Scheduling of Multi-Antenna Broadcast Systems with Heterogeneous Users

Krishna Jagannathan, Sem Borst, Phil Whiting and Eytan Modiano

***Abstract*— We study the problem of efficiently scheduling users in a Gaussian broadcast channel with M transmit antennas and K independent receivers, each with a single antenna. We first focus on a scenario with two transmit antennas and statistically identical users, and analyze 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 selects a second user from the next $L-1$ strongest ones to form the best pair, taking channel orientations into account as well. We prove that the expected rate gap converges to $1/(L-1)$ nats/symbol when the total number of users K tends to infinity. Allowing L to increase with K, it may be deduced that transmitting to a properly chosen pair of users is asymptotically optimal, while considerably reducing the feedback overhead and scheduling complexity. Next, we tackle the problem of maximizing a *weighted* sum rate in a scenario with heterogeneous user characteristics. We establish a novel upper bound for the weighted sum capacity, which we then use to show that the maximum expected weighted sum rate can be asymptotically achieved by transmitting to a suitably selected subset of at most MC users, where C denotes the number of distinct user classes. Numerical experiments indicate that the asymptotic results are remarkably accurate and that the proposed schemes operate close to absolute performance bounds, even for a moderate number of users.**

***Index Terms*— Broadcast channels, MIMO systems, scheduling.**

I. Introduction

IN THE present paper we consider the downlink transmission from a single base station equipped with M transmit antennas to K independent users each with a single receive antenna. In information-theoretic terms, this may be modeled as a multi-antenna Broadcast Channel (BC). Caire & Shamai [1] were the first to obtain the sum capacity expression for the Gaussian BC with two receivers, and to suggest the use of Dirty Paper Coding (DPC) [2] for transmitting over this channel. Viswanath & Tse [24] and Vishwanath *et al.* [23] 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.* [12]. Recently, Weingarten *et al.* [28] showed that DPC in fact achieves the full capacity region of the multi-antenna Gaussian BC, thus providing a characterization of the entire capacity region.

Various researchers have investigated the sum capacity gains achievable in the above-described system by simultaneously transmitting to several users. In particular, Jindal & Goldsmith [10] show that the sum capacity 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 [8] demonstrates that the sum capacity grows with the SNR at rate $\min\{M, K\}$. In other words, multiple transmit antennas can potentially provide an M-fold gain in the sum capacity.

The above capacity results rely on the assumption that perfect channel state information (CSI) is available at the transmitter, which may involve a significant amount of feedback overhead. In addition, DPC is quite a sophisticated technique and challenging to implement in an actual system. Motivated by these issues, extensive efforts have been made to devise practical transmission and coding schemes and find ways to reduce the amount of channel feedback information required. Hochwald *et al.* [4], [5] describe an algorithm based on channel inversion and sphere encoding, and demonstrate that it closely approaches the sum capacity while being simpler to operate than DPC. Jindal [9] considers a multi-antenna BC with limited channel feedback information, and shows that the full sum capacity gain at high SNR values is achievable as long as the number of feedback bits grows linearly with the SNR (in dB).

As mentioned above, multiple transmit antennas can potentially yield an M-fold increase in the sum capacity. However, it is necessary that at least M users are served simultaneously in order to reap the full benefits. Transmitting to fewer than M users falls short of the maximum rate as it fails to fully exploit the available degrees of freedom. Transmitting to more than M users may be necessary to achieve the sum capacity in general, but the upper bound in [10] suggests that transmitting to a suitably selected subset of 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 [19], [20].

Clearly, the above principle allows for a reduction of the amount of channel feedback and coding complexity. In particular, it suggests beam-forming schemes which construct M orthogonal beams and serve the users with the largest channel gains on each of them with equal power. Transmis-

Manuscript received June 1, 2006; revised December 21, 2006. The work of Modiano and Jagannathan was supported by NSF ITR grant CCR-0325401, by DARPA/AFOSR through the University of Illinois grant no. F49620-02-1-0325, by NASA Space Communication Project grant number NAG3-2835, and by a grant from Draper Laboratory.

Krishna Jagannathan and Eytan Modiano are with the Massachusetts Institute of Technology, Cambridge, MA 02139 (e-mail: {krishnaj,modiano}@mit.edu).

Sem Borst and Phil Whiting are with Alcatel-Lucent Bell Labs, P.O. Box 94079, Murray Hill, NJ 07974 (e-mail: {sem,pwhiting}@alcatel-lucent.com).

Sem Borst is also affiliated with the Department of Mathematics & Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Digital Object Identifier 10.1109/JSAC.2007.070915.

sion schemes along these lines are presented in Viswanath *et al.* [25], Sharif & Hassibi [15], and Vakili *et al.* [22]. Viswanathan & Kumaran [26] proposed fixed-beams and adaptive steerable-beams schemes grounded on that principle as well. Further related results may be found in Sharif & Hassibi [16], [17] who derive the asymptotic sum capacity for TDMA, DPC and beam-forming in the limit where the number of users grows large.

In the present paper, we propose scheduling schemes that transmit only to a small subset of users with favorable channel characteristics, and provide near-optimal performance when the total number of users to choose from is large. Extensive numerical experiments reveal that the scheduling schemes in fact operate remarkably close to absolute performance bounds, even when the number of users is fairly moderate. Since the proposed schemes only transmit to a small fraction of the users, they provide significant scope for reducing the feedback overhead and scheduling complexity.

In a recent paper, Yoo & Goldsmith [29] show that employing zero-forcing beam-forming to a set of M nearly orthogonal users with large channel norms is asymptotically optimal as the number of users grows large. The similarity between the user selection algorithm proposed in the above paper and our scheduling schemes for a homogeneous system will become apparent in Section III. However, there are also important differences to be emphasized. The authors of [29] are concerned primarily with circumventing the complexities involved in implementing a DPC encoder, and propose zero-forcing beamforming as a more practical alternative. It is worth noting that their method requires perfect CSI at the transmitter. In contrast, we are interested in reducing the amount of channel feedback by transmitting only to a small group of users. As far as the actual encoding is concerned, we assume that a technique such as the one proposed in [4], [5] can be used to achieve rates very close to DPC, once the optimal set of users to serve has been determined.

We first focus on a simple scenario with two transmit antennas and statistically identical users, and analyze 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 next $L-1$ strongest ones to form the best possible pair with it, taking channel angles into account as well. We prove that the expected rate gap converges to $1/(L-1)$ nats/symbol 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 maximum expected sum rate is achievable by transmitting to a properly chosen pair 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. We remark that our scheme requires full channel feedback (i.e., both magnitude and phase information) only from the L strongest users. Finding the users with the largest channel gains can be accomplished using simple thresholding schemes wherein only the users with good channel gains are required to provide feedback.

Next, we turn our attention to a more general system with M transmit antennas and heterogeneous user characteristics. For a heterogeneous system, the sum capacity is no longer an appropriate performance metric, because it does not reflect the potential fairness issues that arise. Hence, we will focus on maximizing a *weighted* sum rate, where the users with weaker channels would typically be assigned higher weights. Leaving fairness considerations aside, maximizing a weighted sum rate is also of critical importance in so-called queue-based scheduling strategies where the user weights are taken to be functions of the respective queue lengths. Queue-based scheduling strategies are particularly attractive because under mild assumptions they are known to achieve stability whenever feasible without explicit knowledge of the system parameters, see for instance [14], [18], [21].

