as function of the instantaneous channel conditions. This idea follows the basic intuition that a user ought to spend more on feedback at a particular time slot when the expected return for him/her in terms of downlink rate is larger. Conversely, if the odds to be selected by the scheduler for a particular user are low, there is little interest for that user to describe the channel accurately to the base at that time. Therefore the feedback rate optimization can be formalized so as to take into account (i) *the user's channel quality* and (ii) *the probability with which this user will be selected*.

In this paper, we revisit the results obtained in [1] to take a more general scenario for channel quality indicators into consideration, reformulate the feedback rate adaptation problem in this 'corrected' framework, discuss its complexity and describe some suboptimal solutions. Performance is illustrated with Monte Carlo simulations.

*Notation:*  $\mathbb{E}$  denotes statistical expectation.  $\mathbb{C}^n$  denotes the

$n$ -dimensional complex space. Boldface lowercase letters are used to denote vectors, and boldface uppercase denote matrices.  $f_x(\cdot)$  gives the probability density function (pdf) of random variable  $x$ , and  $F_x(\cdot)$  its cumulative density function (cdf). The probability of an event  $A$  occurring is denoted by  $Pr[A]$ . The  $l^2$ -norm of vector  $\mathbf{x}$  is denoted as  $\|\mathbf{x}\|$ , and  $\tilde{\mathbf{x}} \triangleq \frac{\mathbf{x}}{\|\mathbf{x}\|}$ . Finally,  $\log(\cdot)$  is the natural logarithm.

## II. SYSTEM MODEL

We consider a multi-antenna Gaussian broadcast channel, where a transmitter equipped with  $N_t$  antennas serves selected users among  $N \geq N_t$  single-antenna receivers under a total transmit power constraint  $P$ . The latter are assumed to have perfect channel knowledge. The received signal at user  $k$ , denoted  $y_k \in \mathbb{C}$  can be written as:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k \quad (1)$$

where  $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$  is the transmitted signal vector, such that  $\mathbb{E}\|\mathbf{x}\|^2 = P$ ,  $\mathbf{h}_k \in \mathbb{C}^{1 \times N_t}$  and  $n_k \in \mathbb{C}$  represent the channel vector and the noise at the  $k$ th user, respectively. We assume perfect channel knowledge at the receiver, and that the entries of the noise vector are i.i.d. zero mean unit variance complex Gaussian random variables (r.v.'s),  $CN(0, 1)$ . Furthermore, we assume a block-fading channel and focus on the ergodic sum rate as system performance measure.

### A. CSI and Quantization

We assume the  $N$  receivers have i.i.d. Rayleigh fading channels, and assume that the fed back CSI consists of: i) quantized channel direction information (CDI), and ii) channel quality information (CQI) [8], [9], [10].  $\tilde{\mathbf{h}}$  denotes the true direction,  $\hat{\mathbf{h}}$  its quantized version.

We assume different users use different, independent codebooks. The corresponding quantization error is defined as  $\sin^2 \epsilon$ , where  $\epsilon \triangleq \angle(\tilde{\mathbf{h}}, \hat{\mathbf{h}})$ , the angle between the true and quantized channel directions. The cdf of any CDI quantizer may be upper-bounded by [17]:

$$F_{\sin^2 \epsilon}(x) = \begin{cases} \delta^{1-N_t} x^{N_t-1} & 0 \leq x \leq \delta \\ 1 & x > \delta \end{cases} \quad (2)$$

where  $\delta \triangleq 2^{-b/(N_t-1)}$ ,  $b$  being the number of bits used for quantization. For tractability, this distribution will be adopted in our derivations.

Although there are several different options for defining the CQI at the user level, a popular CQI measure, adopted here, is the following SINR estimate given by [8], [9], [10]:

$$\hat{\gamma} \triangleq \frac{P/N_t \alpha \cos^2 \epsilon}{1 + P/N_t \alpha \sin^2 \epsilon} \quad (3)$$

where  $\alpha \triangleq \|\mathbf{h}\|^2$ . This choice is justified in the following section.

As in the cited papers, CQI is assumed to be unquantized. However the effect of quantizing it is investigated through simulations at the end of the paper.

