# Efficient Scheduling of Multi-User Multi-Antenna Systems

Krishna P Jagannathan[†], Sem Borst[*,1], Phil Whiting[*], Eytan Modiano[†]

[*]Bell Laboratories, Lucent Technologies, P.O. Box 94079, Murray Hill, NJ 07974

[†]Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology, Cambridge, MA 02139

***Abstract*— The capacity region of the Gaussian multi-antenna broadcast channel was characterized recently in [19]. It was shown that a scheme based on Dirty Paper Coding [2] achieves the full capacity region when the transmitter has perfect channel state information. However, this scheme potentially involves considerable amounts of feedback and complex algorithms for coding and user selection. This has led to a quest for practical transmission schemes and ways to reduce the amount of channel state information required. In particular, it has been shown that when the total number of users is large, the sum capacity can be closely approached by transmitting to a small subset of near-orthogonal users.**

**In order to further quantify the latter observation, we study a Gaussian broadcast channel with two transmit antennas and $K$ statistically identical, independent users each with a single receive antenna. We obtain an exact asymptotic characterization of the gap between the full sum capacity and the rate that can be achieved by transmitting to a suitably selected pair of users. Specifically, we consider various simple schemes for user-pair selection that take into account the channel norms as well as the relative orientation of the channel vectors. We conclude that a scheme that picks the strongest user and selects a second user to form the best pair, is asymptotically optimal, while also being attractive in terms of feedback and operational complexity. Numerical experiments show that the asymptotic results tend to be remarkably accurate, and that the proposed scheme significantly outperforms a beam-forming strategy for a typical number of users.**

## I. Introduction

The use of wireless communications continues to experience tremendous growth. This continual growth creates increasing pressure to squeeze the most out of the limited amount of wireless spectrum available. The use of antenna arrays offers a promising technique for improving spectrum efficiency so as to achieve higher data rates, larger capacity, better coverage, or a combination of these. The multi-antenna Broadcast Channel (BC) has been the subject of much research interest recently, owing primarily to the substantial capacity benefits that these systems can potentially offer.

In the present paper we consider the downlink transmission from a single base station equipped with $M > 1$ transmit antennas to $K > 1$ independent users each with a single receive antenna. In information-theoretic terms, this may be modeled as a multi-antenna BC. The sum capacity for the Gaussian BC was first obtained for the case of two users with a single receive antenna by Caire & Shamai [1]. Subsequently, Viswanath & Tse [16] and Vishwanath *et al.* [15] extended the result for the sum capacity to an arbitrary number of users and receive antennas by exploiting a powerful duality relation with the multi-access channel which was further explored in Jindal *et al.* [9]. Recently, Weingarten *et al.* [19] provided a characterization of the entire capacity region. They showed that a scheme based on Dirty Paper Coding (DPC) [2] achieves the full capacity region of the multi-antenna BC.

The above capacity results rely on the assumption that perfect channel state information is available at the transmitter. However, the amount of overhead involved in feeding back the channel state information may be prohibitive, especially when the number of users is large. Further, DPC is quite a sophisticated technique and challenging to implement in an actual system.

Motivated by the above issues, extensive efforts have been made to devise practical transmission and coding schemes and find ways to reduce the amount of channel state information required. Hochwald *et al.* [3], [4] describe an algorithm based on channel inversion and sphere encoding, and show that it closely approaches capacity while being simpler to operate than DPC. Jindal [6] considers a multi-antenna BC with limited channel feedback information, and shows that the full capacity gain at high SNR values is achievable as long as the number of feedback bits grows linearly with the SNR (in dB).

In the case of a single receive antenna, it is known that multiple transmit antennas yield substantial capacity gains when several users are served simultaneously. In particular, Jindal & Goldsmith [7] show that the gain over a TDMA strategy is approximately $\min\{M, K\}$, i.e., the minimum of the number of transmit antennas and the number of users. Jindal [5] demonstrates that the sum capacity grows with the SNR at rate $\min\{M, K\}$. In other words, multiple transmit antennas provide potentially huge capacity gains, but it is necessary that at least $M$ users are served simultaneously in order to reap the full benefits. Transmitting to fewer than

[1]The second author is also affiliated with CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands, and Department of Mathematics & Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

$M$ users falls short of the maximum capacity as it fails to fully exploit the available degrees of freedom. Transmitting to more than $M$ users may be necessary to achieve the maximum capacity in general, but the results in [5], [7] indicate that transmitting to a suitably selected subset of exactly $M$ near-orthogonal users is close to optimal. When the total number of users to choose from is sufficiently large, such a subset exists with high probability, see [13], [14] for a rigorous asymptotic characterization.

Clearly, the above principle allows for a reduction of the amount of channel feedback and coding complexity. In particular, it suggests the use of beamforming schemes that construct $M$ (random) orthogonal beams and serve the users with the best channel quality along each of the beams with equal power. Transmission schemes along these lines are presented in [12] and [18].

In the present paper we focus on the case of two transmit antennas and statistically independent, identically distributed (i.i.d) users. We derive an exact asymptotic characterization of the gap between the full sum capacity and the rate that can be achieved by transmitting to a suitably selected pair of users. In particular, we consider a scheme that picks the user with the largest channel gain, and then selects a second user from the $L$ strongest ones to form the best possible pair with it, taking the orientation of the channel vectors into account as well. We prove that the rate gap converges to $1/(L-1)$ when the total number of users $K$ tends to infinity. Allowing $L$ to increase with $K$, we conclude that the gap asymptotically vanishes, and that the sum capacity is achievable by transmitting to a properly chosen pair of users. Numerical results show that the asymptotics tend to be remarkably accurate, even for a relatively moderate number of users. The fact that the rate gap decays as $1/(L-1)$ also suggests that a modest value of $L$ is adequate for most practical purposes. The above results have significant implications for the design of channel feedback mechanisms and transmission techniques.

The remainder of the paper is organized as follows. In Section II we present a detailed model description and review some known results for the sum capacity of the Gaussian multi-antenna BC. In Section III, we provide some useful bounds for the sum capacity and other preparatory results. Our main asymptotic results are contained in Section IV. In Section V we present some numerical results to show that our asymptotic results are quite accurate, even for a moderate number of users. In Section VI we point out some directions for future work.

