

Differentiated Rate Scheduling for the Down-link of Cellular Systems

Amir F. Dana, Masoud Sharif, Ali Vakili, and Babak Hassibi

Abstract—We consider the problem of differentiated rate scheduling for the downlink (i.e., multi-antenna broadcast channel), in the sense that the rates required by different users must satisfy certain constraints on their ratios. When full channel state information (CSI) is available at the transmitter and receivers, the problem can be readily solved using dirty paper coding (DPC) and the application of convex optimization techniques on the dual problem which is the multiple access channel (MAC). Since in many practical application full CSI may not be feasible and computational complexity prohibitive when the number of users is large, we focus on other simple schemes that require very little CSI: time-division opportunistic (TO) beamforming where in different time slots (of different lengths) the transmitter performs opportunistic beamforming to the users requiring the same rate, and weighted opportunistic (WO) beamforming where the random beams are assigned to those users having the largest weighted SINR. For single antenna systems we also look at the capacity-achieving superposition coding (SC) scheme. In all cases, we determine explicit schedules to guarantee the rate constraints and show that, in the limit of large number of users, the throughput loss compared to the unconstrained throughput (sum-rate capacity) tends to zero. We further provide bounds on the rate of convergence of the sum-rates of these schemes to the sum-rate capacity. Finally, we provide simulation results of the performance of different scheduling schemes considered in the paper.

Index Terms—MIMO, broadcast channel, opportunistic beamforming, QoS, dirty paper coding.

I. INTRODUCTION

THE downlink scheduling in cellular systems is known to be one major bottleneck for future broadband communication systems. From an information-theoretic perspective, broadcast channels [1], and in particular the Gaussian broadcast channel, can be used to model the downlink in a cellular system. There exist an abundance of information-theoretic results describing the limits of the achievable rates to the users in single-input single-output (SISO) Gaussian broadcast channels (see e.g., [4], [5]). For example in a homogeneous network, i.e., a network where the fading and noise distributions of all the users are identical, if the transmitter wants to maximize the throughput (or the sum of the rates to all the receivers), it is well known that the optimal strategy is to transmit to the

user with the best channel condition at each channel use. This is often referred to as the *opportunistic* transmission strategy [2].

More recently, there has been growing interest in the use of multiple antennas (at the transmitter, receivers, or both) for wireless communication systems. This has led to an interest in the multiple-input multiple-output (MIMO) Gaussian broadcast channel, where the transmitter and the various users may be equipped with multiple transmit and receive antennas, respectively. First, the sum-rate of the MIMO broadcast channel, i.e., the maximum possible sum of the rates to all users [6], and then the entire capacity region [7] were shown to be achieved by an interference cancellation scheme referred to as *dirty paper coding* [8].

Thus, from a theoretical point of view, the limits of reliable communication in MIMO Gaussian broadcast channels is well understood. Fortunately, the same is true if one takes a *computational* point of view. Using the *duality* of the capacity region of Gaussian broadcast systems and multi-access systems (MAC) introduced first in [9], one can use the polymatroid structure of the problem, to compute any point on the capacity region via standard convex optimization techniques [3].

A crucial assumption in all the aforementioned results is that the channel coefficients to all the users be known—an assumption referred to as full channel state information (CSI)—at the transmitter. In fact, it is easy to show that with no CSI at the transmitter there is no capacity gain to be had by employing multiple antennas at the transmitter (provided all the users have single antennas) [10]. However, in practice, obtaining full CSI at the transmitter may not be feasible, especially for systems where the number of users is large and/or the users are mobile so that the channel coefficients vary rapidly with time. Furthermore, when the number of users is large, the computational complexity of DPC, and even the convex optimization steps required to determine the optimal covariance matrices from the dual MAC, may become prohibitively large. Therefore there is interest in developing *simple* schemes that require *little* CSI at the transmitter, yet deliver on most of the capacity offered by the MIMO broadcast channel.

In homogenous networks, the sum-rate point is a symmetrical point on the boundary of the capacity region and so treats all the users equally. In systems which are provisioned to provide differentiated services to different users, the transmitter has to give different services (or rates) to different subsets of receivers, and yet at the same time, maximize the throughput (see e.g., [11] for a discussion of the SISO case). Giving differentiated rates to users clearly means operating at non-symmetrical boundary points of the capacity region. As

Paper approved by S. A. Jafar, the Editor for Wireless Communication Theory and CDMA of the IEEE Communications Society. Manuscript received September 12, 2006; revised May 15, 2007. This work was supported in part by the National Science Foundation under grant no. CCR-0133818 and CCR-0326554, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking. An early version of these results appeared in IEEE International Symposium on Information Theory, Adelaide, Australia, 2005.

A. F. Dana is with Qualcomm, San Diego, CA, 92121 (e-mail: dana@systems.caltech.edu).

M. Sharif, A. Vakili, and B. Hassibi are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 (e-mail: {masoud, avakili, hassibi}@systems.caltech.edu).

Digital Object Identifier 10.1109/TCOMM.2008.060516

mentioned earlier, this problem can, in principle, be solved since the duality to the MAC allows one to attain any point on the capacity region.

However, since this solution requires full CSI at the transmitter and potentially prohibitive computations when the number of users is large, the main goal of this paper is to develop *simple* schemes, that require very *little* CSI, guarantee *differentiated rates* to the users, and that operate *close* to the boundary of the capacity region. We will also be interested in quantifying the rate loss, compared to the sum rate, for various differentiated rate schemes.

The remainder of the paper is organized as follows. Section II describes the model and states the main problem we are considering. Section III reviews the known result for Gaussian broadcast channels and the scaling laws of the optimal sum-rate (achieved by DPC) as well as the sum-rate achieved by opportunistic beamforming. Sections IV, V, and VI give the main results of the paper. Simulation results are presented in Section VII and Section VIII concludes the paper.