Although the sum rate expression for the multi-antenna Gaussian BC and associated bounds have been thoroughly investigated, the problem of maximizing a general function over the capacity region has not attracted nearly as much attention. To the best of our knowledge, Viswanathan *et al.* [27] are among the few authors who consider the problem of attaining more general points on the boundary of the capacity region. In particular, they present an algorithm for finding the power allocation to achieve any weighted sum rate maximizing point. However, the optimization procedure is computationally demanding, especially for large numbers of users, and requires perfect CSI. Lee & Jindal [13] study the problem of obtaining the symmetric capacity, i.e., the maximum rate that can be provided to each of the users simultaneously. Vakili *et al.* [22] address the problem of differentiated rate scheduling, i.e., maximizing the sum rate subject to certain target ratios for the long-term expected rates received by users of various classes. The results and schemes proposed in [22] rely on the fact that as the number of users grows large, the rates can be divided in arbitrary proportions through time sharing among various classes, while retaining the maximum sum rate and only transmitting to groups of M users at a time, provided the users have identical channel statistics. As it turns, the latter no longer applies when the users have distinct channel statistics, and in particular transmitting to M users at a time no longer suffices then to achieve the maximum (weighted) sum rate.

Here, we consider an M-antenna broadcast system with a user population that consists of C distinct classes, where each class is assigned a non-negative weight. In this setting, we derive a generic upper bound for the weighted sum capacity, which includes as a special case the sum capacity bound in [10]. We then proceed to show that the upper bound is in fact attained for certain 'ideal' configurations of MC channel vectors. Finally, we prove that a nearly ideal configuration of such channel vectors exists with high probability, and that the maximum expected weighted sum rate can thus be asymptotically achieved, when the total number of users grows large.

The remainder of the paper is organized as follows. In Section II we present a detailed model description and review some relevant results for the capacity region of the Gaussian multi-antenna BC. In Section III we focus on the problem of maximizing the sum rate in a system with homogeneous users. Section IV addresses the weighted sum rate maximization problem in a scenario with heterogeneous users. In Section V we discuss the numerical experiments we conducted, which

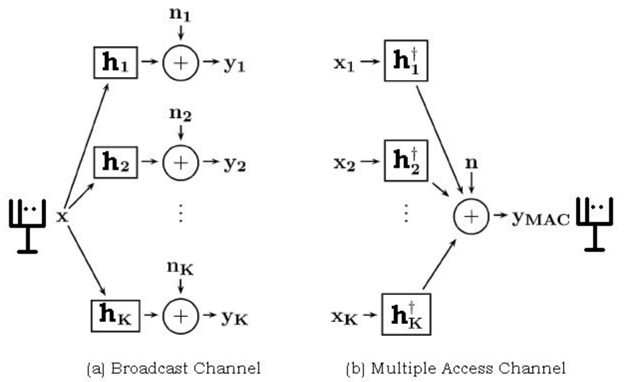

Fig. 1. The multi-antenna BC (left) and the MAC (right) have the same capacity region.

indicate that the asymptotics are surprisingly accurate, even for a moderate number of users. Throughout the paper, we omit most of the proofs due to page constraints, and refer the reader to [6] for detailed proofs and other supplementary material.

II. Model description and capacity region

A. Model description

We consider a broadcast channel (BC) with $M > 1$ transmit antennas and K 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 fixed. We temporarily assume that the transmitter has perfect CSI, i.e., exact knowledge of the matrix $\boldsymbol{H}$. The noise at each of the receivers is assumed to be circularly symmetric complex Gaussian, distributed according to $\mathcal{CN}(0,1)$. Thus the received signal at the k-th receiver is $y_k = \boldsymbol{h}_k\boldsymbol{x} + n_k$. 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 implies $Tr(\Sigma_x) \le P$. (Here Tr denotes the trace operator, which is the sum of the diagonal elements of a square matrix.)

B. Known information-theoretic results

Now we review some known results regarding the capacity region and the sum capacity of the multi-antenna Gaussian BC.

Let $\pi(k)$, $k = 1,\ldots,K$, be a permutation of $k = 1,\ldots,K$. As shown in [23], the following rate vector is achievable using Dirty Paper Coding (DPC):

$$R_{\pi(k)} = \log\left(\frac{1 + \boldsymbol{h}_{\pi(k)}(\sum_{l\le k}\Sigma_{\pi(l)})\boldsymbol{h}_{\pi(k)}^\dagger}{1 + \boldsymbol{h}_{\pi(k)}(\sum_{l<k}\Sigma_{\pi(l)})\boldsymbol{h}_{\pi(k)}^\dagger}\right), \quad k = 1,\ldots,K.$$

The DPC region is defined as the closed convex hull of the union of all such rate vectors, over all positive semi-definite covariance matrices that satisfy the power constraint $\sum_{k=1}^K Tr(\Sigma_{\pi(k)}) \le P$, and over all possible permutations $\pi(k)$. As shown in [1], [28], DPC in fact achieves the entire capacity region denoted as $\mathscr{C}_{BC}$. The weighted sum capacity $C_{BC}^{\mathbf{w}}(\boldsymbol{H},P)$ for any weight vector $\mathbf{w} \in \mathbb{R}_+^K$ can therefore be written as in equation (1).

Unfortunately, the maximization in (1) involves a non-concave function of the covariance matrices, which makes it hard to deal with analytically as well as numerically. However, in [23], [24], 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.

Let $S_k := \sum_{l=1}^k R_l$ be the partial sum rate of the first k users. Note that the weighted capacity can be written in terms of the partial sum rates as $C_{BC}^{\mathbf{w}}(\boldsymbol{H},P) = \sum_{k=1}^K \Delta w_k S_k$, where $\Delta w_k := w_k - w_{k+1}$, with the convention that $w_{K+1} = 0$. Without loss of generality we may assume that $w_1 \ge w_2 \ge \cdots \ge w_K$. Using the duality result, the weighted sum capacity (1) of the BC can be rewritten in terms of the dual MAC weighted sum rate. Given the above ordering of the weights, the optimal decoding order for the dual MAC is $K, K-1,\ldots,1$. In other words, user k experiences interference only from users $1,2,\ldots,k-1$, for each $k = 1,\ldots,K$. This ordering gives rise to the following expression for the weighted sum capacity of the BC:

$$C_{BC}^{\mathbf{w}}(\boldsymbol{H},P) = \max_{\sum_{k=1}^K P_k \le P} \sum_{k=1}^K \Delta w_k \log\det\left(I_M + \sum_{l=1}^k P_l \boldsymbol{h}_l^\dagger \boldsymbol{h}_l\right), \quad (2)$$

where $P_k \ge 0$ denotes the power allocated to the k-th receiver. As a special case of (2) with $w_k = 1$, $k = 1,\ldots,K$, the sum capacity is obtained as

$$C_{BC}^{\text{sum}}(\boldsymbol{H},P) = \max_{\sum_{k=1}^K P_k \le P} \log\det\left(I_M + \sum_{k=1}^K P_k \boldsymbol{h}_k^\dagger \boldsymbol{h}_k\right). \quad (3)$$

Since $\log\det(\cdot)$ is a concave function on the set of positive-definite matrices, the problems in (2) and (3) only involve maximizing a concave objective function subject to convex constraints. Specialized algorithms have been developed to solve these problems [11], [27].