## II. System model and known results

### A. Model

We consider a broadcast channel with $M > 1$ transmit antennas and $K$ independent receivers each with a single antenna, as schematically represented in Figure 1(a).

Let $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$ be the transmitted vector signal and let $\boldsymbol{h}_k \in \mathbb{C}^{1\times M}$ be the channel-gain vector of the $k$-th receiver. Denote by $\boldsymbol{H} = [\boldsymbol{h}_1^\dagger \boldsymbol{h}_2^\dagger \cdots \boldsymbol{h}_K^\dagger]^\dagger$ the concatenated channel matrix of all $K$ receivers. For now, the matrix $\boldsymbol{H}$ is arbitrary but

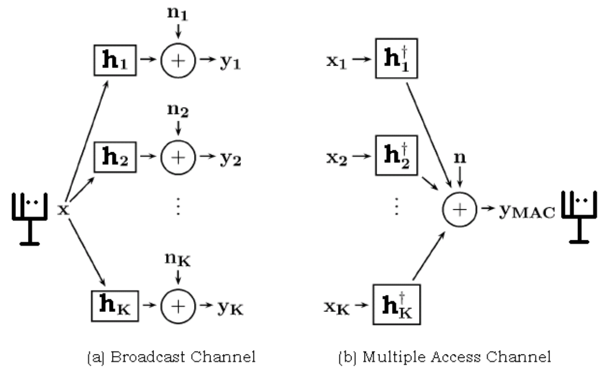


Fig. 1. The multi-antenna broadcast channel (left) and the multiple access channel (right) have the same capacity region

assumed to be fixed. We further assume that the transmitter has perfect channel state information, i.e., exact knowledge of the matrix $\boldsymbol{H}$. The circularly symmetric complex Gaussian noise at the $k$-th receiver is $n_k \in \mathbb{C}$ where $n_k \sim \mathcal{CN}(0,1)$. Thus the received signal at the $k$-th receiver is $y_k = \boldsymbol{h}_k \boldsymbol{x}_k + n_k$. The covariance matrix of the transmitted signal is $\Sigma_x = \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\dagger\right]$. The transmitter is subject to a power constraint $P$, which means $Tr(\Sigma_x) \leq P$.

### B. Sum capacity

The sum capacity is a key metric of interest for the BC as it measures the total achievable system throughput. Since it only considers the aggregate rate, it does not reflect potential fairness issues that arise when users with widely disparate channel characteristics obtain vastly different rates. In the present paper, however, we focus on the case of statistically identical users, which by symmetry will obtain equal long-term rate shares.

In case of a single transmit antenna, the sum capacity is simply equal to the largest single-user capacity in the system [11], i.e., the sum rate is maximized by transmitting only to the user with the largest channel gain. However, this is not true when there are multiple transmit antennas. In that case, the sum capacity is achieved by using DPC techniques to simultaneously transmit to several users.

From the results in [1], [15], [16], the sum capacity, denoted by $\mathcal{C}_{BC}(\boldsymbol{H}, P)$, can be expressed in terms of the following maximization problem:

$$\mathcal{C}_{BC}(\boldsymbol{H}, P) = \max \sum_{k=1}^{K} \log \frac{\det\left(I_M + \boldsymbol{h}_k (\sum_{j\leq k} \boldsymbol{\Sigma}_j) \boldsymbol{h}_k^\dagger\right)}{\det\left(I_M + \boldsymbol{h}_k (\sum_{j<k} \boldsymbol{\Sigma}_j) \boldsymbol{h}_k^\dagger\right)}, \quad (1)$$

where the optimization is performed over the set of all positive semi-definite covariance matrices $\boldsymbol{\Sigma}_k$, $k = 1, \ldots, K$, such that $\sum_{k=1}^{K} Tr(\boldsymbol{\Sigma}_k) \leq P$.

Unfortunately, the objective function in the above problem is non-concave, which makes it hard to deal with analytically as well as numerically. However, in [15], a duality is shown to exist between the BC and the Gaussian multiple access channel (MAC) with a sum-power constraint $P$. That is, the dual MAC which is formed by reversing the roles of transmitters and receivers, as represented in Figure 1(b), has the same capacity region as the BC.

Using this duality, the sum capacity of the BC can be written in terms of the dual MAC sum capacity as

$$\mathcal{C}_{BC}(\boldsymbol{H}, P) = \max_{\sum_{k=1}^{K} P_k \leq P} \log\det\left(I_M + \sum_{k=1}^{K} P_k \boldsymbol{h}_k^\dagger \boldsymbol{h}_k\right), \quad (2)$$

where $P_k \geq 0$ denotes the power allocated to the $k$-th receiver. Note that the objective function in (2) is indeed concave in the values of the $P_k$'s. Specialized algorithms for calculating the BC sum capacity have been developed in [8].

## III. Preliminary results

### A. Bounds for the sum capacity

We now present some useful upper and lower bounds for the BC sum capacity. Denote by $\boldsymbol{h}_{(i)}$ the channel vector of the receiver with the $i$-th largest norm, i.e., $||\boldsymbol{h}_{(1)}||^2 \geq ||\boldsymbol{h}_{(2)}||^2 \geq \cdots \geq ||\boldsymbol{h}_{(K)}||^2$. The next upper bound for the sum capacity is established in [7].

$$\mathcal{C}_{BC}(\boldsymbol{H}, P) \leq M \log\left(1 + \frac{P}{M}||\boldsymbol{h}_{(1)}||^2\right). \quad (3)$$

Observe that the above upper bound can be achieved when there are $M$ receivers with orthogonal channel vectors tied for the maximum norm $||\boldsymbol{h}_{(1)}||^2$.

Next, we obtain a simple lower bound for the sum capacity for the case of $M = 2$ transmit antennas.

$$\begin{aligned} \mathcal{C}_{BC}(\boldsymbol{H}, P) &\geq C(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \quad (4) \\ &:= \log\det\left(I_2 + \frac{P}{2}(\boldsymbol{h}_i^\dagger \boldsymbol{h}_i + \boldsymbol{h}_j^\dagger \boldsymbol{h}_j)\right). \end{aligned}$$

Observe that the above lower bound corresponds to scheduling any two users $i$ and $j$ with equal power.