II. PROBLEM FORMULATION

We consider a fading Gaussian broadcast channel (GBC) with M antennas at the transmitter and n users, each with $N = 1$ receive antennas.¹ The channels to each user are assumed to be block fading with a coherence interval of T ; in other words, the channels remain constant for T channel uses after which they change to different independent values.² Furthermore, over different users the fading is assumed to be independent. Thus, during any coherence interval, the signal to the i -th user, $i = 1, 2, \dots, n$, can be written as

$$x_i(t) = \sqrt{\rho} H_i s(t) + w_i(t), \quad t = 1, \dots, T \quad (1)$$

where $H_i \in \mathcal{C}^{1 \times M}$ is constant during the coherence interval and has i.i.d $\mathcal{CN}(0, 1)$ entries. Here, $w_i(t)$ is additive white noise with distribution $\mathcal{CN}(0, 1)$, and $s(t) \in \mathcal{C}^{M \times 1}$ is the transmit symbol satisfying $E\|s(t)\|^2 = M$. Therefore, the received signal to noise ratio (SNR) of the i -th user will be $E\rho|H_i s(t)|^2 = P = M\rho$; however, to simplify the notation we refer to ρ as the SNR of the users.

We assume that H_i is known perfectly at the receiver. We denote the (average) rate of the i -th user, $i = 1, \dots, n$, over the different channel realizations by R_i .

In this paper, we are interested in analyzing differentiated rate scheduling schemes for broadcast systems with n users. We consider a partitioning of the users into K groups $\mathcal{G}_1, \dots, \mathcal{G}_K$, where different groups require different rates from the transmitter. The sizes of the groups are all of the same order and hence, the cardinality of \mathcal{G}_k is $\alpha_k n$ where K and α_k 's are fixed numbers such that $\sum_{i=1}^K \alpha_i = 1$.

Assuming that the average rate of a user in the k 'th group is denoted by R^k ,³ we impose the constraint that the average rate of a user in the k -th group is β_k times the average rate of a user

in the K -th group. The β_k 's are fixed numbers independent of n . In general we are interested in the following problem.

Problem 1. Consider the fading MIMO GBC with users as described above. Let R_i denote the rate to the i -th user and R^k denote the rate to a user in group k . Then construct a transmission scheme such that

$$\begin{aligned} & \max_{(R_1, \dots, R_n) \in \mathcal{C}_{BC}} \sum_{i=1}^n R_i \\ & \text{subject to } \frac{R^k}{R^K} = \beta_k, \quad k = 1, \dots, K \end{aligned} \quad (2)$$

where \mathcal{C}_{BC} is the capacity region of the broadcast channel given in [7], [13].

Clearly, the solution to Problem 1 is given by the intersection of the line $R^k/R^K = \beta_k$, $k = 1, \dots, K-1$ with the boundary of the capacity region of the broadcast channel. Using the duality between the capacity region of Gaussian broadcast channel and multiple access channel [9], [14], [15], Problem 1 can be solved using bisection method in the following way.

- 1) Choose a set of rates R'^k satisfying the rate constraints of (2).
- 2) By appealing to the dual MAC, solve the problem, $\min \sum p_i$, subject to the rates R'^k .
- 3) If the minimum sum of powers, $\min \sum p_i$, is less than total transmit power, ρM , then the rate vector is achievable. Increase the rate proportionately (according to vector $\underline{\beta} = (\beta_1, \dots, \beta_K)$) and go to 1.
- 4) Otherwise decrease the rates proportionately (according to vector $\underline{\beta} = (\beta_1, \dots, \beta_K)$) and go to 1.

While this is all fine, the algorithm is computationally-intensive, requires full CSI and finally requires implementation of DPC. Furthermore, the solution to Problem 1 does not give us insight into how much throughput loss we would incur by imposing the rate constraints.

In this paper, we look into devising simpler scheduling schemes such as time-sharing and opportunistic transmission. We also compare the performance of different scheduling schemes in terms of their sum-rate. It is clear that there is a price to pay in terms of throughput (sum-rate) to maintain the rate constraints. This is due to the fact that we are not working on the sum-capacity point and therefore, the throughput will be reduced compared to the case where we had no rate constraint. Therefore, we look at the difference of the achievable throughput subject to the rate constraints of (2) with the sum-rate when no rate constraints are present.⁴

In this paper, we consider a system with many users and, rather than attempting to solve Problem 1 directly, we propose three scheduling schemes, namely, weighted opportunistic beamforming (WO), time division opportunistic beamforming (TO), and superposition coding (SC) for single antenna systems. In WO, a generalization of opportunistic random beamforming, each beam is assigned to the user that has

¹It is possible to extend our results to $N \neq 1$ in a straightforward fashion. However, for simplicity, we shall not do so here. From a practical point of view $N = 1$ is also very reasonable.

²We should remark that, although the assumption of a constant channel for T channel uses is critical, the requirement that the channels vary independently from one coherence interval to the next is not.

³Throughout the paper, we use superscript k to refer to any user in \mathcal{G}_k .

⁴For single antenna systems we look at the difference between the sum-capacity and the achievable throughput with rate constraints of (2). For multiple antennas we look at the difference of scheduling schemes with the opportunistic beamforming schedule.

the largest “weighted” signal to noise and interference ratio (SINR) corresponding to that beam. In TO, each group has its own time slot in which the transmitter chooses the user with the best SINR from the corresponding group. Superposition coding (SC) is the scheme that achieves the information-theoretic capacity region for a single antenna broadcast channel. For each scheduling we give an explicit scheme to guarantee the rate constraints.

III. PRELIMINARY RESULTS FOR MIMO GBC

In this section, we review some of the results for MIMO Gaussian broadcast channels. In particular, we will look at the sum-capacity of these channels and revisit the concept of “opportunistic” beamforming and analyze the difference between the achievable sum-rate using this scheme and the sum-capacity point for large systems.

A. The Capacity Region of MIMO GBC

It is shown in [7] that the capacity region of MIMO Gaussian broadcast channels is achieved by a precoding scheme called Dirty Paper Coding (DPC)[8]. Furthermore it is shown in [9], [14], [15] that the capacity region of the MIMO GBC is equal to the capacity region of a *dual* multiple access channel (MAC) with sum-power constraint equal to the GBC power constraint. In particular the ergodic sum-capacity can be written as

$$C_{sum} = \mathbb{E} \max_{0 \leq P_i: \sum_{i=1}^n P_i \leq M\rho} \log \det(I + \sum_{i=1}^n H_i^* P_i H_i) \quad (3)$$

where the expectation is over all the channel realizations.