### B. User Selection and Precoding Scheme

Zero-forcing beamforming (ZFBF) with uniform power allocation is the adopted precoding scheme. Thus, the transmitted signal  $\mathbf{x}$  is given by:

$$\mathbf{x} = \sqrt{\frac{P}{K}} \mathbf{W} \mathbf{s} \quad (4)$$

where  $K \leq N_t$  is the number of users scheduled,  $\mathbf{W} \in \mathbb{C}^{N_t \times K}$  is a zero-forcing matrix with respect to the quantized channel matrix, having unit norm columns, and  $\mathbf{s}$  is the vector of  $K$  symbols to be transmitted, its entries being independently generated zero-mean unit-variance complex Gaussian variables.

The scheduling scheme tries to maximize the sum rate achieved by ZFBF to a subset of users, based on the fed back CSI: in the optimal case, this is done through exhaustive search over all groups of up to  $N_t$  users; a suboptimal scheme would use a greedy algorithm such as those of [4], [19]. As the transmitter relies on limited feedback information, and the CDI will involve some quantization error, only taking the channel norm into consideration when scheduling (i.e. using the product of CDI and true channel norm as a channel estimate and using that for quantization) will lead to the MUD gain eventually being lost [9]. Thus the scheduling algorithm should rely on a CQI, such as the one in (3), which takes both channel norm and quantization error into consideration [9], [10], [18].

## III. ADAPTIVE FEEDBACK RATE ALLOCATION

To maximize the total throughput of the system, the transmitter needs to i) determine the best group of users, and ii) design the corresponding precoding matrix for transmission to those users. Under an average feedback rate constraint, it thus makes sense for a given user to quantize its channel more accurately if it is more likely to be scheduled, since this would lead to a higher throughput being achieved. As a given user only has access to partial channel information (it knows its own local channel state information, and possibly the channel statistics of the other users as assumed here), this amounts to adapting the feedback rate so as to maximize a user's *expected* rate, based on its current local knowledge. As first argued in [1], this adaptation could be made as a function of the channel energy  $\alpha \triangleq \|\mathbf{h}\|^2$ , since the channel direction of an individual user does not provide information on its separability from other users, the directions of which being unknown to the user under consideration.

Denoting the event of being scheduled by  $S$ , this can be approximated as:

$$\mathbb{E}R = \int_0^\infty \mathbb{E}[Pr[S]R|\alpha = a] f_\alpha(a) da, \quad (5)$$

where the expectation is over all unknowns/random variables at the user.

Note that the probability of being scheduled and the achieved rate will both depend on the scheduling algorithm, and on the knowledge at the transmitter. Thus, for a scheduling

rule based on the fed back CQI  $\hat{\gamma}_\alpha$  (cf. (3), the subscript emphasizes the dependence on the channel energy), we have:

$$\begin{aligned} & \mathbb{E}[Pr[S]R|\alpha = a] \\ &= \int_0^\infty \mathbb{E}[Pr[S]R|\alpha = a, \hat{\gamma}_a] f_{\hat{\gamma}_\alpha|\alpha}(\hat{\gamma}_a|a) d\hat{\gamma}_a. \end{aligned} \quad (6)$$

Given  $\hat{\gamma}_a$ , and that  $\alpha = a$ , the achievable rate may be approximated by:  $\log_2(1 + \hat{\gamma}_\alpha)$ . This amounts to assuming that  $N_t$  users, whose quantized channels are orthogonal, will be scheduled, an approximation which will become more accurate as the number of users increases. The conditional distribution  $f_{\hat{\gamma}_\alpha|\alpha}(\hat{\gamma}_a|a)$  may be obtained by combining (2) and (3).