We now present a lemma that will be useful in simplifying the expression for the BC sum capacity in (2) as well as the lower bound in expression (4).

*Lemma 3.1:* For any $K$, $M$,

$$\det\left(I_M + \sum_{k=1}^{K} Q_k \boldsymbol{h}_k^\dagger \boldsymbol{h}_k\right) = \det(I_K + J),$$

with $J_{kl} := \sqrt{Q_k Q_l}\boldsymbol{h}_k \boldsymbol{h}_l^\dagger$, $k, l = 1, \ldots, K$.

**Proof:** Define the $K \times M$ matrix $\boldsymbol{H}$ by $H_{km} := \sqrt{Q_k} h_{km}$, $k = 1, \ldots, K$, $m = 1, \ldots, M$. The proof then follows easily from the determinant identity $\det(I_M + \boldsymbol{H}^\dagger \boldsymbol{H}) = \det(I_K + \boldsymbol{H}\boldsymbol{H}^\dagger)$. Indeed,

$$\begin{aligned} &\det(I_M + \sum_{k=1}^{K} Q_k \boldsymbol{h}_k^\dagger \boldsymbol{h}_k) \\ &= \det\left(I_M + \begin{bmatrix} \sqrt{Q_1}\boldsymbol{h}_1 \\ \cdots \\ \sqrt{Q_K}\boldsymbol{h}_K \end{bmatrix}^\dagger \begin{bmatrix} \sqrt{Q_1}\boldsymbol{h}_1 \\ \cdots \\ \sqrt{Q_K}\boldsymbol{h}_K \end{bmatrix}\right) \\ &= \det(I_M + \boldsymbol{H}^\dagger \boldsymbol{H}) \\ &= \det(I_K + \boldsymbol{H}\boldsymbol{H}^\dagger) \\ &= \det\left(I_K + \begin{bmatrix} \sqrt{Q_1}\boldsymbol{h}_1 \\ \cdots \\ \sqrt{Q_K}\boldsymbol{h}_K \end{bmatrix} \begin{bmatrix} \sqrt{Q_1}\boldsymbol{h}_1 \\ \cdots \\ \sqrt{Q_K}\boldsymbol{h}_K \end{bmatrix}^\dagger\right) \\ &= \det(I_K + J). \end{aligned}$$

□

Using Lemma 3.1, we can simplify the lower bound in (4) to obtain

$$\mathcal{C}_{BC}(\boldsymbol{H}, P) \geq \log\det\left(I_2 + \frac{P}{2}(\boldsymbol{h}_i^\dagger \boldsymbol{h}_i + \boldsymbol{h}_j^\dagger \boldsymbol{h}_j)\right) =$$

$$\log\left(1 + \frac{P}{2}\left(||\boldsymbol{h}_i||^2 + ||\boldsymbol{h}_j||^2\right) + \frac{P^2}{4}||\boldsymbol{h}_i||^2||\boldsymbol{h}_j||^2 U_{ij}\right),$$

where $U_{ij} := 1 - \frac{|<\boldsymbol{h}_i, \boldsymbol{h}_j>|^2}{||\boldsymbol{h}_i||^2||\boldsymbol{h}_j||^2}$.

The above lower bound reflects the fact that the sum capacity for two users crucially depends on the norms of the respective channel vectors, and their degree of orthogonality. In particular, the sum capacity is large when the two users are nearly orthogonal and have large channel gains.

### B. Random channel vectors

So far, we have assumed the channel vectors to be arbitrary but fixed. In order to derive meaningful asymptotic results, we henceforth assume the channel vectors to be random, and primarily consider the *expected* sum capacity. To be specific, we assume that the various components of the channel vector of a user are independent and identically distributed according to $\mathcal{CN}(0, 1)$, which corresponds to independent Rayleigh fading.

*Remark 3.1:* The randomness in the channel vectors may be interpreted as variations resulting from fast fading due to multi-path propagation effects. The impact of distance-related path loss and slow fading (shadow fading due to obstacles) is supposed to be reflected in the mean signal-to-noise ratio $P$. The expected sum capacity then represents the long-term system throughput. Implicitly, we make here the usual block fading assumption, where the frame length is short enough for the channel to remain (nearly) constant over the duration of a frame, yet sufficiently long to achieve a transmission rate close to the theoretical capacity.

The next two lemmas characterize the distribution of the squared-normalized inner product of two arbitrary channel vectors, and the order statistics of the norms respectively, under the independent Rayleigh fading assumption. We first deal with the squared-normalized inner product.

*Lemma 3.2:* For any two users $i, j = 1, \ldots, K$, $i \neq j$, the squared-normalized inner product $\frac{|<\boldsymbol{h}_i, \boldsymbol{h}_j>|^2}{||\boldsymbol{h}_i||^2 ||\boldsymbol{h}_j||^2}$ of the respective channel vectors is distributed as the minimum of $(M-1)$ i.i.d. uniform random variables in [0,1]. In particular, when $M = 2$, the above quantity is uniform in [0,1]. The above statement also applies when the two users are selected based on their channel norms, since the norms and phases of the channel vectors are independent.

**Proof:** We prove the result for $M = 2$. The result can be proved for arbitrary values of $M$ using similar arguments. Since the phase between the vectors is independent of the norms, we may assume without loss of generality that $\boldsymbol{h}_i = [1 \quad 0]$ and $\boldsymbol{h}_j = [X_1 + iY_1 \quad X_2 + iY_2]$. Here $X_1, X_2, Y_1$, and $Y_2$ are i.i.d. normal variables. We are thus concerned with the behavior of the ratio $\frac{X_1^2+Y_1^2}{X_1^2+Y_1^2+X_2^2+Y_2^2}$. Using the fact that the square of a Gaussian variable is a Chi-squared variable and that the sum of two i.i.d. Chi-squared variables is exponential, we see that the squared-normalized inner product has the same distribution as $\frac{A}{A+B}$, where $A$ and $B$ are i.i.d. exponential variables. This is also the ratio of the first and second event times in a Poisson process; therefore, this quantity is seen to be uniform in [0,1]. In general, the squared-normalized inner product can be interpreted as the ratio of the first event time to the $M$-th event time in a Poisson process, and the result follows readily. □

Next, we turn our attention to the order statistics of the channel norms. The next lemma shows that the difference between the $L$-th largest and the maximum channel norm is asymptotically negligible in a certain sense compared to the $L$-th largest norm itself, as long as $L$ grows sufficiently slowly with $K$.