Sum-Capacity Scaling Laws: In point-to-point multi-antenna systems the throughput scaling is often equivalent to the “multiplexing gain” defined as $\lim_{P \rightarrow \infty} \frac{C_{sum}}{\log P}$ where C_{sum} denotes the ergodic sum-rate capacity (or throughput) of the channel achieved by coding over several coherence intervals.

In broadcast channels as the number of users can also be large, two different throughput scaling laws can be envisioned with respect to $P = M\rho$ (or equivalently SNR) and with respect to the number of users n . It is shown in [16] that for a MIMO GBC with fixed M and n , $\lim_{P \rightarrow \infty} \frac{C_{sum}}{\log P} = M$. This gain which is referred to as *multiplexing gain*, requires channel knowledge with very high fidelity at the transmitter (indeed a fidelity that grows with the transmit power) [19].

Another gain is obtained for a system with large number of users n . In this case, for large n , fixed M , and P , we have $\lim_{n \rightarrow \infty} \frac{C_{sum}}{\log \log n} = M$ [17]. It is shown in [10] that this gain can be achieved with very little CSI at the transmitter. These are clearly two very different regimes and both confirm that the sum-rate is linearly scaled with the number of transmit antennas M .

In this paper we will focus on the *large n* regime. Note that, in order to perform DPC to achieve the sum-capacity (or any other point in the capacity region) the transmitter requires exact knowledge of the channels. In the following, we briefly describe “opportunistic” beamforming, a simple scheme that achieves most of the sum-capacity in some regimes and yet requires very little CSI at the transmitter. For a complete study of this scheme, see [10], [20].

B. Opportunistic Beamforming

The main idea behind opportunistic beamforming is to exploit the multi-user diversity available in the network. In this beamforming, during any coherence interval, the transmitter constructs M random beams and transmits each beam to the user with the highest signal to interference plus noise ratio (SINR). Let $\phi_m(M \times 1)$, for $m = 1, \dots, M$, be M random orthonormal vectors generated according to an isotropic distribution [18]. The transmitted signal is

$$s(t) = \sum_{m=1}^M \phi_m s_m(t), \quad t = 1, \dots, T,$$

where each $s_m(t)$ is a scalar signal (with average unit power, i.e., $\mathbb{E}|s_m|^2 = 1$) intended for one of the users. Assuming the users know their own channel coefficients (a much more reasonable assumption than the transmitter knowing *all* the channel gains to the different users), each user can compute its signal-to-interference-plus-noise-ratio (SINR) for every beam as

$$\text{SINR}_{i,m} = \frac{|H_i \phi_m|^2}{\frac{1}{\rho} + \sum_{l \neq m} |H_i \phi_l|^2}, \quad m = 1, \dots, M. \quad (4)$$

If each user feeds back its best SINR and corresponding beam index to the transmitter, the transmitter can assign each beam to the user that has the best SINR for that beam. It is shown in [10] that opportunistic beamforming is order-optimal in the large n regime, but not in the large P regime. The reason is that opportunistic beamforming is interference dominated and so the sum-rate does not scale with the logarithm of the power.

Before we proceed further, it is useful to mention that the probability distribution function (PDF) and the cumulative distribution function (CDF) of $\text{SINR}_{i,m}$, denoted by $f_s(x)$ and $F_s(x)$ respectively, can be written as [10]

$$\begin{aligned} f_s(x) &= \frac{e^{-x/\rho}}{(1+x)^M} \left(\frac{1}{\rho}(1+x) + M - 1 \right), \\ F_s(x) &= 1 - \frac{e^{-x/\rho}}{(1+x)^{M-1}}, \quad x \geq 0. \end{aligned} \quad (5)$$

C. Tighter Scaling Laws

In this section, we give a tighter result regarding the convergence of the sum-rate of opportunistic beamforming scheme to the sum-capacity.⁵

Theorem 1. Consider the fading MIMO Gaussian broadcast channel of Section II. For fixed ρ and M

$$C_{sum} = M \log \log n + M \log \rho + O\left(\frac{\log \log n}{\log n}\right), \quad (6)$$

where C_{sum} , as defined in (3), refers to the sum-capacity of the broadcast channel with n users.

⁵The following notation will be used in this paper. For two functions f, g defined on natural numbers we have $f(n) = O(g(n))$ if $\lim_{n \rightarrow \infty} \inf f(n)/g(n) < \infty$, we have $f(n) = \Omega(g(n))$ if $\lim_{n \rightarrow \infty} \inf f(n)/g(n) > 0$. Finally we have $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$

Proof: We use the following upper bound on the sum-capacity of GBC derived in [17]

$$C_{sum} \leq E M \log \left(1 + \rho \max_{1 \leq i \leq n} \|H_i\|^2 \right). \quad (7)$$

Now $\|H_i\|^2$'s for $i = 1, \dots, n$ have $\chi^2(2M)$ distribution and they are independent. Therefore, using order statistics results [21], it can be shown that $\max_i \|H_i\|^2$ with high probability behaves as $\log n$. More precisely,

$$\begin{aligned} \Pr \left(2(M-2) \leq \frac{\max_i \|H_i\|^2 - \log n}{\log \log n} \leq 2M \right) \\ = 1 - O \left(\frac{1}{(\log n)^2} \right) \end{aligned}$$

We can split the expectation in the right hand side of (7) in two parts; one is the expectation conditioned on $\max_i \|H_i\|^2$ being less than or equal to $\log n + 2M \log \log n$ and the other is the expectation conditioned on $\max_i \|H_i\|^2$ being greater than $\log n + 2M \log \log n$. This way, we can upper bound the expectation in (7) as,

$$\begin{aligned} E \log \left(1 + \rho \max_i \|H_i\|^2 \right) \\ \leq \log (1 + \rho (\log n + 2M \log \log n)) \\ + \int_{\log n + 2M \log \log n}^{\infty} \log (1 + \rho x) f_m(x) dx, \end{aligned}$$

where $f_m(x)$ is the distribution of $x = \max_i \|H_i\|^2$. It is quite straightforward to show that the second term in the right hand side behaves like $O(\frac{\log \log n}{\log n})$ for large n and the first term scales like $\log \log n + \log \rho + O(\frac{\log \log n}{\log n})$. This shows that, $C_{sum} \leq M \log \log n + M \log \rho + O(\frac{\log \log n}{\log n})$.