Equation (5) becomes<sup>1</sup>:

$$\begin{aligned} \mathbb{E}R &\approx \int_0^\infty da f_\alpha(a) \\ &\left[ \int_0^\infty d\hat{\gamma}_a Pr[S|\alpha = a, \hat{\gamma}_a] \log_2(1 + \hat{\gamma}_a) f_{\hat{\gamma}_\alpha|\alpha}(\hat{\gamma}_a|a) \right]. \end{aligned} \quad (7)$$

Ideally, a user should maximize its expected rate, as described above, i.e. find the best function  $b(\alpha)$ , which specifies the number of feedback bits as a function of channel energy  $\alpha$ , subject to an average feedback rate constraint  $\bar{B}$ :

$$\int_0^\infty b(a) f_\alpha(a) da = \bar{B}. \quad (8)$$

But expressing  $Pr[S|\alpha = a, \hat{\gamma}_a]$  in a closed form manner turns out to be quite untractable: the probability of being scheduled will be a function of the distribution of the CQI  $\hat{\gamma}$ , which in turn depends on the way the bits are allocated over the entire domain of the channel norm random variable. This leads us to consider sub-optimal solutions.

#### A. Scheduling-Independent Adaptation

The simplest approach would be to ignore the scheduling probability in the optimization (as if it were independent of any of the parameters, or always equal to 1). This could be a reasonable approach if the user does not know the distributions of the other users, or if the number of users in the system changes too fast to be tracked.

The expected rate (7) becomes:

$$\int_0^\infty da f_\alpha(a) \left[ \int_0^\infty d\hat{\gamma}_a \log_2(1 + \hat{\gamma}_a) f_{\hat{\gamma}_\alpha|\alpha}(\hat{\gamma}_a|a) \right]. \quad (9)$$

The inner integral is given by (cf. Eq. (15) in [1]):

$$\begin{aligned} & \int_0^\infty d\hat{\gamma}_a \log_2(1 + \hat{\gamma}_a) f_{\hat{\gamma}_\alpha|\alpha}(\hat{\gamma}_a|a) \\ &= \frac{1}{\log_2 e} \left\{ \log(1 + c_a) - \left( 1 + \frac{(-1)^{N_t}}{(c_a \delta_a)^{N_t-1}} \right) \log(1 + c_a \delta_a) \right. \\ &+ \left. \sum_{i=0}^{N_t-2} \frac{(-1)^i}{(c_a \delta_a)^i} \frac{1}{N_t - 1 - i} \right\}, \end{aligned} \quad (10)$$

where  $c_a = \frac{P_a}{N_t}$  and  $\delta_a = 2^{-b(a)/(N_t-1)}$ .

<sup>1</sup>This formulation assumes that exactly  $N_t$  users are scheduled, which will depend on the type of scheduling algorithm used, and on other system parameters such as transmit power.

This is a concave function of  $b(a)$  so that relaxing the integer constraint on it transforms the problem into a variational problem for which an optimum can be guaranteed (negative second variation).

#### B. Quantization-Error Independent Adaptation

Another, alternative, approach is to instead assume the probability of being scheduled is some increasing function of the channel energy alone  $\alpha$ , which will be denoted  $Pr[S|\alpha = a]$ . The intuition behind why this would work better is that for fixed CDI quantization error, the CQI  $\hat{\gamma}$  is an increasing function of  $\alpha$ . Thus, at higher  $\alpha$ , there is more to gain (in terms of rate achieved) by assigning more bits to CDI quantization and consequently feeding back a higher CQI value, so that it makes sense to assign higher weight (the scheduling probability) to these values of  $\alpha$ .

Equation (7) is thus approximated by:

$$\int_0^\infty da f_\alpha(a) Pr[S|\alpha = a] \left[ \int_0^\infty d\hat{\gamma}_a \log_2(1 + \hat{\gamma}_a) f_{\hat{\gamma}_\alpha|\alpha}(\hat{\gamma}_a|a) \right] \quad (11)$$

Though conceptually different, this last formulation is similar to that of the problem solved in [1], for which a 'water-filling' solution was derived, by relaxing the integer constraint on the channel norm to feedback bit rate mapping function  $b(\cdot)$ . The solution is restated here for completeness. Reformulating the quantity in brackets in (11) as a function of the channel energy instance  $a$  and the corresponding number of bits  $b(a)$ , and denoting it by  $g(a, b(a))$ , the problem can be reduced to solving for a positive scalar, the optimal 'water level'  $\lambda^*$ , using a line search method such as the bisection method, so as to meet the average bit rate constraint with equality.