*Lemma 3.3:* Let $L(K)$ be a sequence such that $L(K) = \mathrm{o}(K^\delta)$ for any $0 < \delta < 1$ as $K \to \infty$ and $A, B, Q > 0$ positive constants. Then

$$\lim_{K\to\infty} \mathbb{E}\left[\log\left(A + Q||\boldsymbol{h}_{(1)}||^2\right)\right] - \mathbb{E}\left[\log\left(B + Q||\boldsymbol{h}_{(L)}||^2\right)\right] = 0.$$

**Proof:** See Appendix B. □

*Remark 3.2:* It is worth observing that the Rayleigh fading assumptions are not essential, and that with appropriate modifications the asymptotic results extend to a wider range of distributions of the channel vectors, as long as the phases between users are independent. We briefly mention a few important special cases. First of all, our results can be readily generalized to the case of Ricean fading. A second important case is the class of Increasing Failure Rate (IFR) distributions, with positive densities, for which the Erlang distribution is just a special case. (A distribution function $F(\cdot)$ with density $f(\cdot)$ is said to be IFR when $f(x)/(1 - F(x))$ is increasing.) Although the lognormal distribution is not IFR, its extremal behavior can be understood from that of the Gaussian, and it can be shown that in this case too our results apply.

## IV. Asymptotics

As mentioned earlier, the upper bound in (3) for $M = 2$ can be achieved when there is a pair of orthogonal users, each with the maximum channel norm $||\boldsymbol{h}_{(1)}||^2$. Intuitively, when the number of users is large, there exists with high probability a pair of users which are nearly orthogonal and have norms close to the maximum. This suggests that the sum rate can be closely approached by transmitting to a pair of 'good' users and allocating equal power to each of them. We are now ready to formalize this assertion.

We will consider three heuristic selection schemes. Scheme I selects two arbitrary users among the $L$ strongest ones. Scheme II selects an arbitrary user among the $L$ strongest ones, and a second one from the same group such that the sum rate achieved by the pair is maximized. Scheme III picks the best pair among the $L$ strongest users, i.e., it picks the pair that maximizes the sum rate. Note that scheme II dominates scheme I and that scheme III in turn dominates scheme II, and that all three schemes coincide when $L = 2$. In all the three schemes, the selected users are scheduled with equal power.

### *A. Rough large-$K$ asymptotics*

The next theorem considers ratio-asymptotics for scheme I. Specifically, it shows that the ratio of the rate obtained by using scheme I to the upper bound in (3) converges to unity as $K$ grows large.

*Theorem 4.1:* For any fixed value of $L \geq 2$,

$$\lim_{K\to\infty} \frac{\mathbb{E}\left[C(\boldsymbol{h}_{(i)}, \boldsymbol{h}_{(j)}, P)\right]}{\mathbb{E}\left[2\log\left(1 + \frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]} = 1, \qquad (5)$$

for all $i, j \leq L$, $i \neq j$.

**Proof:** It follows from Equation (3) and (4) that the ratio is smaller than one for any fixed $K$ and $L$. Thus, it suffices to show that the liminf of the ratio is larger than one as $K \to \infty$. It follows from Lemma 1.1 in Appendix A that

$$C(\boldsymbol{h}_{(i)}, \boldsymbol{h}_{(j)}, P) \geq 2\log\left(1 + \frac{P}{2}||\boldsymbol{h}_{(L)}||^2\right) + \log(U_{ij}),$$

with $U_{ij} := 1 - \frac{|<\boldsymbol{h}_{(i)}, \boldsymbol{h}_{(j)}>|^2}{||\boldsymbol{h}_{(i)}||^2 ||\boldsymbol{h}_{(j)}||^2}$.

Now, Lemma 3.2 implies

$$\mathbb{E}\left[\log(U_{ij})\right] = \int_{x=0}^{1} \log(x)\mathrm{d}x = \left[x(\log(x) - 1)\right]_{x=0}^{1} = -1.$$

The proof is then completed using Lemma 3.3 with $A = B = 1$, $Q = P/2$, and noting that $\mathbb{E}\left[\log\left(1 + \frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right] \to \infty$ as $K \to \infty$. □

The above results shows that scheme I (and consequently all the schemes) are asymptotically optimal in the ratio sense. However, it should be noted that this asymptotic result is too crude to capture the relative importance of the degree of orthogonality versus the magnitude of the channel vectors. Thus the ratio-asymptotics provide no indication of the relative performance of the various schemes and little guidance as to what a suitable choice of $L$ might be for a given finite value of $K$. Also, it is possible that there is a gap between the sum rate achieved by any of these schemes and the capacity limit, which cannot be discerned using ratio-asymptotics.

### B. Refined large-$K$ asymptotics

In order to discriminate among the various selection schemes and gain a better sense of the performance impact of the parameter $L$, we now proceed to consider finer asymptotics. In particular, we consider the *difference* between the sum rate and the upper bound for the capacity in (3).

*Theorem 4.2:* For any fixed value of $L \geq 2$, $l \leq L$, the difference

$$\mathbb{E}\left[2\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]-\mathbb{E}\left[\max_{k=1,\ldots,L,k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\right]$$

approaches $1/(L-1)$ as $K\to\infty$.

**Proof:** We first prove that the limsup of the difference is no larger than $1/(L-1)$.

Using Lemma 1.1 in Appendix A, we obtain

$$\begin{aligned}
&\max_{k=1,\ldots,L,k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\\
\geq\ & \max_{k=1,\ldots,L,k\neq l} 2\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(L)}||^2\right)+\log(U_{lk})\\
=\ & 2\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(L)}||^2\right)+\max_{k=1,\ldots,L,k\neq l}\log(U_{lk}),
\end{aligned}$$

with $U_{lk} := 1-\frac{|<\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)}>|^2}{||\boldsymbol{h}_{(l)}||^2||\boldsymbol{h}_{(k)}||^2}$.