To lower bound the sum-capacity we employ an opportunistic beamforming schedule. In order to write the sum-rate achieved by random beamforming, we have to take into account the probability that there can be a user that has the best SINR for two different beams. However, based on [10], the probability that a user has the best SINR for two beams is at most of order $\frac{1}{n}$ and the contribution of this event is asymptotically negligible in all our analysis. Furthermore, we can show that with probability of at least $1 - O(\frac{1}{\log n})$, the maximum SINR for each beam is at least $\log n - 2M \log \log n$. Therefore, the sum-rate of the opportunistic beamforming asymptotically behaves like

$$\begin{aligned} C_{ob} &= ME \log \left(1 + \rho \max_{1 \leq i \leq n} \text{SINR}_{1,i} \right) \\ &\geq M \log (1 + \rho (\log n - 2M \log \log n)) \\ &\quad \cdot \Pr \left(\max_{1 \leq i \leq n} \text{SINR}_{1,i} > \log n - 2M \log \log n \right) \\ &= M \log \log n + M \log \rho + O(\frac{\log \log n}{\log n}). \end{aligned}$$

Therefore, C_{ob} matches the upper bound on the sum-capacity. ■

Theorem 1 implies that the difference of the sum-rate achieved by beamforming and DPC tends to zero at least as fast as $\frac{\log \log n}{\log n}$, i.e. $(C_{sum} - C_{ob}) = O(\frac{\log \log n}{\log n})$. This is a much stronger result than being simply order optimal.

Next we look at the differentiated rate scheduling problem.

IV. TIME-DIVISION OPPORTUNISTIC (TO) BEAMFORMING

The simplest scheme to give differentiated rates to different users is to assign different numbers of channel uses to different users, i.e., time-sharing. This should be done opportunistically to maximize the sum-rate. In particular, we divide each coherence interval into K slots of duration t_k each, $k = 1, \dots, K$. During the k -th subinterval the transmitter performs opportunistic beamforming to *only* the $\alpha_k n$ users in the k -th group. If the α_k 's are fixed and n grows, it can be shown that to satisfy the rational rate constraints, we must have

$$\frac{t_k}{T} = \frac{\alpha_k \beta_k}{\sum_{l=1}^K \alpha_l \beta_l}, \quad k = 1, \dots, K \quad (8)$$

Intuitively, since each group has a size of order n , the sum-rate for each group scales like $M \log \log n$. Therefore in order to satisfy the rate constraints, we need to only take into account the size of the group, i.e., α_i , and the rate ratio β_i . Therefore, using (8), we can easily show that

$$\lim_{n \rightarrow \infty} \frac{R^k}{R^K} = \beta_k, \quad k = 1, \dots, K-1.$$

The following result quantifies the sum-rate loss due to the rate constraints and also the sum-optimality of the scheduling.

Theorem 2. Consider the fading MIMO Gaussian broadcast channel of Section II. Let M , ρ , α_k , and β_k be fixed and t_k 's be chosen as (8). Then the rate constraints of (2) are met and

$$\lim_{n \rightarrow \infty} (C_{ob} - C_{tdob}) = \Theta \left(\frac{1}{\log n} \right), \quad (9)$$

where C_{tdob} represents the sum-rate for the time-division opportunistic scheme.

Proof: That the rate constraints are met is fairly straightforward. Now in order to find the difference between the throughputs in the two cases we have

$$C_{ob} - C_{tdob} = \int_0^\infty h(x) dx,$$

where

$$h(x) = M \log(1+x) n f_s(x) \left(F_s(x)^{n-1} - \sum_{k=1}^K \frac{\alpha_k t_k}{T} F_s(x)^{\alpha_k n-1} \right).$$

Define

$$\begin{aligned} l^- &= \rho (\log n - (M+3) \log \log n), \\ l^+ &= \rho (\log n + (M+3) \log \log n). \end{aligned} \quad (10)$$

We break the integral into the following three regions: $\mathcal{I}_1 = [0, l^-]$, $\mathcal{I}_2 = [l^-, l^+]$ and $\mathcal{I}_3 = [l^+, \infty]$. As for the third region we have

$$\begin{aligned} \left| \int_{\mathcal{I}_3} h(x) dx \right| &\leq \int_{\mathcal{I}_3} M n \log(1+x) f_s(x) dx \\ &\leq \frac{1}{(\rho \log n)^{M-1}} \int_{\mathcal{I}_3} M n \log(1+x) \frac{e^{-x/\rho}}{\rho} dx \\ &= O \left(\frac{n(-\log(1+x) e^{-\frac{x}{\rho}})|_{l^+}^\infty)}{\rho (\log n)^{M-1}} - e^{\frac{1}{\rho}} \text{Ei} \left(-\frac{l^+}{\rho} \right) \right), \end{aligned} \quad (11)$$

where $\text{Ei}(-x) = -\int_x^\infty \frac{e^{-y}}{y} dy$ is the exponential integral. For

large x we have $\text{Ei}(-x) = \Theta(-\frac{e^{-x}}{x})$. Using this identity in the previous inequality gives $\left| \int_{\mathcal{I}_3} h(x) dx \right| = O(\frac{\log \log n}{(\log n)^{2M+2}})$. Now we move on to the integrals over \mathcal{I}_1 and \mathcal{I}_2 . It can be easily checked that

$$\begin{aligned} & \left| \int_{\mathcal{I}_1} h(x) dx \right| \\ &= O(\log \log n \int_{\mathcal{I}_1} M n f_s(x) F_s(x)^{\min\{\alpha_k\}n-1} dx) \\ &= O(\log \log n F_s(x)^{\min\{\alpha_k\}n} \big|_0^{l^-}) \\ &= O\left(\log \log n \left(1 - \frac{(\log n)^{M+3}}{n(1+\rho \log n)^{M-1}}\right)^{\min\{\alpha_k\}n}\right) \\ &= O(\log \log n \cdot e^{-(\log n)^4}). \end{aligned} \quad (12)$$

Using integration by parts we have

$$\begin{aligned} \int_{\mathcal{I}_2} h(x) dx &= \log(1+x) (F_s(x)^n - \sum_{k=1}^K \frac{t_k}{T} F_s(x)^{\alpha_k n}) \big|_{l^-}^{l^+} \\ &\quad - M \int_{\mathcal{I}_2} \frac{F_s(x)^n - \sum_{k=1}^K \frac{t_k}{T} F_s(x)^{\alpha_k n}}{1+x} dx. \end{aligned}$$