III. Homogeneous users

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

We will show that the sum capacity can be closely approached by transmitting to a suitably selected pair of users as the total number of users grows large. In preparation for that, we first present some useful lower and upper bounds for the sum capacity.

$$C_{BC}^{\mathbf{w}}(\boldsymbol{H},P) = \max_{\boldsymbol{R}\in\mathscr{C}_{BC}} \sum_{k=1}^{K} w_k R_k = \max_{\pi} \max_{\sum_{k=1}^{K} Tr(\Sigma_{\pi(k)})\leq P} \sum_{k=1}^{K} w_{\pi(k)} \log\left(\frac{1+\boldsymbol{h}_{\pi(k)}(\sum_{l\leq k}\Sigma_{\pi(l)})\boldsymbol{h}_{\pi(k)}^{\dagger}}{1+\boldsymbol{h}_{\pi(k)}(\sum_{l<k}\Sigma_{\pi(l)})\boldsymbol{h}_{\pi(k)}^{\dagger}}\right). \quad (1)$$

A. Bounds for the sum capacity

Denote by $\boldsymbol{h}_{(k)}$ the channel vector of the receiver with the k-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 [10]:

$$C_{BC}^{\text{sum}}(\boldsymbol{H},P) \leq M\log\left(1+\frac{P}{M}||\boldsymbol{h}_{(1)}||^2\right). \quad (4)$$

Observe that the above bound can be achieved when there are M receivers with orthogonal channel vectors tied for the maximum norm $||\boldsymbol{h}_{(1)}||^2$. For the remainder of this section, we focus on the case of $M = 2$ transmit antennas. The upper bound for the sum capacity in (4) then becomes

$$C_{BC}^{\text{sum}}(\boldsymbol{H},P) \leq 2\log\left(1+\frac{P}{2}||\boldsymbol{h}_{(1)}||^2\right). \quad (5)$$

Taking $P_i = P_j = P/2$ and $P_k = 0$ for all $k \neq i, j$ in Equation (3), we obtain a simple lower bound for the sum capacity

$$C_{BC}^{\text{sum}}(\boldsymbol{H},P) \geq C(\boldsymbol{h}_i,\boldsymbol{h}_j,P) := \log\det\left(I_2+\frac{P}{2}(\boldsymbol{h}_i^{\dagger}\boldsymbol{h}_i+\boldsymbol{h}_j^{\dagger}\boldsymbol{h}_j)\right), \quad (6)$$

which corresponds to transmitting to users i and j at equal power.

For any two vectors $\boldsymbol{g},\boldsymbol{h} \in \mathbb{C}^2$, let $U(\boldsymbol{g},\boldsymbol{h}) := \frac{|<\boldsymbol{g},\boldsymbol{h}>|^2}{||\boldsymbol{g}||^2||\boldsymbol{h}||^2}$ be the squared normalized inner product. By expanding the determinant in (6), we obtain

$$C(\boldsymbol{h}_i,\boldsymbol{h}_j,P) = \log\left(1+\frac{P}{2}(||\boldsymbol{h}_i||^2+||\boldsymbol{h}_j||^2)+\frac{P^2}{4}||\boldsymbol{h}_i||^2||\boldsymbol{h}_j||^2 V_{ij}\right), \quad (7)$$

with $V_{ij} = 1 - U(\boldsymbol{h}_i,\boldsymbol{h}_j)$.

The lower bound expression (7) reflects the fact that the sum rate for two users critically depends on the norms of the respective channel vectors and their degree of orthogonality. In particular, the sum rate is large when the channel vectors are nearly orthogonal and have large norms. Indeed, the lower bound coincides with the upper bound (5) when users i and j are orthogonal and tied for the maximum norm, i.e., $||\boldsymbol{h}_i||^2 = ||\boldsymbol{h}_j||^2 = ||\boldsymbol{h}_{(1)}||^2$ and $<\boldsymbol{h}_i,\boldsymbol{h}_j>= 0$.

B. Random channel vectors

The lower and upper bounds for the sum capacity in the previous subsection hold for any arbitrary but fixed set of channel vectors. In order to derive meaningful asymptotic results, we will in the remainder of the section assume the channel vectors to be random and focus on the *expected* sum rate. We will adhere to the common assumption that the components of the channel vectors of the various users to be independent and distributed according to $\mathcal{CN}(0,1)$, which corresponds to independent Rayleigh fading, although this assumption is actually not that essential for most of the results to hold.

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 expected sum rate 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.

As mentioned earlier, the two-user sum rate critically depends on the norms of the channel vectors and their squared normalized inner product, and the statistical properties of these two quantities will therefore play a crucial role. The next lemma characterizes the distribution of the squared normalized inner product of two arbitrary channel vectors.

Lemma 3.1: For any two users $i, j = 1,\ldots,K$, $i \neq j$, the squared normalized inner product $U(\boldsymbol{h}_i,\boldsymbol{h}_j) := \frac{|<\boldsymbol{h}_i,\boldsymbol{h}_j>|^2}{||\boldsymbol{h}_i||^2||\boldsymbol{h}_j||^2}$ is independent of the norms of the respective channel vectors and 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].

We now turn the 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, as long as L grows sufficiently slowly with K.

Lemma 3.2: Let $L(K)$ be a sequence such that $L(K) = o(K^{\delta})$ as $K \to \infty$ for *every* $0 < \delta < 1$ (e.g., $L(K) = \log(K)$, but not $L(K) = \sqrt{K}$) and $A, B, Q > 0$ positive constants. Then

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

C. Large-K asymptotics

As mentioned earlier, the upper bound in (5) for the sum capacity can be achieved when there is a pair of orthogonal users, each with the maximum channel norm $||\boldsymbol{h}_{(1)}||^2$. Intuitively, when the total 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 capacity can be closely approached by transmitting to such a pair of users and allocating equal power to each of them.

We are now ready to formalize the above assertion. We will consider three heuristic selection schemes for scheduling a pair of users with equal power. Scheme I picks two arbitrary users among the L strongest ones. Scheme II selects an arbitrary user among the L strongest ones[1], and a second one from the same group to form the best pair, i.e., the pair that maximizes the sum rate. Scheme III picks the best pair among the L strongest users, i.e., 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 $L = 2$.

[1] For example, we may choose the user with the largest channel norm.

Theorem 3.1: For any fixed value of $L \geq 2$, $l \leq L$, the difference

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

converges to $\frac{1}{L-1}$ as $K \to \infty$.

Theorem 3.2: For any fixed value l and sequence $L(K)$ with $\lim_{K\to\infty} L(K) \to \infty$, the difference

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

tends to zero, as $K \to \infty$.

We now present some corollaries, which are direct consequences of Theorems 3.1 and 3.2.

Corollary 3.1: The upper bound for the sum capacity (5) is asymptotically tight.

Corollary 3.2: For any fixed value of L, $l \leq L$,

$$\mathbb{E}\left[C_{BC}^{\text{sum}}(\boldsymbol{H},P)\right]-\mathbb{E}\left[\max_{k=1,\dots,L,k\neq l} C(\boldsymbol{h}_{(l)},\boldsymbol{h}_{(k)},P)\right] \to \frac{1}{L-1}$$

as $K \to \infty$.

The above corollary 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.

Corollary 3.3:

$$\mathbb{E}\left[C_{BC}^{\text{sum}}(\boldsymbol{H},P)\right]-\mathbb{E}\left[C(\boldsymbol{h}_{(1)},\boldsymbol{h}_{(2)},P)\right] \to 1$$

as $K \to \infty$.

The above corollary corresponds to a special case of scheme I with $L = 2$, and shows that simply selecting the two strongest users leaves a performance gap of 1 nat/symbol.