$\lambda^*$  determines a threshold channel energy  $a_{thres}$ , at which:

$$P_S(a_{thres}) \frac{\partial g(a_{thres}, b(a_{thres}))}{\partial b(a_{thres})} \Big|_{b(a_{thres})=0} = \lambda^*. \quad (12)$$

For  $a \leq a_{thres}$ , no bits are allocated for feedback, whereas for  $a > a_{thres}$ , the optimal mapping  $b^*(\cdot)$  is such that:

$$P_S(a) \frac{\partial g(a, b(a))}{\partial b(a)} \Big|_{b(a)=b^*(a)} = \lambda^* \quad (13)$$

The scheduling probability function  $Pr[S|\alpha = a]$  used is the probability of belonging to the group of  $N_t$  users with the best channel norms:

$$Pr[S|\alpha = a] = \sum_{i=0}^{N_t-1} \binom{N-1}{i} (F_\alpha(a))^{N-1-i} (1 - F_\alpha(a))^i. \quad (14)$$

Though this yields a bit rate allocation similar to the one in [1], we emphasize that there is a fundamental difference in the actual scheduling scheme implemented in both systems: in [1] the CQI used for scheduling was the channel norm, as opposed to the SINR estimate used here.

#### IV. SIMULATION RESULTS

To illustrate the performance gains of the suggested scheme, Monte Carlo simulations were carried out, for different numbers of users and antennas under different average feedback bit rate constraints. For the CDI quantization, random vector quantization (RVQ) [16] was used: since the cdf of the quantization error is known for this particular scheme, the latter can be used to generate the quantized CDI, thus speeding up simulations [20]; using RVQ also shows that the results based on the model (2) are still meaningful for a more realistic quantization scheme. Note that in this case, the quantization error model used in the bit rate adaptation algorithm and the actual one differ. The scheduling algorithm used is the one from [18], which is a greedy algorithm that schedules up to  $N_t$  users, stopping when adding one more user no longer increases the *estimated* total throughput.

Figure 1 illustrates the importance of taking the scheduling probability (even if it is actually only an approximation) into consideration: the gains obtained from applying the scheme described in section III-A are negligible for low numbers of users (mainly due to the fact that the expected rate approximation is quite inaccurate), and relatively small for higher numbers, since this scheme essentially wastes bits on cases where the likelihood of being scheduled is small. We thus focus on the scheme from section III-B in what follows.

Figures 2 and 3 show the performance of the latter scheme, where additionally quantization of the CQI was considered, and the performance degradation it causes tested. Thus, instead of assuming the CQI is known perfectly at the base station, a fixed codebook is generated by applying Lloyd's algorithm to CQI samples corresponding to the given channel model and the CDI feedback bit rate allocation obtained from the algorithms considered. Denoting by  $b_{CQI}$  the number of bits used for CQI quantization, this could be viewed as a suboptimal feedback allocation over the entire CSI (CDI + CQI), under an average total feedback rate equal to  $\bar{B} + b_{CQI}$ .

Though quantizing CQI causes losses in total throughput, comparing to the scheme where a constant feedback rate is used and the CQI is perfectly known, the adaptive scheme still performs better, sometimes even under a lower average bit rate (4 bits/user vs. 6 bits/user in figure 2). Moreover, as expected, the more users in the system, the more there is to gain from adapting to instantaneous channel conditions, as this leads to the probability of being scheduled to be concentrated within a shrinking (with the number of users) interval of the channel norm range. On the other hand, the greater the number of antennas at the transmitter, the greater the loss with respect to the full CSIT case: this is in accordance with non-adaptive schemes such as [7], [9] which show that the number of bits used needs to be scaled linearly with the number of antennas and the SNR in order to maintain a fixed performance loss gap; however, as figure 3 shows, the contrast between the adaptive scheme and a constant feedback rate scheme may still be quite significant in this case (more transmit antennas), for sufficiently many users in the system.