We can further simplify the integral to

$$\begin{aligned} \int_{\mathcal{I}_2} h(x) dx &= \Theta\left(\frac{\log \log n}{(\log n)^{2M+2}}\right) - \frac{M}{\rho \log n} \left(1 - O\left(\frac{\log \log n}{\log n}\right)\right) \\ &\quad \cdot \int_{l^-}^{l^+} \left(F_s(x)^n - \sum_{k=1}^K \frac{t_k}{T} F_s(x)^{\alpha_k n}\right) dx. \end{aligned}$$

Next, we show that the last integral in the above equation is of order constant. Suppose $i = \arg \min_k \{\alpha_k\}$. Note that with this condition $\alpha_i \leq \frac{1}{2}$. Now the series of equations in (13) verify that the integral is of order constant, where (a) follows by a change of variable $y = \frac{x}{\rho} - \log n + (M-1) \log \log n$ and (b) holds because $(1 - \frac{e^{-y}}{n(1+\Theta(\frac{\log \log n}{\log n}))^{M-1}})^n = \Theta\left((1 - O(\frac{(\log n)^4}{n})) \exp(-\nu(n)e^{-y})\right)$, for large n , $\nu(n) = (1 + O(\frac{\log \log n}{\log n}))^{M-1}$, and for y in the range defined in the integral. Also (c) follows after change of variable of $z = \nu(n)e^{-y}$ and (d) is a direct consequence of the following asymptotic expansions for $-\text{Ei}(-x)$ for small and large x respectively [22],

$$\begin{aligned} \text{Ei}(-x) &= \ln x + \gamma_0 + O(x), & x \ll 1, \\ \text{Ei}(-x) &= -\frac{e^{-x}}{x} (1 + O(\frac{1}{x})) & x \gg 1. \end{aligned}$$

Putting Equations (12) to (13) together, we get that $(C_{ob} - C_{tdob}) = \Theta(\frac{1}{\log n})$. ■

V. WEIGHTED OPPORTUNISTIC (WO) BEAMFORMING

In the weighted opportunistic scheme we weigh the SINR of each user according to its group by μ_k , $k = 1, \dots, K$. Then during each coherence interval, the transmitter assigns the M random beams to the M users that have the largest *weighted* SINR. More specifically, let the SINR corresponding to beam m and user i be denoted by $x_{i,m}$. The distribution of $x_{i,m}$ was given in (5). With this notation, beam m is

assigned to the user l_0 from group k_0 such that $(k_0, l_0) = \arg \max_{(k,i): i \in \mathcal{G}_k} \mu_k x_{i,m}$. Equivalently, one can see that beam m assigned to group k_0 such that

$$k_0 = \arg \max_{1 \leq k \leq K} \mu_k \max_{i \in \mathcal{G}_k} x_{i,m} \quad (14)$$

In the WO scheme there are two questions to be answered. First, how to determine the weights explicitly such that the rational rate constraints are met. Here, unlike the TO case, the answer is not trivial. And second, what is the rate loss compared to the unconstrained sum-rate capacity of the broadcast channel itself. The following theorems settle the aforementioned questions.

Theorem 3. Consider the fading MIMO Gaussian broadcast channel of Section II. Consider the WO beamforming scheme with

$$\mu_k = 1 + \frac{\log \beta_k}{\log n - (M-1) \log \log n}. \quad (15)$$

Assuming, M , ρ , α_k 's and β_k 's are fixed, we have

$$\lim_{n \rightarrow \infty} \frac{R^k}{R^K} = \beta_k, \quad k = 1, \dots, K. \quad (16)$$

Proof: Let us look at the average transmitted rate to the first user in the first group. Clearly,

$$\begin{aligned} R^1 &= M \int_0^\infty \log(1+x) f_s(x) \\ &\quad \cdot \Pr \left(\mu_1 x_{1,m} \geq \mu_k x_{i,m}, \quad 1 \leq k \leq K, i \in \mathcal{G}_k \mid x_{1,m} = x \right) dx \end{aligned}$$

Using the independence of the SINRs for different users, the probability of the event defined inside the above integral can be written as

$$\prod_{k=1}^K \prod_{i \in \mathcal{G}_k, i \neq 1} \Pr \left(x_{i,m} \leq \frac{\mu_1}{\mu_k} x \right) = F_s(x)^{\alpha_1 n - 1} \prod_{k=2}^K F\left(\frac{\mu_1}{\mu_k} x\right)^{\alpha_k n}.$$

Accordingly, we have

$$R^1 = M \underbrace{\int_0^\infty \log(1+x) f_s(x) F_s(x)^{\alpha_1 n - 1} \prod_{k=2}^K F\left(\frac{\mu_1}{\mu_k} x\right)^{\alpha_k n} dx}_{h_1(x)}.$$

We further split the above integral to three integrals over the intervals $\mathcal{I}_1 = [0, l^-]$, $\mathcal{I}_2 = [l^-, l^+]$ and $\mathcal{I}_3 = [l^+, \infty]$, where l^- and l^+ are defined in (10). The integral over the first region can be written as

$$\begin{aligned} \int_{\mathcal{I}_1} h_1(x) dx &\stackrel{(a)}{=} O(\log \log n \int_{\mathcal{I}_1} M f_s(x) F_s(x)^{\alpha_1 n - 1} dx) \\ &= O\left(\frac{\log \log n}{n} F_s(x)^{\alpha_1 n} \big|_0^{l^-}\right) \\ &= O(\log \log n \cdot e^{-(\log n)^4}), \end{aligned}$$

where (a) follows from the fact that $0 \leq F_s(x) \leq 1$. Similarly, it can be shown that for the integral over the third region we have $\int_{\mathcal{I}_3} h_1(x) dx = O(\frac{\log \log n}{(\log n)^{2M+2}})$. Hence, the main contribution is due to the integral over the second interval.