Corollary 3.4: For any fixed value l and sequence $L(K)$ with $\lim_{K\to\infty} L(K) = \infty$,

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

as $K \to \infty$.

The above corollary shows that scheme II is asymptotically optimal when the list size L grows suitably large, and thus implies that the dominating scheme III is asymptotically optimal as well. Since $\lim_{K\to\infty}\mathbb{E}\left[\log\left(1+\frac{P}{2}\|\boldsymbol{h}_{(1)}\|^2\right)\right]=\infty$, it follows as a further simple consequence that for any fixed value of $L \geq 2$,

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

Thus, scheme I is asymptotically optimal in a ratio sense, and hence so are schemes II and III.

In conclusion, the above results show that scheme II is asymptotically optimal in the sense that the absolute gap to 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 them 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 when the users are the two strongest ones.

IV. Heterogeneous users

In the previous section, we studied the problem of maximizing the sum rate in a system with statistically identical users. We now turn our attention to a more general system with M transmit antennas and heterogeneous user characteristics (i.e., channels are not necessarily i.i.d). As mentioned earlier, the sum capacity is no longer an appropriate performance metric now, because it does not reflect the potential fairness issues that arise when users with different channel statistics obtain vastly different throughput portions. Hence, we focus on the problem of maximizing a *weighted* sum rate expression, and demonstrate that transmitting to a properly selected group of users asymptotically achieves the maximum expected weighted sum rate, although scheduling just M users will no longer be sufficient in the heterogeneous case.

A. Bounds for the weighted sum rate

We first establish a generic upper bound for the weighted sum rate for an arbitrary number of M transmit antennas. Let w_k be the weight associated with the k-th user. For notational convenience, define $\Delta w_k := w_k - w_{k+1}$ with the convention that $w_{K+1} = 0$. Without loss of generality, we assume that the users are indexed such that $w_1 \geq w_2 \geq \cdots \geq w_K$.

Theorem 4.1: For any given set of channel vectors,

$$C_{BC}^{\mathbf{w}}(\boldsymbol{H},P) \leq \max_{\sum_{k=1}^K P_k \leq P} \Delta w_1 \log(1+P_1\|\boldsymbol{h}_1\|^2)+ M\sum_{k=2}^{K}\Delta w_k \log\left(1+\sum_{l=1}^{k}\frac{P_l}{M}\|\boldsymbol{h}_l\|^2\right).$$

Proof

Equation (2) yields that $C_{BC}^{\mathbf{w}}(\boldsymbol{H},P)=\sum_{k=1}^K \Delta w_k S_k$, with

$$S_k = \log\det\left(I_M + \sum_{l=1}^{k} P_l \boldsymbol{h}_l^{\dagger}\boldsymbol{h}_l\right).$$

Clearly,

$$S_1 = \log(1+P_1\|\boldsymbol{h}_1\|^2). \tag{8}$$

Using Hadamard's inequality for Hermitian positive semi-definite matrices [3, p. 502], and the concavity of the log function, we obtain

$$S_k \leq \sum_{m=1}^{M}\log\left(1+\sum_{l=1}^{k}P_l h_{lm}^{\dagger}h_{lm}\right) \leq M\log\left(1+\sum_{l=1}^{k}\frac{P_l}{M}\|\boldsymbol{h}_l\|^2\right) \tag{9}$$

for all $k = 2, \dots, K$.

Substituting inequalities (8) and (9), the statement of the theorem follows.

□

The next upper bound follows as a straightforward corollary of Theorem 4.1.

Corollary 4.1: For any given set of channel vectors,

$$C_{BC}^{\mathbf{w}}(\boldsymbol{H},P) \leq M \max_{\sum_{k=1}^K P_k \leq P/M}\sum_{k=1}^{K}\Delta w_k \log(1+\sum_{l=1}^{k}P_l\|\boldsymbol{h}_l\|^2). \tag{10}$$

In order to develop a suitable asymptotic framework, we assume that there are C classes of users, with K_c the number of class-c users and $\sum_{c=1}^{C} K_c = K$. Let $\boldsymbol{h}_k^{(c)}$ be the channel vector of the k-th class-c user. With minor abuse of notation, we let w_c be the weight associated with class c, and define $\Delta w_c := w_c - w_{c+1}$, with the convention that $w_{C+1} = 0$ as before. Let T_c be the total rate received by class c. Thus the weighted sum rate is $T := \sum_{c=1}^{C} w_c T_c$. Without loss of generality, we assume that the classes are indexed such that $w_1 \geq w_2 \geq \cdots \geq w_C$. Let $\boldsymbol{h}_{(k)}^{(c)}$ be the channel vector of the class-c user with the k-th largest norm, i.e., $\|\boldsymbol{h}_{(1)}^{(c)}\|^2 \geq \|\boldsymbol{h}_{(2)}^{(c)}\|^2 \geq \cdots \geq \|\boldsymbol{h}_{(K_c)}^{(c)}\|^2$.

The next corollary specializes the upper bound in (10) to a class-based system.

Corollary 4.2: For any given set of channel vectors,

$$\sum_{c=1}^{C} w_c T_c \leq U(w_c; \|\boldsymbol{h}_{(1)}^{(c)}\|^2; P) \tag{11}$$

$$:= \max_{\sum_{c=1}^{C} P_c \leq P/M} \sum_{c=1}^{C} \Delta w_c \log\left(1 + \sum_{d=1}^{c} P_d \|\boldsymbol{h}_{(1)}^{(d)}\|^2\right).$$

Note that when all weights are taken equal to one, the upper bound in (10) reduces to that in equation (4) for the sum rate. Recall that the upper bound in (4) is tight in the sense that it can actually be achieved when there are M users with orthogonal channel vectors tied for the maximum norm. Likewise, the upper bound in (11) is also attainable for certain configurations of channel vectors, as described below.

Assume that there exists a set of MC users, M from each class, with channel vectors $\boldsymbol{h}_m^{(c)}$, $c = 1, \ldots, C$, $m = 1, \ldots, M$, that satisfy the following two properties:

(i) within each class, all M users are tied for the maximum norm, i.e., $\|\boldsymbol{h}_m^{(c)}\|^2 = \|\boldsymbol{h}_{(1)}^{(c)}\|^2$ for all $c = 1, \ldots, C$, $m = 1, \ldots, M$; (ii) for each $c = 1, \ldots, C$, the vectors $\boldsymbol{h}_m^{(c)}, m = 1, \ldots, M$ are mutually orthogonal. That is, the M users within any class are orthogonal to each other.

Note that the relative orientation of channel vectors across different classes does not matter in the above description. That is, the upper bound can be achieved even if the users in different classes are arbitrarily oriented, as long as they are mutually orthogonal within their respective classes, and have the same norms. Figure 2 shows an example of an upper bound achieving channel configuration, for the case of $C = 2$ user classes and $M = 2$ transmit antennas.

In order to see that the above configuration indeed achieves the upper bound, let $P_1^*(K), \ldots, P_C^*(K)$ be the optimizing power levels of the upper bound in (11) for given values of $\|\boldsymbol{h}_{(1)}^{(c)}\|^2$, $c = 1, \ldots, C$, i.e.,

$$P^*(K) = (P_1^*(K), \ldots, P_C^*(K)) :=$$

$$\arg \max_{\sum_{c=1}^{C} P_c \leq P/M} \sum_{c=1}^{C} \Delta w_c \log\left(1 + \sum_{d=1}^{c} P_d \|\boldsymbol{h}_{(1)}^{(d)}\|^2\right).$$

Now suppose that we assign power $P_c^*(K)$ to each of the M class-c users in the optimal configuration, and arrange the users in order of increasing class index in the DPC sequence.