For compactness, denote

$$\Delta(L) := \log\left(1+\frac{P}{2}||\boldsymbol{h}_{(L)}||^2\right)-\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right).$$

Then,

$$\begin{aligned}
&2\mathbb{E}\left[\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]-\mathbb{E}\left[\max_{k=1,\ldots,L,k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\right]\\
&\leq -2\mathbb{E}\left[\Delta(L)\right]-\mathbb{E}\left[\max_{k=1,\ldots,L,k\neq l}\log(U_{kl})\right].
\end{aligned}$$

Taking $A=B=1$, $Q=P/2$ in Lemma 3.3, it follows that $\limsup_{K\to\infty} -\mathbb{E}\left[\Delta(L)\right]=0$.

Using Lemma 3.2, a straightforward calculation yields

$$\begin{aligned}
\mathbb{E}\left[\max_{k=1,\ldots,L,k\neq l}\log(U_{kl})\right] &= \int_{x=0}^{1}\log(x)(L-1)x^{L-2}\mathrm{d}x\\
&=\left[x^{L-1}\left(\log(x)-\frac{1}{L-1}\right)\right]_{x=0}^{1}=-\frac{1}{L-1}.
\end{aligned}$$

We now show that the liminf of the difference is no smaller than $1/(L-1)$.

Using Lemma 1.4 in Appendix A, we obtain

$$\begin{aligned}
&\max_{k=1,\ldots,L,k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\\
\leq\ & \max_{k=1,\ldots,L,k\neq l} 2\log\left(\frac{1}{\epsilon}+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)+\log(\max\{\epsilon,U_{kl}\})\\
\leq\ & 2\log\left(\frac{1}{\epsilon}+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)+\log(\max\{\epsilon,\max_{k=1,\ldots,L,k\neq l}U_{kl}\}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&2\mathbb{E}\left[\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]-\mathbb{E}\left[\max_{k=1,\ldots,L,k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\right]\\
&\geq 2\mathbb{E}\left[\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]-2\mathbb{E}\left[\log\left(\frac{1}{\epsilon}+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]\\
&\quad -\mathbb{E}\left[\log\left(\max\{\epsilon,\max_{k=1,\ldots,L,k\neq l}U_{kl}\}\right)\right].
\end{aligned}$$

Taking $A=1$, $B=\frac{1}{\epsilon}$, $Q=P/2$ in Lemma 3.3, it follows that for any $\epsilon>0$,

$$\limsup_{K\to\infty}\left[\mathbb{E}\left[\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]-\mathbb{E}\left[\log\left(\frac{1}{\epsilon}+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right)\right]\right]=0.$$

Using Lemma 3.2, a straightforward computation yields

$$\begin{aligned}
&\mathbb{E}\left[\log\left(\max\{\epsilon,\max_{k=1,\ldots,L,k\neq l}U_{kl}\}\right)\right]\\
&= \int_{x=\epsilon}^{1}\log(x)(L-1)x^{L-2}\mathrm{d}x+\epsilon^{L-1}\log(\epsilon)\\
&= \left[x^{L-1}\left(\log(x)-\frac{1}{L-1}\right)\right]_{x=\epsilon}^{1}+\epsilon^{L-1}\log(\epsilon)\\
&= \frac{\epsilon^{L-1}-1}{L-1}.
\end{aligned}$$

Letting $\epsilon\downarrow 0$, the result follows. □

The above theorem shows that the asymptotic performance gap of scheme II decays as $1/(L-1)$, which suggests that a relatively moderate value of $L$ may be adequate for most practical purposes. The next corollaries follow as direct consequences from Theorem 4.2.

*Corollary 4.1:* For any fixed value $l$ and sequence $L(K)$ with $\lim_{K\to\infty} L(K)=\infty$,

$$\mathbb{E}\left[\mathcal{C}_{BC}(\boldsymbol{H},P)\right]-\mathbb{E}\left[\max_{k=1,\ldots,L(K),k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\right]\to 0$$

as $K\to\infty$.

The above corollary shows that scheme II is asymptotically optimal when sufficiently many users are considered. Because of the dominance relationship, it immediately follows that scheme III is asymptotically optimal as well. As a by-product, we conclude that the upper bound (3) is asymptotically tight.

*Corollary 4.2:*

$$\mathbb{E}\left[\mathcal{C}_{BC}(\boldsymbol{h}_1,\ldots,\boldsymbol{h}_K,P)\right]-\mathbb{E}\left[C(\boldsymbol{h}_{(1)},\boldsymbol{h}_{(2)},P)\right]\to 1$$

as $K\to\infty$.

The above corollary corresponds to a special case of scheme I, i.e., $L=2$. We see that simply selecting the two strongest users leaves a performance gap of 1 nats/symbol.

In conclusion, the above results show that scheme II is asymptotically optimal in the sense that the absolute gap with the sum capacity vanishes to zero provided $L(K) \to \infty$ as $K \to \infty$. Thus, transmitting to a suitably selected pair of users is asymptotically optimal, where one of the users may in fact be arbitrarily chosen from a fixed short list. The gain from considering all pairs of users, as in scheme III, is asymptotically negligible. However, picking an arbitrary pair of users, as in scheme I, is not optimal even the users are the two strongest ones.