$$\begin{aligned}
& \int_{l^-}^{l^+} \left(F_s(x)^n - \sum_{k=1}^K \frac{t_k}{T} F_s(x)^{\alpha_k n} \right) dx = \Theta \left(\int_{l^-}^{l^+} (F_s(x)^n - F_s(x)^{\alpha_i n}) dx \right) \\
& \stackrel{(a)}{=} \Theta \left(\int_{-4 \log \log n}^{(2M+2) \log \log n} \left(1 - \frac{e^{-y}}{n(1 + O(\frac{\log \log n}{\log n}))^{M-1}} \right)^n - \left(1 - \frac{e^{-y}}{n(1 + O(\frac{\log \log n}{\log n}))^{M-1}} \right)^{\alpha_i n} dy \right) \\
& \stackrel{(b)}{=} \Theta \left(\int_{-4 \log \log n}^{(2M+2) \log \log n} \exp(-\nu(n)e^{-y}) - \exp(-\nu(n)\alpha_i e^{-y}) dy \right) \\
& \stackrel{(c)}{=} \Theta \left(\int_{\frac{1}{(\log n)^{2M+2} \nu(n)}}^{(\log n)^4 \nu(n)} \frac{e^{-z} - e^{-\alpha_i z}}{z} dz \right) \stackrel{(d)}{=} \Theta(\ln \alpha_i + \frac{\log \log n}{\log n}) = \Theta(1),
\end{aligned} \tag{13}$$

Looking at the behavior of $F_s(\frac{\mu_1}{\mu_k}x)^{\alpha_k n}$ for $x \in \mathcal{I}_2$ we have where

$$\begin{aligned}
F_s(\frac{\mu_1}{\mu_k}x) &= F_s \left(\frac{\mu_1}{\mu_k} (y + \rho(\log n - (M-1) \log \log n)) \right) \\
&\stackrel{(a)}{=} F_s \left(\rho \left(y(1 + O(\frac{1}{\log n})) + \log(\frac{\beta_1}{\beta_k}) + \Delta \right) + O(\frac{1}{\log n}) \right),
\end{aligned} \tag{17}$$

where $\Delta = \log n - (M-1) \log \log n$ and (a) follows by defining $y = \frac{x}{\rho} - \log n + (M-1) \log \log n$, where $y \in \mathcal{I}'_2 = [(2M+2) \log \log n, -4 \log \log n]$, and using the definition of μ_k 's in the theorem statement. Further simplification gives

$$\begin{aligned}
F_s(\frac{\mu_1}{\mu_k}x)^{\alpha_k n} &\stackrel{(b)}{=} \left(1 - \frac{\beta_k e^{-y(1+O(\frac{1}{\log n}))} (1 - O(\frac{\log \log n}{\log n}))}{\beta_1 \rho^{M-1}} \right)^{\alpha_k n} \\
&= \exp \left(\alpha_k n \log \left(1 - \frac{\beta_k e^{-y(1+O(\frac{1}{\log n}))} (1 - O(\frac{\log \log n}{\log n}))}{\beta_1 \rho^{M-1}} \right) \right) \\
&= \left(1 - O(\frac{\log \log n}{\log n}) \right) \exp \left(-\frac{\alpha_k \beta_k}{\beta_1 \rho^{M-1}} e^{-y} \right),
\end{aligned}$$

where (b) follows by substituting $F_s(x)$ in (17) with its expression from (5) and noting that for $y \in \mathcal{I}'_2$ we have, $(1 + \frac{y}{\log n})^{1-M} = 1 - O(\frac{\log \log n}{\log n})$. The last equality is obtained by expanding the logarithm. Using a similar argument, it can be shown that for $y \in \mathcal{I}'_2$,

$$f_s(\rho(y + (\log n - (M-1) \log \log n))) = \frac{e^{-y}}{n} \left(\frac{1}{\rho^M} - O(\frac{\log \log n}{\log n}) \right).$$

Using the above expressions for $F_s(\frac{\mu_1}{\mu_k}x)^{\alpha_k n}$ and $f_s(x)$, the integral over \mathcal{I}_2 can be written as

$$\begin{aligned}
\int_{\mathcal{I}_2} h_1(x) dx &= B_0 \int_{\mathcal{I}'_2} e^{-y} \exp \left(-\frac{\sum_{k=1}^K \alpha_k \beta_k}{\beta_1 \rho^{M-1}} e^{-y} \right) dy \\
&= B_0 \int_{(\log n)^{-2M-2}}^{(\log n)^4} \exp \left(-\frac{\sum_{k=1}^K \alpha_k \beta_k}{\beta_1 \rho^{M-1}} z \right) dz \\
&\stackrel{(a)}{=} B_0 \frac{e^{-A_0 z}}{A_0} \Big|_{(\log n)^4}^{(\log n)^{-2M-2}} \\
&= M \frac{\log \log \rho n}{\rho^{M-1} A_0 n} (1 - O(\frac{\log \log n}{n \log n})) \\
&= M \frac{\beta_1}{\sum_{k=1}^K \beta_k \alpha_k} \frac{\log \log \rho n}{n} + O(\frac{\log \log n}{n \log n}),
\end{aligned}$$

$$\begin{aligned}
B_0 &= M \frac{\log \log \rho n}{n \rho^{M-1}} \left(1 + O(\frac{\log \log n}{\log n}) \right), \\
A_0 &= -\frac{\sum_{k=1}^K \alpha_k \beta_k}{\beta_1 \rho^{M-1}}.
\end{aligned}$$

and $z = e^{-y}$ in (a). Putting these integrals together we have

$$R^1 = M \frac{\beta_1}{\sum_{k=1}^K \beta_k \alpha_k} \frac{\log \log \rho n}{n} + O(\frac{\log \log n}{n \log n}).$$

Similarly, we can check that for a user in group k the average transmitted rate is

$$R^k = M \frac{\beta_k}{\sum_{k=1}^K \beta_k \alpha_k} \frac{\log \log \rho n}{n} + O(\frac{\log \log n}{n \log n}),$$

and hence the rational constraints are readily obtained. The achievable sum-rate in this case is

$$\begin{aligned}
C_{wob} &= \sum_{k=1}^K \alpha_k n R^k \\
&= M n \frac{\sum_{k=1}^K \beta_k \alpha_k}{\sum_{j=1}^K \beta_j \alpha_j} \frac{\log \log \rho n}{n} + O(\frac{\log \log n}{\log n}) \\
&= M \log \log \rho n + O(\frac{\log \log n}{\log n}).
\end{aligned}$$

Comparing this with the sum-capacity we see that $\lim_{n \rightarrow \infty} (C_{wob} - C_{ob}) = 0$. ■

Remark 4. Theorem 3 asserts that the average rates of users are quite sensitive to the change of μ_i 's. Following the methodology in the proof of Theorem 3, we can prove that for a two group system

$$\begin{aligned}
\frac{\mu_1}{\mu_2} &= 1 - o\left(\frac{1}{\log n}\right) \implies \lim_{n \rightarrow \infty} \frac{R^1}{R^2} = 1, \\
\frac{\mu_1}{\mu_2} &= c < 1 \implies \lim_{n \rightarrow \infty} \frac{R^1}{R^2} = 0,
\end{aligned}$$

where c is a constant independent of n .