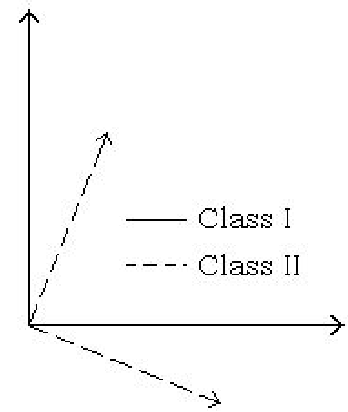

Fig. 2. An optimal channel configuration for two user classes.

Then, the partial sum rate $S_c := \sum_{d=1}^{c} T_d$ of the first c classes will be

$$S_c = \log\det\left(I_M + \sum_{d=1}^{c} P_d^*(K) \left[\sum_{m=1}^{M} \boldsymbol{h}_m^{(d)\dagger} \boldsymbol{h}_m^{(d)}\right]\right). \tag{12}$$

Consider the matrix

$$A_d := \begin{bmatrix} \boldsymbol{h}_1^{(d)} \\ \vdots \\ \boldsymbol{h}_M^{(d)} \end{bmatrix}.$$

Since the rows of the above matrix are orthogonal and have the same norms, it is clear that A_d is a scaled version of a unitary matrix. Thus, $A_d^\dagger A_d = \|\boldsymbol{h}_{(1)}^{(d)}\|^2 I_M$. Next, note that the expression inside the square brackets in (12) can be written as

$$\sum_{m=1}^{M} \boldsymbol{h}_m^{(d)\dagger} \boldsymbol{h}_m^{(d)} = \left[\boldsymbol{h}_1^{(d)\dagger} \ldots \boldsymbol{h}_M^{(d)\dagger}\right] \begin{bmatrix} \boldsymbol{h}_1^{(d)} \\ \vdots \\ \boldsymbol{h}_M^{(d)} \end{bmatrix}$$

$$= A_d^\dagger A_d = \|\boldsymbol{h}_{(1)}^{(d)}\|^2 I_M.$$

Thus, the partial sum rate in (12) becomes

$$\begin{aligned} S_c &= \log\det\left(I_M \left[1 + \sum_{d=1}^{c} P_d^*(K) \|\boldsymbol{h}_{(1)}^{(d)}\|^2\right]\right) \\ &= \log\left(1 + \sum_{d=1}^{c} P_d^*(K) \|\boldsymbol{h}_{(1)}^{(d)}\|^2\right)^M \\ &= M \log\left(1 + \sum_{d=1}^{c} P_d^*(K) \|\boldsymbol{h}_{(1)}^{(d)}\|^2\right). \end{aligned}$$

Finally, since the total weighted sum rate may be written as $\sum_{c=1}^{C} w_c T_c = \sum_{c=1}^{C} \Delta w_c S_c$, it follows that the optimal configuration indeed achieves the upper bound in (11).

B. Random channel vectors

The lower and upper bounds for the weighted sum rate in the previous subsection hold for any arbitrary but fixed set of channel vectors.

In order to derive asymptotic results, we will as before assume the channel vectors to be random and focus on the *expected* weighted sum rate. Within each class we assume the channel vectors to be independent and identically distributed, i.e., $\boldsymbol{h}_1^{(c)}, \boldsymbol{h}_2^{(c)}, \ldots$ are i.i.d. copies of some random vector $\boldsymbol{h}^{(c)} \in \mathbb{C}^2$. Among the various classes, the channel vectors may however have different statistical characteristics. The numbers of users of the various classes are assumed to grow large in fixed proportions, i.e., $K_c = \alpha_c K$ for fixed coefficients $\alpha_1, \ldots, \alpha_C$ with $\sum_{c=1}^{C} \alpha_c = 1$.

We assume that the channel vectors of all users in class c are Rayleigh faded with parameter $\beta_c, c = 1, \ldots, C$. In other words, for each class c, $\boldsymbol{h}^{(c)} = \beta_c \boldsymbol{h}$, where the components of $\boldsymbol{h}$ are independent and distributed according to $\mathcal{CN}(0,1)$ as in the homogeneous case.

C. Large-K asymptotics

We now proceed to show that the upper bound in (11) is asymptotically achievable by transmitting to a judiciously chosen subset of MC users. In the case of homogeneous users, the key observation was that when the total 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 intuitive insight was then formalized by establishing that selecting such a pair of users and allocating equal power to each of them asymptotically achieves the maximum expected sum rate. Likewise, there exists with high probability a group of MC users with channel vectors close to the optimal configuration in the heterogeneous case when the total number of users is large. Thus, we will show that selecting such a group of MC users and allocating power P_c^* to the M class-c users, where

$$P^* = (P_1^*, \ldots, P_C^*) := \arg \max_{\sum_{c=1}^{C} P_c \le P/M} \sum_{c=1}^{C} \Delta w_c \log \left(\sum_{d=1}^{c} P_d \beta_d^2 \right)$$

asymptotically achieves the upper bound in (11). Define

$$\begin{aligned} V(w_c; \beta_c^2; P) \quad &:= \max_{\sum_{c=1}^{C} P_c \le P/M} \sum_{c=1}^{C} \Delta w_c \log \left(\sum_{d=1}^{c} P_d \beta_d^2 \right) \\ &= \sum_{c=1}^{C} \Delta w_c \log \left(\sum_{d=1}^{c} P_d^* \beta_d^2 \right). \end{aligned} \tag{13}$$

Note that the power levels $(P_1^*, \ldots, P_C^*)$ are the limiting values of $(P_1^*(K), \ldots, P_C^*(K))$ when the norms $\|\boldsymbol{h}_{(1)}^{(c)}\|^2$ grow large in proportion to β_c^2. It may in fact be shown that $(P_1^*(K), \ldots, P_C^*(K))$ converge to $(P_1^*, \ldots, P_C^*)$ in probability, as $K \to \infty$. Finally, note that the asymptotic powers $(P_1^*, \ldots, P_C^*)$ depend only on the values of the weights w_c and the relative channel qualities β_c of the various classes. This implies that the asymptotically optimal power assignments can be *pre-computed*, without having to solve the optimization (11) for each set of channel realizations.

We will now prove that transmitting to a carefully selected subset of MC users asymptotically achieves the upper bound (11) and thus maximizes the expected weighted sum rate. Motivated by the knowledge of the optimal channel configuration, we propose the following two user selection schemes which will be referred to as the 'list' scheme and the 'cone' scheme, respectively.

Even though the optimal configuration admits arbitrary angles between users of different classes, the following user selection schemes pick users that lie along the *same* orthogonal directions for each class, thus making each of the C orthogonal sets nearly collinear. This particular choice, though not necessary, simplifies the exposition of our user selection schemes and the ensuing proof techniques, while preserving asymptotic optimality.

1) List scheme: The 'list' scheme first identifies for each class the users with norms close to the maximum, and then selects a nearly orthogonal set of users among these. Specifically, the list scheme first selects the ML_c strongest users from class c, and divides them into M 'groups' of size L_c each, say in a round-robin fashion. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_M \in \mathbb{C}^M$ be an arbitrary set of orthonormal vectors. From the M groups of L_c users formed above, the scheme picks, from group i, the user whose channel is most collinear to $\boldsymbol{u}_i$. That is, it selects the user in group i who maximizes the normalized inner product with $\boldsymbol{u}_i, i = 1, 2, \ldots, M$. This procedure is repeated for every class $c = 1, 2, \ldots, C$. This leads to a set of MC users, M from each class, which have a geometry close to the optimal configuration described earlier.