## V. Numerical results

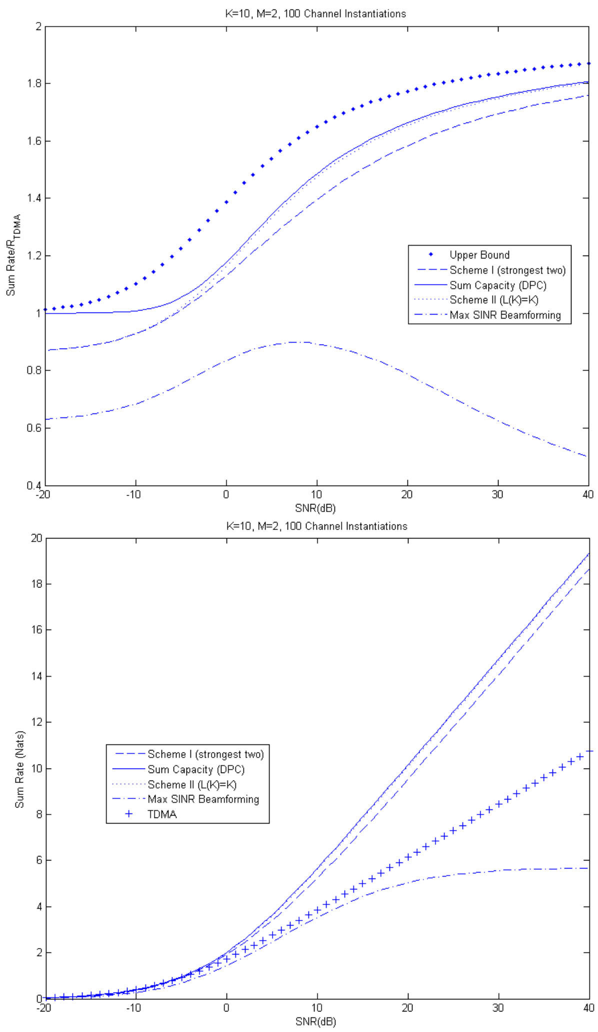


Fig. 2. (a) Comparison of various user selection heuristics with TDMA, $K = 10$ users. (b) Absolute sum rate in nats vs SNR, $K = 10$ users.

In this section, we compare the sum rate obtained by the various user selection schemes with the TDMA rate. We also make a comparison with a beam-forming (BF) scheme along the lines described in [12] and [18].

We present numerical results for a system with two transmit antennas and $K = 10$ users in Figure 2. The corresponding results for a system with $K = 25$ users are shown in Figure 3. In Figures 2(a) and 3(a), we plot the ratio of the sum rate obtained by the various schemes to the TDMA sum rate,

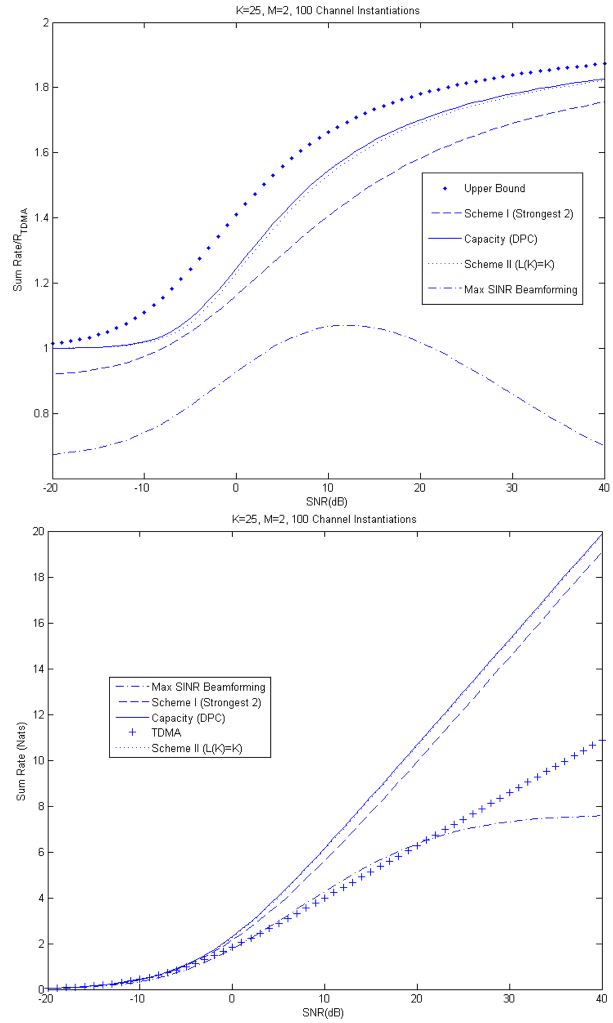


Fig. 3. (a) Comparison of various user selection heuristics with TDMA, $K = 25$ users. (b) Absolute sum rate in nats vs SNR, $K = 25$ users.

versus the SNR (in dB). The results shown here were average over 100 channel realizations. The solid line corresponds to the optimal DPC scheme. The dotted line just underneath the solid line corresponds to a special case of scheme II. Specifically, we schedule the user with the largest channel norm, and the second user to form the best possible pair with it (i.e., $L(K) = K$). It is clear that even for this moderate value of $K$, scheme II performs very well, in addition to being asymptotically optimal. The broken line corresponds to a special case of scheme I, where the two strongest users are scheduled with equal power. It is clear that scheme II dominates scheme I quite significantly. It is also interesting to note that the upper bound in (3) (shown in the Figures 2(a) and 3(a) with diamonds), although asymptotically tight, is quite loose for practical values of $K$ and SNR. We finally observe that TDMA is optimal in the very low SNR regime. The absolute sum rate (in nats) for the two scenarios are graphed as a function of SNR in Figures 2(b) and 3(b).

The BF scheme proposed in [12] selects two users which have the best Signal-to-Interference-and-Noise Ratio (SINR) on each of the antennas. In particular, the transmitter forms random beams along the direction of two orthonormal vectors $\phi_1$ and $\phi_2$, and selects two users $k_m^* := \arg\max_{k=1,\dots,K} SINR_{k,m}$, $m = 1, 2$, where

$$SINR_{k,m} := \frac{| < \boldsymbol{h}_k, \phi_m > |^2}{2/P + | < \boldsymbol{h}_k, \phi_{3-m} > |^2}.$$

The expected sum rate obtained (ignoring potential complications when $k_1^* = k_2^*$), is therefore

$$R_{BF} := \mathbb{E}\left[\log(1 + SNR_{k_1^*,1}) + \log(1 + SNR_{k_2^*,2})\right].$$

The lower curves in Figures 2 and 3 plot the sum rate of this BF scheme compared with the other schemes. We observe that transmitting along two pre-determined beams without using actual phase information performs poorly, even though it is known to be asymptotically optimal in the limit of a large number of users. However, a plot of the quantity $C(\boldsymbol{h}_{k_1^*}, \boldsymbol{h}_{k_2^*})$ (*not* shown in the figure) revealed that this particular scheme does well in terms of *selecting* a pair of users.