The above Theorem also shows that, as in the case of TO beamforming, WO beamforming achieves the sum-rate of the unconstrained broadcast channel as $n \rightarrow \infty$. As a matter of fact the simulation results of Section VII suggest that throughput of the WO beamforming is much closer to the sum-capacity than the throughput of the TO beamforming. In the next lemma we look at the difference between the throughput

of WO and the sum-rate of opportunistic beamforming with no rate constraints for a system with two groups. Carrying the proof over to more than two groups is a possible, however, cumbersome task.

Lemma 5. *Consider the setting of Theorem 3 and assume that there are two groups present in the network. Then for any group size $\alpha_i n$, and any β_i 's for $i = 1, 2$, the sum-rate loss of WO tends to zero with a convergence rate at least as fast as $\frac{1}{\log n}$, i.e.,*

$$\lim_{n \rightarrow \infty} (C_{ob} - C_{wob}) = O\left(\frac{1}{\log n}\right). \quad (18)$$

Proof: For each beam m , let x_m and y_m denote the maximum SINR in groups \mathcal{G}_1 and \mathcal{G}_2 respectively. Furthermore, assume that $\beta_1 \geq \beta_2$. We can write $C_{ob} - C_{wob}$, the loss in the throughput, as

$$\sum_{m=1}^M \mathbb{E} \log \left(\frac{1 + x_m \mathbb{1}(x_m \geq y_m) + y_m \mathbb{1}(x_m < y_m)}{1 + x_m \mathbb{1}(x_m \geq \frac{\mu_2}{\mu_1} y_m) + y_m \mathbb{1}(x_m < \frac{\mu_2}{\mu_1} y_m)} \right).$$

The difference can be further simplified to get⁶

$$C_{ob} - C_{wob} = \sum_{m=1}^M \mathbb{E} \left(\log\left(\frac{1 + y_m}{1 + x_m}\right) | y_m \geq x_m \geq \frac{\mu_2}{\mu_1} y_m \right).$$

Note that for $y_m \geq x_m \geq \frac{\mu_2}{\mu_1} y_m$ we have $\frac{1+y_m}{1+x_m} \leq \frac{1+y_m}{1+\frac{\mu_2}{\mu_1} y_m} \leq \frac{\mu_1}{\mu_2}$. Therefore, the sum-rate loss is upperbounded as $C_{ob} - C_{wob} \leq \sum_{m=1}^M \log\left(\frac{\mu_1}{\mu_2}\right) \cdot \Pr(y_m \geq x_m \geq \frac{\mu_2}{\mu_1} y_m)$. Using the fact that $\Pr(y_m \geq x_m \geq \frac{\mu_2}{\mu_1} y_m) = \Theta(1)$ and $\frac{\mu_1}{\mu_2} = 1 + \Theta(\frac{1}{\log n})$ we get $C_{ob} - C_{wob} = O(\frac{1}{\log n})$. ■

VI. SUPERPOSITION CODING FOR SINGLE ANTENNA BROADCAST CHANNELS

In this section, we analyze the performance of superposition coding for the case when there are only two groups of users $\mathcal{G}_1, \mathcal{G}_2$ with sizes $\alpha_1 n$ and $\alpha_2 n$. We assume that the average rate provided to a user in the first group is required to be $\beta > 1$ times the rate provided to a user in the second group. It should be mentioned that the ergodic capacity region of a broadcast channel with two users has been studied in [4]; here we look at a generalization of the result of [4] in which we have n users divided into two groups with different rate demands.

In order to maximize the rate (sum-rate) while keeping the ratio of different group rates fixed and equal to β , we need to find the point on the boundary of the capacity region of the Gaussian broadcast channel that satisfies the rate constraint. We know that every boundary point is the solution to a maximization problem of form

$$\max_{(R_1, \dots, R_n) \in \mathcal{C}_{BC}} \sum_{i=1}^n \mu_i R_i$$

for some positive values of μ_1, \dots, μ_n . In our case because of the symmetry among the users in each group, the values of μ_i 's will be the same for the users in the same group.

⁶In this paper we use the notation $\mathbb{E}(f(x)|x \in \mathcal{A})$ to denote $\Pr(x \in \mathcal{A}) \cdot \int_{x \in \mathcal{A}} f(x) dx$.

Therefore, we only need to characterize the boundary points that are the maximizing solution to the problem

$$\max_{(R_1, \dots, R_n) \in \mathcal{C}_{BC}} \mu_1 \left(\sum_{i \in \mathcal{G}_1} R_i \right) + \mu_2 \left(\sum_{i \in \mathcal{G}_2} R_i \right),$$

for $\mu_1, \mu_2 > 0$. The following lemma characterizes such boundary points.

Lemma 6. *Consider a scalar Gaussian broadcast system with the model described in Section II. Consider the following optimization problem*

$$\max_{(R_1, \dots, R_n) \in \mathcal{C}_{BC}} \mu_1 \left(\sum_{i \in \mathcal{G}_1} R_i \right) + \mu_2 \left(\sum_{i \in \mathcal{G}_2} R_i \right), \quad (19)$$

where \mathcal{C}_{BC} is the ergodic capacity region of broadcast channel with power constraint ρ and $\mu_1 \geq \mu_2 \geq 0$. Then the solution of the above optimization problem is

$$\begin{aligned} \alpha_1 n R_i &= E(\log(1 + \rho x) | \mu_1 x \geq \mu_2 y) \\ &+ E(\log(\frac{(\mu_1 - \mu_2)y(1 + \rho x)}{\mu_1(y - x)}) | (x, y) \in \mathcal{R}), \end{aligned} \quad (20)$$

for $i \in \mathcal{G}_1$. Similarly, for $i \in \mathcal{G}_2$, we have

$$\begin{aligned} \alpha_2 n R_i &= E(\log(1 + \rho y) | \mu_1 x \leq \mu_2 y) \\ &- E(\log(\frac{(\mu_1 - \mu_2)x(1 + \rho y)}{\mu_2(y - x)}) | (x, y) \in \mathcal{R}), \end{aligned} \quad (21)$$

where $x = \max_{i \in \mathcal{G}_1} |h_i|^2$, $y = \max_{i \in \mathcal{G}_2} |h_i|^2$ and region \mathcal{R} is defined as

$$\mathcal{R} = \{(x, y) \in \mathbb{R}^+ \times \mathbb{R}^+ | 0 \leq \frac{\mu_2}{(\mu_1 - \mu_2)x} - \frac{\mu_1}{(\mu_1 - \mu_2)y} \leq \rho\}. \quad (22)$$