2) Cone scheme: The 'cone' scheme first identifies a group of users that are close to orthogonal, and then selects the ones with the largest norms among these.

Definition: Fix a $0 < \lambda < 1$. Then, two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *λ-aligned* if

$$U(\boldsymbol{u}, \boldsymbol{v}) := \frac{| < \boldsymbol{u}, \boldsymbol{v} > |^2}{\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2} \ge 1 - \lambda.$$

If this is true, we also say that $\boldsymbol{u}$ *lies in the λ-cone* of $\boldsymbol{v}$ and *vice-versa*.

Now, let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_M \in \mathbb{C}^M$ be an arbitrary set of orthonormal vectors, and let $0 < \lambda < 1$ be a small tolerance margin. The cone scheme first considers the set of all channel vectors that are λ-aligned with $\boldsymbol{u}_i$, for $i = 1, 2, \ldots, M$. From each of these M nearly orthogonal 'cones' of channel vectors, the scheme picks the strongest user from each class.

After selecting the users in the above-described fashion, both the list and the cone schemes allocate power P_c^* (defined in (13)) to all M class-c users.

Define $\widehat{T}_c$ as the rate received by class c under the list scheme, i.e., the sum rate of the M class-c users selected, and denote by $\widehat{T} := \sum_{c=1}^{C} w_c \widehat{T}_c$ the total weighted sum rate. The next theorem shows that the list scheme asymptotically maximizes the expected weighted sum rate, as long as the list size L_c grows with the system size at a particular rate.

Theorem 4.2: In the list scheme described above, assume that the list size $L_c(K)$ for each class is such that $\lim_{K \to \infty} L_c(K) = \infty$ and $L_c(K) \le o(K^\delta)$ as $K \to \infty$ for *every* $\delta > 0$. Then, the List scheme is asymptotically optimal in the sense that it closes the gap to the weighted sum capacity. Specifically, the gap between the upper bound in (11) and the weighted sum rate achieved using the List scheme converges to 0 as K becomes large:

$$\lim_{K \to \infty} \mathbb{E}\left[U(w_c; \|\boldsymbol{h}_{(1)}^{(c)}\|^2; P) \right] - \mathbb{E}\left[\widehat{T} \right] = 0.$$

The above theorem shows that scheduling a suitably selected group of *MC* users asymptotically achieves the upper bound (11) and thus maximizes the expected weighted sum rate. In fact, it shows that scheduling M users from each of the classes $c \in C^*$ is sufficient to asymptotically achieve the maximum expected weighted sum rate, where $C^* := \{c : P_c^* > 0\}$. Further, in [7], we derived bounds for the gap to capacity corresponding to the List scheme, for any constant finite value of the List size in case $M = 2$.

Remark 4.1: In case the list size scales as $L_c(K) = O(K^\delta)$, there is a strictly positive rate gap which however decreases to 0 as δ tends to 0. Inspection of the proof [6] further reveals that $\mathbb{E}\left[U(w_c; \|\boldsymbol{h}_{(1)}^{(c)}\|^2; P)\right]$ asymptotically behaves as $M w_1 \log\log K + MV(w_c; \beta_c^2; P)$ in the sense that the difference decays to zero asymptotically, and hence so does $\mathbb{E}\left[\widehat{T}\right]$. In fact, it may be deduced that the total rate received by class 1 grows as $M[\log\log K + \log(P_1^*) + 2\log(\beta_1)]$, while the total rate received by class c, $c = 2, \ldots, C$, asymptotically converges to $M\left[\log\left(\sum_{d=1}^{c} P_d^* \beta_d^2\right) - \log\left(\sum_{d=1}^{c-1} P_d^* \beta_d^2\right)\right]$. Thus, asymptotically, the lion's share of the aggregate throughput is accounted for by the class with the largest weight.

In a similar fashion, it can be shown that the cone scheme described above also asymptotically achieves the maximum weighted sum rate, as long as the tolerance margin λ is scaled down at an appropriate rate. Let $\widetilde{T}_c$ denote the throughput obtained by class c under the cone scheme, and define $\widetilde{T} := \sum_{c=1}^{C} w_c \widetilde{T}_c$ to be the total weighted sum rate.

Theorem 4.3: In the cone scheme described above, assume that the tolerance margin λ scales with the system size as $\lambda = \frac{\kappa}{(\log K)^2}$ for some $\kappa > 0$. Then, the cone scheme is asymptotically optimal in the sense that it closes the gap to the weighted sum capacity:

$$\lim_{K\to\infty} \mathbb{E}\left[U(w_c; \|\boldsymbol{h}_{(1)}^{(c)}\|^2; P)\right] - \mathbb{E}\left[\widetilde{T}\right] = 0.$$

V. Numerical results

A. Homogeneous case

Here, 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 [15] and [25].

We present numerical results for a system with two transmit antennas and $K = 25$ users in Figure 3. In Figure 3(a), we plot the ratio of the sum rate obtained by the various schemes to the TDMA sum rate, versus the SNR (in dB). The results shown here were an average over 100 channel realizations. The solid line corresponds to the optimal DPC scheme. The dotted line just underneath the solid line corresponds to scheme II with $L = 5$. 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 (5), 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 this system is graphed as a function of SNR in Figure 3(b).

The BF scheme proposed in [15] 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 ϕ_1 and ϕ_2, and selects two users $k_m^* := \arg\max_{k=1,\ldots,K} SINR_{k,m}$, $m = 1, 2$, 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 Figure 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 actually 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 Figure 3(b). In contrast, the TDMA sum rate R_{TDMA} grows without bound, albeit slowly.

B. Heterogeneous case

1) Background for the numerical results: The simulation results which are provided below are for a two-class system. The weights are taken to be $w_1 = 2$, $w_2 = 1$ (although we usually equivalently normalized these to sum to 1 over the users), and the coefficients $\beta_1 = 0.5$, $\beta_2 = 1.0$ determine the mean SNR's. The two populations of users are of equal size, $K_1 = K_2 = 10$. Under these circumstances, the asymptotically optimal power values are $P_1^* = 1/3$, $P_2^* = 1/6$, scaling out P, which is varied in most of the results below. We will state its value when necessary.