Note that as $P \downarrow 0$, we have

$$R_{BF} \approx \frac{P}{2}\mathbb{E}\left[| < \boldsymbol{h}_{k_1^*}, \phi_1 > |^2 + | < \boldsymbol{h}_{k_2^*}, \phi_2 > |^2\right] =$$
$$P\mathbb{E}\left[| < \boldsymbol{h}_{k_1^*}, \phi_1 > |^2\right] \leq P\mathbb{E}\left[||\boldsymbol{h}_{(1)}||^2\right] \approx R_{TDMA}.$$

Denoting $g_{ij} := | < \boldsymbol{h}_{k_i^*}, \phi_j > |^2$, we find that $R_{BF}$ approaches

$$\mathbb{E}\left[\log\left(1 + \frac{g_{11}}{g_{12}}\right) + \log\left(1 + \frac{g_{22}}{g_{21}}\right)\right] = 2\mathbb{E}\left[\log\left(1 + \frac{g_{11}}{g_{12}}\right)\right].$$

as $P \to \infty$. This shows that for any fixed number of users, the sum rate of the BF scheme saturates at a finite value as the transmit power becomes large, as is shown in Figures 2(b) and 3(b). In contrast, the TDMA sum rate $R_{TDMA}$ grows without bound, albeit slowly.

## VI. Future work

Several natural topics for further investigation present themselves. First of all, the above results have evident implications for the design of channel feedback mechanisms and transmission techniques. It would be interesting to address these aspects in more detail. A second major avenue that would be worth pursuing is to generalize the results to an arbitrary number of transmit antennas, and possibly several receive antennas. A further challenging subject that is under ongoing investigation, concerns the extension to a scenario with heterogeneous users and maximizing a *weighted* sum rate or achieving an optimal fair operating point of the capacity region. Some interesting results along the latter lines may be found in [10], [17].

## Acknowledgment

The authors wish to express their gratitude to Gerhard Kramer and Hanan Weingarten for many helpful suggestions and interesting discussions. The second author also gratefully acknowledges insightful comments of Nihar Jindal and Harish Viswanathan.

## Appendix

### A. Bounds for the sum capacity

Here, we gather a few further bounds for the sum capacity that are useful in proving Theorems 4.1 and 4.2.

We first prove a lower bound.

*Lemma 1.1:* For any $i, j$,

$$C(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \quad \geq \quad 2\log\left(1 + \frac{P}{2}||\boldsymbol{h}_{i\wedge j}||^2\right) + \log(U_{ij}),$$

with $||\boldsymbol{h}_{i\wedge j}||^2 := \min\{||\boldsymbol{h}_i||^2, ||\boldsymbol{h}_j||^2\}$.

**Proof:** By definition,

$$\begin{aligned}
& C(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \\
= \ & \log\left(1 + \frac{P}{2}\left(||\boldsymbol{h}_i||^2 + ||\boldsymbol{h}_j||^2\right) + \frac{P^2}{4}||\boldsymbol{h}_i||^2||\boldsymbol{h}_j||^2 U_{ij}\right) \\
\geq \ & \log\left[\left(1 + \frac{P}{2}||\boldsymbol{h}_i||^2\right)\left(1 + \frac{P}{2}||\boldsymbol{h}_j||^2\right)U_{ij}\right] \\
\geq \ & 2\log\left(1 + \frac{P}{2}||\boldsymbol{h}_{i\wedge j}||^2\right) + \log U_{ij}.
\end{aligned}$$

□

We now turn to some upper bounds. Define

$$F(\boldsymbol{h}_i, \boldsymbol{h}_j, P) := 1 + P||\boldsymbol{h}_{i\vee j}||^2 + \frac{P^2}{4}||\boldsymbol{h}_{i\vee j}||^4 U_{ij},$$

with $||\boldsymbol{h}_{i\vee j}||^2 := \max\{||\boldsymbol{h}_i||^2, ||\boldsymbol{h}_j||^2\}$.

*Lemma 1.2:* For any $i, j$,

$$\mathcal{C}_{BC}(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \leq \log(F(\boldsymbol{h}_i, \boldsymbol{h}_j, P)).$$

**Proof:** Using Equation (2) and Lemma 3.1,

$$\begin{aligned}
& \mathcal{C}_{BC}(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \\
= \ & \max_{P_i + P_j \leq P} \log\left(1 + P_i||\boldsymbol{h}_i||^2 + P_j||\boldsymbol{h}_j||^2 + P_iP_j||\boldsymbol{h}_i||^2||\boldsymbol{h}_j||^2 U_{ij}\right) \\
\leq \ & \max_{P_i + P_j \leq P} \log\left(1 + (P_i + P_j)||\boldsymbol{h}_{i\vee j}||^2 + P_iP_j||\boldsymbol{h}_{i\vee j}||^4 U_{ij}\right) \\
= \ & \log\left(1 + P||\boldsymbol{h}_{i\vee j}||^2 + \frac{P^2}{4}||\boldsymbol{h}_{i\vee j}||^4 U_{ij}\right).
\end{aligned}$$

□

*Lemma 1.3:* For any $i, j$, $\epsilon \in (0, 1)$,

$$F(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \leq \left(\frac{1}{\epsilon} + \frac{P}{2}||\boldsymbol{h}_{i\vee j}||^2\right)^2 \max\{\epsilon, U_{ij}\}.$$

**Proof:** By definition, for any $\epsilon \in (0, 1)$,

$$\begin{aligned}
F(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \ &= \ 1 + P||\boldsymbol{h}_{i\vee j}||^2 + \frac{P^2}{4}||\boldsymbol{h}_{i\vee j}||^4 U_{ij} \\
&\leq \ \left(\frac{1}{\epsilon} + \frac{P}{2}||\boldsymbol{h}_{i\vee j}||^2\right)^2 \max\{\epsilon, U_{ij}\}.
\end{aligned}$$

□

*Lemma 1.4:* For any $i, j$, $\epsilon \in (0, 1)$,

$$C(\boldsymbol{h}_i, \boldsymbol{h}_j, P) \leq 2\log\left(\frac{1}{\epsilon} + \frac{P}{2}||\boldsymbol{h}_{i\vee j}||^2\right) + \log(\max\{\epsilon, U_{ij}\}).$$

**Proof:** Follows from (4) and Lemmas 1.2 and 1.3. □

## *B. Proof of Lemma 3.3*

**Proof:** First note that the difference is bounded from below by

$$\mathbb{E}\left[\log\left(1+\frac{A-B}{B+Q||\boldsymbol{h}_{(1)}||^2}\right)\right],$$

so the liminf is non-negative since $||\boldsymbol{h}_{(1)}||^2 \to \infty$ a.s. as $K \to \infty$.