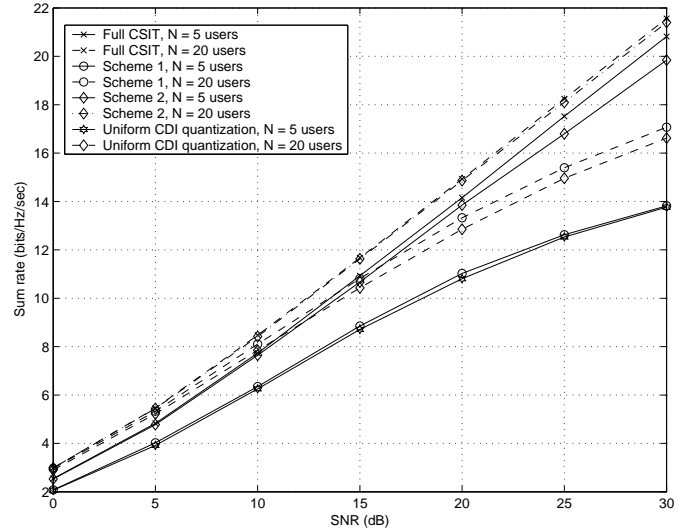


Fig. 1. Achievable sum rate for  $N=5, 20$  users, and average feedback rates per user of 6 bits for both suboptimal schemes (scheme 1 corresponds to section III-A, scheme 2 to section III-B)

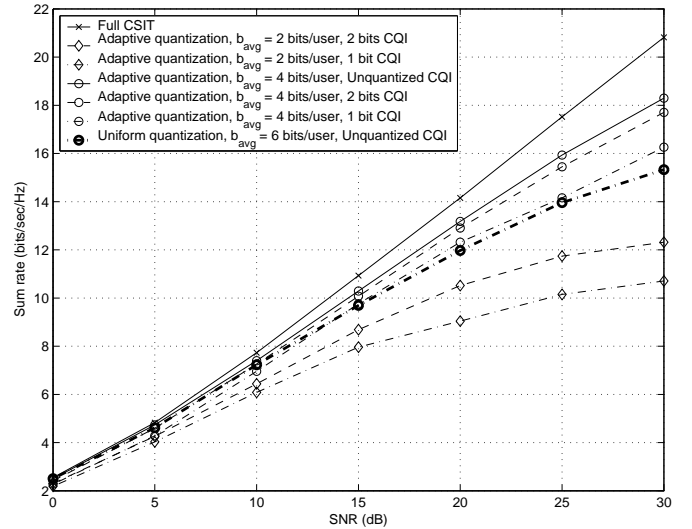


Fig. 2. Achievable sum rate for  $N=10$  users and  $N_t = 2$  antennas

#### V. CONCLUSION

The adaptive scheme in which the feedback rate is optimized as a function of the channel quality first proposed in [1] was reformulated to account for a better scheduling algorithm, and suboptimal algorithms to implement it were proposed for users with Rayleigh fading i.i.d. channels in the system. The associated performance gains were illustrated through Monte Carlo simulations, which included investigating the effect of quantizing the CQI. Future work should investigate adaptation in more general channels, for example in the case where user channels are no longer identically distributed.

#### ACKNOWLEDGMENT

This work was supported in part by the ORMAC project.