We now describe the schemes themselves. As far as the list and cone schemes are concerned, these are detailed in the text. Throughout, the asymptotically optimal power settings will be used, no power optimization is being employed. We will also consider TDMA, by which we mean the scheme that picks the user which has the maximum weighted rate when

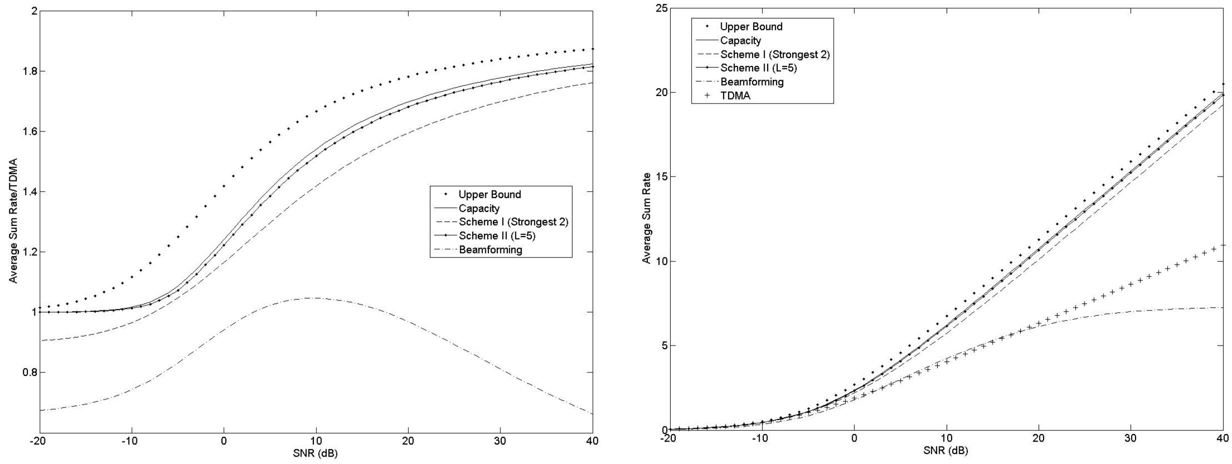

Fig. 3. (a) Comparison of various user selection heuristics with TDMA, $K = 25$ users. (b) Absolute sum rate in nats vs SNR, $K = 25$ users.

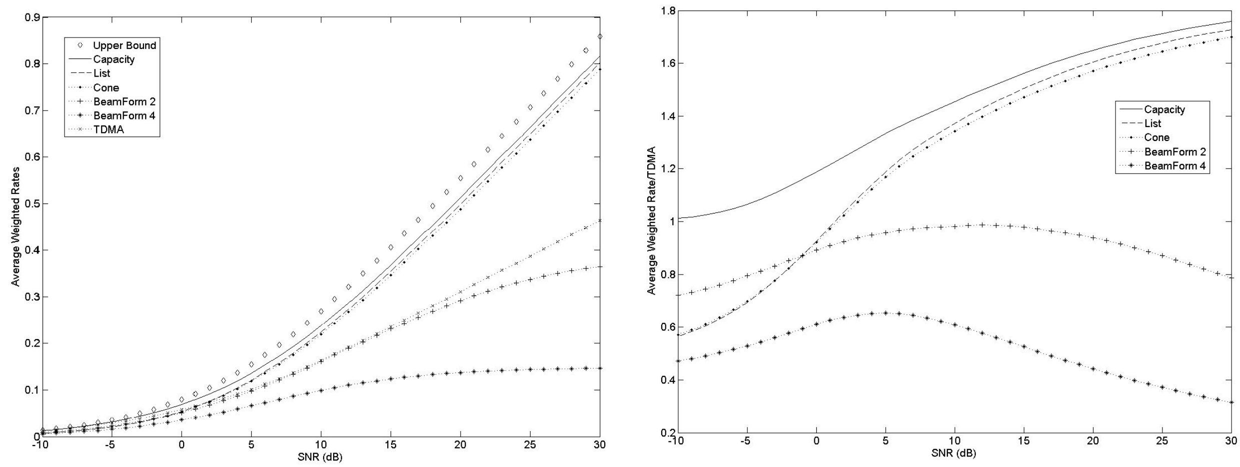

Fig. 4. (a) Absolute weighted rates for various schemes and upper bound (b) Ratio to TDMA for various schemes

assigned full power, over all the users. Thus, it selects the k-th class-c user which maximizes

$$\max_{c=1,\ldots,C} \max_{k=1,\ldots,K_c} w_c \log(1 + P\|\boldsymbol{h}_k^{(c)}\|^2)$$
$$= \max_{c=1,\ldots,C} w_c \log(1 + P\|\boldsymbol{h}_{(1)}^{(c)}\|^2).$$

Finally, we consider two BF versions. The first version (referred to as BeamForm 2 in the figure) schedules one user in each beam, with the powers equally split and the user with the maximum weighted rate as determined by the SNR being the one selected for each beam. The second version (referred to as BeamForm 4) schedules one user from each class in each of the beams. In this case each user is assigned half its classes asymptotic power. The latter scheme is not expected to perform well as the interference between users on the same beam cannot be resolved except by using DPC or some equivalent approach. As in the homogeneous case, rates are given in nats throughout.

2) Graphs for basic schemes: Figure 4(a) shows results for all the main schemes as well as the upper bound and the average maximum weighted capacity limit. $L = 5$ was set for the list scheme and $\delta = 0.2$ for the cone scheme. (Further numerical experiments indicated that the performance of the list scheme is quite robust with respect to the list size L, so that the exact value is not that critical.) As expected, the upper bound (11) is loose and the list and cone schemes perform well at high SNR values. For low SNR values, TDMA outperforms these schemes. The BF schemes fall off at very high SNR as the figure shows.

As far as the list and cone schemes are concerned, good performance at high SNR is expected. However, at low SNR TDMA is close to optimal. (This latter conclusion follows from the linearity of the log.) Thus for low to moderate SNR's one could make up for the loss of rate in the list scheme by optimizing the powers. Similarly, the cone scheme does well at high SNR but not at low SNR. This loss in performance can also be addressed by assigning the powers optimally. This is a concave optimization in three independent variables, and is therefore potentially a time-consuming calculation, since we have no explicit formula for determining the optimal powers.

Figure 4(b) shows the same results, but gives the ratio to TDMA. Note that unlike the homogeneous case, BF is not asymptotically optimal in terms of differences as the number of users is increased at fixed SNR. However, at low SNR's (below 0 dB) BeamForm 2 does better than cone or list. Figure 4(b) shows that BeamForm 2 performs consistently worse than TDMA, which was also observed in the homogeneous example which had a similar number of users. The results

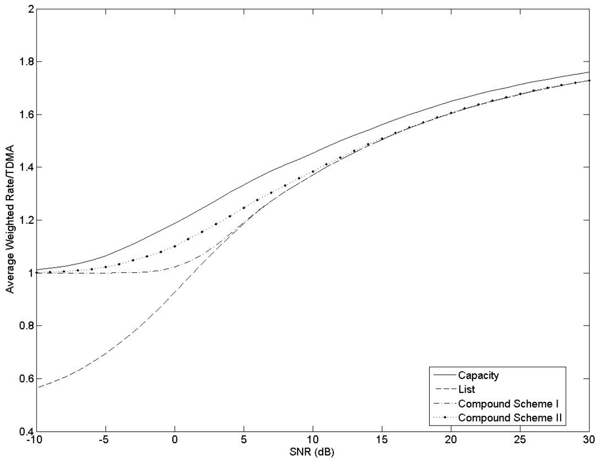

Fig. 5. Relative weighted sum rates for compound schemes compared to TDMA.

for BeamForm 4 are worse than those for BeamForm 2 as expected.

3) Additional compound schemes: We now look at simpler enhancements to avoid power optimization. One such enhancement to the list scheme is to identify the best possible pair among the already selected four users. Consider the two-user weighted sum rates obtained by scheduling all possible pairs of these users. The power is split equally while scheduling two users of the same class, but when scheduling one user from each class, we allocate them powers $2P_1^*$ and $2P_2^*$ respectively. The two-user scheme picks the pair that corresponds to the highest weighted sum rate among the six possible pairs.

We thus arrive at the following heuristic schemes. Compound scheme I selects the better among TDMA and the list schemes. Compound scheme II goes further and selects the best among TDMA, the two-user scheme above, and the original list scheme.