We now show that the limsup is non-positive. Denoting $m_K := \mathbb{E}\left[||\boldsymbol{h}_{(1)}||^2\right]$ and applying Jensen's inequality, we obtain

$$\mathbb{E}\left[\log\left(A+Q||\boldsymbol{h}_{(1)}||^2\right)\right] \leq \log(A+Qm_K).$$

For any $\epsilon > 0$,

$$\mathbb{E}\left[\log\left(B+Q||\boldsymbol{h}_{(L)}||^2\right)\right] \geq \log(B)+$$

$$\log\left(B+m_K(1-\epsilon)\right)-\log(B)]\mathbb{P}\{||\boldsymbol{h}_{(L)}||^2 \geq m_K(1-\epsilon)\}.$$

Since $L(K) = \mathrm{o}(K^\delta)$ for any $\delta > 0$, it follows that

$$\liminf_{K\to\infty} \mathbb{P}\{||\boldsymbol{h}_{(L)}||^2 \geq m_K(1-\epsilon)\} \geq$$

$$\liminf_{K\to\infty} \mathbb{P}\{||\boldsymbol{h}_{(K^{\epsilon/4})}||^2 \leq m_K(1-\epsilon)\}.$$

It can be shown (proof omitted for brevity) that there exists a constant $C_{\epsilon/4,\epsilon}$ such that

$$\mathbb{P}\{||\boldsymbol{h}_{(K^{\epsilon/4})}||^2 \leq m_K(1-\epsilon)\} \leq \frac{C_{\epsilon/4,\epsilon}}{(\log(K))^2}.$$

Combining the above inequalities and observing that $\log(1+Qm_k(1-\epsilon)) = \mathrm{o}((\log(K))^2)$ as $K \to \infty$, we deduce that the limsup is bounded from above by

$$\limsup_{K\to\infty} \log(A+Qm_K) - \log(B+Qm_K(1-\epsilon)).$$

The latter quantity is no larger than

$$\lim_{K\to\infty} \log\left(1+\frac{A-B}{B+Qm_K}\right) - \log(1-\epsilon) = -\log(1-\epsilon),$$

because $m_K \to \infty$ as $K \to \infty$.

Letting $\epsilon \downarrow 0$, the result follows. □

## References


[1] G. Caire, S. Shamai (2003). On the achievable throughput of a multi-antenna Gaussian broadcast channel. *IEEE Trans. Inf. Theory* **49** (7), 1691–1706.

[2] M. Costa (1983). Writing on dirty paper. *IEEE Trans. Inf. Theory* **29** (3), 439–441.

[3] B.M. Hochwald, C.B. Peel, A.L. Swindlehurst (2005). A vector-perturbation technique for near-capacity multiantenna multiuser communication – part I: channel inversion and regularization. *IEEE Trans. Commun.* **53** (1), 195–202.

[4] B.M. Hochwald, C.B. Peel, A.L. Swindlehurst (2005). A vector-perturbation technique for near-capacity multiantenna multiuser communication – part II: perturbation. *IEEE Trans. Commun.* **53** (3), 537–544.

[5] N. Jindal (2005). A high SNR analysis of MIMO broadcast channels. Preprint.

[6] N. Jindal (2005). MIMO broadcast channels with finite rate feedback. Preprint.

[7] N. Jindal, A. Goldsmith (2005). Dirty-paper coding versus TDMA for MIMO broadcast channels. *IEEE Trans. Inf. Theory* **51** (5), 1783–1794.

[8] N. Jindal, W. Rhee, S. Vishwanath, S.A. Jafar, A. Goldsmith (2005). Sum power iterative water-filling for multi-antenna Gaussian broadcast channels. *IEEE Trans. Inf. Theory* **51** (4), 1570–1580.

[9] N. Jindal, S. Vishwanath, A. Goldsmith (2004). On the duality of Gaussian multiple-access and broadcast channels. *IEEE Trans. Inf. Theory* **50** (5), 768–783.

[10] Lee, J., Jindal, N. (2005). Symmetric capacity of MIMO downlink channels. Preprint.

[11] L. Li, A. Goldsmith (2001). Capacity and optimal resource allocation for fading broadcast channels – Part I: Ergodic capacity. *IEEE Trans. Inf. Theory* **47** (3), 1083–1102.

[12] M. Sharif, B. Hassibi (2005). On the capacity of MIMO broadcast channel with partial side information. *IEEE Trans. Inf. Theory* **51** (2), 506–522.

[13] C. Swannack, E. Uysal-Biyikoglu, G. Wornell (2004). Low-complexity multi-user scheduling for maximizing throughput in the MIMO broadcast channel. In: *Proc. 42nd Annual Allerton Conf. Commun. Control, Comput.*

[14] C. Swannack, E. Uysal-Biyikoglu, G. Wornell (2004). Finding NEMO: near-orthogonal sets for multiplexing and opportunistic scheduling in MIMO broadcast. Preprint.

[15] S. Vishwanath, N. Jindal, A. Goldsmith (2003). Duality, achievable rates and sum-rate capacity of MIMO broadcast channels. *IEEE Trans. Inf. Theory* **49** (10), 2658–2668.

[16] P. Viswanath, D.N.C. Tse (2003). Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Trans. Inf. Theory* **49** (8), 1912–1921.

[17] H. Viswanathan, S. Venketesan, H. Huang. Downlink capacity evaluation of cellular networks with known interference cancellation. *IEEE J. Sel. Areas Commun.* **21** (5), 802–811.

[18] P. Viswanath, D.N.C. Tse, R. Laroia (2002). Opportunistic beamforming using dumb antennas. *IEEE Trans. Inf. Theory* **48** (6), 1277–1294.

[19] H. Weingarten, Y. Steinberg, S. Shamai (2005). The capacity region of the Gaussian MIMO broadcast channel. *IEEE Trans. Inf. Theory*, to appear.