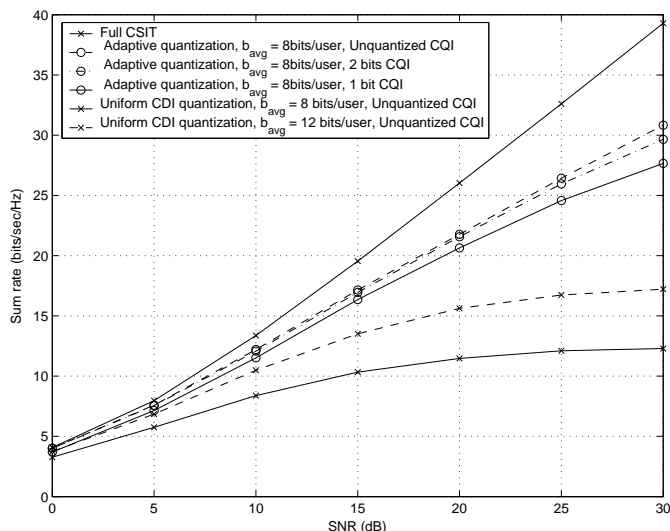


Fig. 3. Achievable sum rate for  $N=20$  users and  $N_t = 4$  antennas

## REFERENCES

- [1] R. Zakhour and D. Gesbert, "Adaptive Feedback Rate Control in MIMO Broadcast Systems", submitted to *ITW 2008*
- [2] R. Knopp and P.A. Humblet, "Information capacity and power control in single-cell multiuser communications", in *Proc. of Int. Conf. on Communications*, Vol. 1, June 1995.
- [3] D. Gesbert, M. Kountouris, R. Heath, C-B. Chae and T. Salzer, "From Single User to Multiuser Communications: Shifting the MIMO paradigm", *IEEE Signal Processing Magazine*, Oct. 2007
- [4] T. Yoo and A. Goldsmith, "On the optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming", *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 3, March 2006
- [5] P. Viswanath, D.N.C. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas", *IEEE Transactions on Information Theory*, Vol. 48, No. 6, June 2002
- [6] M. Sharif, B. Hassibi, "On the Capacity of MIMO Broadcast Channels with Partial Side Information", *IEEE Transactions on Information Theory*, Vol. 51, No. 2, 506-522, Feb. 2005.
- [7] N. Jindal, "MIMO Broadcast Channels with Finite Rate Feedback", *IEEE Transactions on Information Theory*, Vol. 52, No. 11, Nov. 2006
- [8] T. Yoo, N. Jindal and A. Goldsmith, "Finite-Rate Feedback MIMO Broadcast Channels with a Large Number of Users", *2006 IEEE International Symposium on Information Theory*, July 2006
- [9] T. Yoo, N. Jindal and A. Goldsmith, "Multi-antenna Downlink Channels with Limited Feedback and User Selection", *IEEE Journal Selected Areas in Communications*, Sept. 2007.
- [10] M. Kountouris, R. de Francisco, D. Gesbert, D.T.M. Slock and T. Sizer, "Efficient Metrics for Scheduling in MIMO Broadcast Channels with Limited Feedback", *IEEE ICASSP 2007*, April 2007
- [11] R. Zakhour and D. Gesbert, "A two-stage approach to feedback design in MU-MIMO channels with limited channel state information", *18th IEEE Annual International Symposium on Personal Indoor and Mobile Radio Communications*, Sept. 2007
- [12] K. Huang, R. W. Heath Jr. and J. G. Andrews, "Multi-user Aware Limited Feedback for MIMO Systems", submitted to *IEEE Transactions on Signal Processing*, January 2007.
- [13] C. Swannack, G. W. Wornell and E. Uysal-Biyikoglu, "MIMO Broadcast Scheduling with Quantized Channel State Information", *2006 IEEE International Symposium on Information Theory*, July 2006.
- [14] K. Agarwal, C. Hwand and J. Cioffi, "Scalable Feedback Protocol Asymptotically Achieving Broadcast Channel Sum-capacity", *Stanford Technical Report*, Sep. 2006
- [15] M. Kountouris, D. Gesbert and T. Sizer, "Enhanced Multiuser Random Beamforming: Dealing with the Not So Large Number of Users Case" submitted to *IEEE Journal on Selected Areas in Communications, Special Issue on Limited Feedback Communications*, Nov. 2007
- [16] W. Santipach and M. Honig, "Asymptotic Capacity of Beamforming with Limited Feedback" *emph2004 IEEE International Symposium on Information Theory*, July 2004.
- [17] S. Zhou, Z. Wang, and G. B. Giannakis, "Quantifying the Power Loss When Transmit Beamforming Relies on Finite-Rate Feedback", *IEEE Transactions on Wireless Communications*, July 2005
- [18] M. Trivellato, F. Boccardi and H. Huang, "On Transceiver Design and Channel Quantization for Downlink Multiuser MIMO Systems with Limited Feedback", submitted to *IEEE Journal on Selected Areas in Communications, Special Issue on Limited Feedback Communications*, Nov. 2007
- [19] G. Dimic and N. Sidiropoulos, "On Downlink Beamforming with Greedy User Selection: Performance Analysis and Simple New Algorithm", *IEEE Transactions on Signal Processing*, Oct. 2005
- [20] N. Ravindran and N. Jindal, "Multi-User Diversity vs. Accurate Channel Feedback for MIMO Broadcast Channels", *Submitted to ICC 08*, Sept. 2007