A three-user heuristic scheme was also considered, but since it did not provide any appreciable improvement, it has been omitted from the results.

In Figure 5 we compare the list scheme with the two heuristic schemes, Compound I and Compound II. These results are more clearly seen as a ratio to TDMA rather than the absolute rates which are difficult to distinguish. Since Compound I takes the best of TDMA and the list scheme, it cannot do worse than TDMA at any point and list at any point. Hence, it does well at low SNRs and at high SNRs. There is nevertheless a significant rate gap for this scheme for moderate SNR's, roughly in the range 0–5dB. Here TDMA falls off, but the list scheme is not yet in its most advantageous range. However, Compound II closes most of this gap as can be seen. The results in Figures 4 and 5 were averaged over 50 channel realizations.

Acknowledgments

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.

References

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

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

[3] T. Cover, J. Thomas (1991). *Elements of Information Theory*. Wiley & Sons.

[4] 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.

[5] 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.

[6] K.P. Jagannathan (2006). Efficient Scheduling of Multi-Antenna Broadcast Systems. S.M thesis, LIDS, Mass. Inst. of Tech., Cambridge, MA.

[7] K.P. Jagannathan, S.C. Borst, P.A. Whiting, E. Modiano (2006). Scheduling of multi-antenna broadcast systems with heterogeneous users. In: *Proc. 44th Allerton Conference on Communication, Control, and Computing, Monticello, IL.*

[8] N. Jindal (2005). A high SNR analysis of MIMO broadcast channels. In: *IEEE International Symposium on Information Theory, Adelaide, Australia.*

[9] N. Jindal (2006). Finite rate feedback MIMO broadcast channels. In: *Proc. Workshop on Information Theory and its Applications, UC San Diego.*

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

[11] 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. Inform. Theory* **51** (4), 1570–1580.

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

[13] J. Lee, N. Jindal (2006). Symmetric capacity of MIMO downlink channels. In: *IEEE International Symposium on Information Theory, Seattle, WA.*

[14] M.J. Neely, E. Modiano, C.E. Rohrs (2003). Dynamic power allocation and routing for time-varying wireless networks. In: *Proc. Infocom 2003*.

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

[16] M. Sharif, B. Hassibi (2005). Scaling laws of sum rate using time sharing, DPC and beamforming for MIMO broadcast channels. In: *IEEE International Symposium on Information Theory, Adelaide, Australia.*

[17] M. Sharif, B. Hassibi (2006). A Comparison of time-sharing, beamforming, and DPC for MIMO broadcast channels with many users. To appear in *IEEE Trans. Comm.*

[18] A.L. Stolyar (2004). MaxWeight scheduling in a generalized switch: state space collapse and workload minimization in heavy traffic. *Ann. Appl. Prob.* **14**, 1–53.

[19] 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, Computing, Monticello, IL*

[20] C. Swannack, E. Uysal-Biyikoglu, G. Wornell (2005). Finding NEMO: near-orthogonal sets for multiplexing and opportunistic scheduling in MIMO broadcast. In: *International Conference on Wireless Networks, Communications and Mobile Computing, Maui, HI.*

[21] L. Tassiulas, A. Ephremides (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. Automrat. Contr.* **37**, 1936–1948.

[22] A. Vakili, A. Dana, M. Sharif, B. Hassibi (2005). Differentiated rate scheduling for MIMO broadcast channels. In: *Proc. 43rd Annual Allerton Conf. Commun. Control, Computing, Monticello, IL.*

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

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

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

[26] H. Viswanathan, K. Kumaran (2005). Rate scheduling in multiple-antenna downlink wireless systems. *IEEE Trans. Commun.* **53** (4), 645–655.

[27] H. Viswanathan, S. Venketesan, H. Huang (2005). Downlink capacity evaluation of cellular networks with known interference cancellation. *IEEE J. Select. Areas Commun.* **21** (5), 802–811.

[28] H. Weingarten, Y. Steinberg, S. Shamai (2006). The capacity region of the Gaussian MIMO broadcast channel. *IEEE Trans. Inform. Theory* **52** (9), 3936–3964.

[29] T. Yoo, A. Goldsmith (2006). On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming. *IEEE J. Select. Areas Commun.* **24** (3), 528–541.

Krishna Jagannathan received his B.Tech degree in Electrical Engineering from the Indian Institute of Technology (IIT) Madras, India , with silver medal honors in 2004, and his S.M degree from the Massachusetts Institute of Technology (MIT) in 2006. He is currently pursuing a doctorate in the department of Electrical Engineering and Computer Science at MIT. He worked as a summer consultant at the Mathematical Sciences Research Center, Bell Laboratories, Murray Hill, NJ in 2005 and a summer research fellow at the Indian Institute of Science (IISc), Bangalore, India in 2003. His research interests lie in wireless communications, networking and information theory.

Sem Borst received the MSc degree in applied mathematics from the University of Twente, The Netherlands, in 1990, and the PhD degree from the University of Tilburg, The Netherlands, in 1994. During the fall of 1994, he was a visiting scholar at the Statistical Laboratory of the University of Cambridge, England. In 1995, he joined the Mathematics of Networks and Systems department of Bell Laboratories, Lucent Technologies in Murray Hill, USA. He also holds a part-time appointment as a professor of Stochastic Operations Research at Eindhoven University of Technology. Sem Borst is a member of IFIP Working Group 7.3, and serves or has served as a member of several program committees and editorial boards. His main research interests are in the area of performance evaluation and resource allocation algorithms for communication networks.

Phil Whiting received his BA degree from the University of Oxford, his MSc from the University of London and his Ph. D. was in queueing theory from the University of Strathclyde. After a post-doc at the University of Cambridge, Phil's interests centered on wireless. In 1993 Phil participated in the Telstra trial of Qualcomm CDMA in South Eastern Australia. He then joined the Mobile research Centre at the University of South Australia Adelaide. Since 1997 he has been with Bell Labs.

His main interests are the mathematics of wireless networks, particularly stochastic models for resource allocation and information theory. Phil's current research include large deviations asymptotics for balls and urns models (with no wireless applications found yet!) and the application of stochastic control to scheduling in wireless data networks.

Eytan Modiano received his B.S. degree in Electrical Engineering and Computer Science from the University of Connecticut at Storrs in 1986 and his M.S. and PhD degrees, both in Electrical Engineering, from the University of Maryland, College Park, MD, in 1989 and 1992 respectively. He was a Naval Research Laboratory Fellow between 1987 and 1992 and a National Research Council Post Doctoral Fellow during 1992-1993. Between 1993 and 1999 he was with MIT Lincoln Laboratory where he was the project leader for MIT Lincoln Laboratory's Next Generation Internet (NGI) project. Since 1999 he has been on the faculty at MIT; where he is presently an Associate Professor. His research is on communication networks and protocols with emphasis on satellite, wireless, and optical networks.

He is currently an Associate Editor for Communication Networks for IEEE Transactions on Information Theory and for The International Journal of Satellite Communications. He had served as a guest editor for IEEE JSAC special issue on WDM network architectures; the Computer Networks Journal special issue on Broadband Internet Access; the Journal of Communications and Networks special issue on Wireless Ad-Hoc Networks; and for IEEE Journal of Lightwave Technology special issue on Optical Networks. He served as the Technical Program co-chair for Wiopt 2006, IEEE Infocom 2007, and ACM MobiHoc 2007.