Proof: The duality between the broadcast channel and the multi-access channel for the scalar case in [9] states that $\mathcal{C}_{BC} = \bigcup_{\sum_i P_i(\underline{h}) = \rho} \mathcal{C}_{MAC}(P_1(\underline{h}), \dots, P_n(\underline{h}))$, where $\underline{h} = (h_1, \dots, h_n)$, $P_i(\underline{h})$ is the power allocation function of user i and the union is over all the permissible power allocation functions. Furthermore $\mathcal{C}_{MAC}(P_1(\underline{h}), \dots, P_n(\underline{h}))$ is

$$\{\underline{R} : \sum_{i \in \mathcal{S}} R_i \leq \log(1 + \sum_{i \in \mathcal{S}} P_i(\underline{h}) |h_i|^2), \forall \mathcal{S} \subset \{1, \dots, n\}\}.$$

Using the above region, we can rewrite (19) as a maximization problem over all the power allocation functions and all the corresponding rate vectors in the capacity region of the dual multi-access channel. Based on this, it can be verified that the maximum of (19) occurs when we send only to the users with the best channel in each group. Therefore,

$$V_o = \mathbb{E} \max_{\substack{P_x, P_y \\ P_x + P_y = \rho}} (\mu_1 - \mu_2) \log(1 + P_x x) + \mu_2 \log(1 + P_x x + P_y y),$$

where $x = \max_{i \in \mathcal{G}_1} |h_i|^2$, $y = \max_{i \in \mathcal{G}_2} |h_i|^2$. Performing the maximization over P_x, P_y we have one of the following possibilities:

- 1) If $\mu_1 x \geq \mu_2 y$, we assign all the power to the best user of the first group.
- 2) If $0 \leq \frac{\mu_2}{(\mu_1 - \mu_2)x} - \frac{\mu_1}{(\mu_1 - \mu_2)y} \leq \rho$ then we split the power between the two best users in the two groups as $P_x = \frac{(\mu_1 - \mu_2)P_x y + \mu_1 x - \mu_2 y}{\mu_1(y - x)}$ and $P_y = \rho - P_x$.
- 3) If $\frac{\mu_2}{(\mu_1 - \mu_2)x} - \frac{\mu_1}{(\mu_1 - \mu_2)y} > \rho$, all the power is assigned

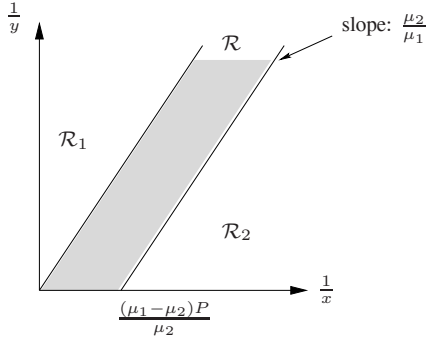


Fig. 1. The decision region for power allocation in the superposition coding in two group case: If $(x, y) \in \mathcal{R}_1$, all the power is allocated to the best user of group one. If $(x, y) \in \mathcal{R}_2$, all the power is allocated to the best user of group two. If $(x, y) \in \mathcal{R}$ then power is split between the best users of the two groups.

to the best user in \mathcal{G}_2 .

We have plotted the decision region for power allocation in the $(\frac{1}{x}, \frac{1}{y})$ region in Figure 1. ■

The question that is remained to be answered is to figure out how to choose μ_1 and μ_2 such that the rate constraint in (2) is satisfied. This is answered in the following theorem.

Theorem 7. Suppose $\beta > 1$ is fixed, $\mu_1 = 1$, and $\mu_2 = 1 - \frac{1}{(\log n)^\gamma}$ where $\gamma = 1 + \frac{1}{\alpha_1\beta + \alpha_2}$. Then

$$\lim_{n \rightarrow \infty} \frac{R^1}{R^2} = \beta. \quad (23)$$

Proof: Consider the first terms in the rate expressions for groups one and two in (20) and (21) respectively. Based on Remark 4, since $\frac{\mu_2}{\mu_1} = 1 - o(\frac{1}{\log n})$, it can be easily checked that these two terms are of order $\alpha_1 \log \log n$ and $\alpha_2 \log \log n$ respectively. Now we look at the second expectation in (20),

$$A^* = E \left(\log \left(\frac{y(1 + \rho x)}{((\log n)^\gamma - 1)(y - x)} \right) | (x, y) \in \mathcal{R} \right).$$

It can be verified that for $(x, y) \in \mathcal{R}$ the term inside the logarithm is greater than one. Hence, A^* is positive. Furthermore, similar to previous cases, the main contribution of the expectation comes from values of x and y around $\log n$. Therefore, we can simplify A^* as

$$\begin{aligned} A^* &= E \left(\log \left(\frac{y(1 + \rho x)}{((\log n)^\gamma - 1)} \right) | (x, y) \in \mathcal{R} \right) \\ &\quad - \underbrace{E(\log(y - x) | (x, y) \in \mathcal{R})}_{B^*} \\ &= (2 - \gamma) \log \log n (1 - O(\frac{\log \log n}{\log n})) \cdot \Pr((x, y) \in \mathcal{R}) \\ &\quad - B^*. \end{aligned}$$

It can be shown that $\Pr((x, y) \in \mathcal{R}) = \alpha_2 - o(1)$. In the following we show that the expectation term in the above equation is $O(1)$. We have

$$\begin{aligned} B^* &= \Theta \left(\int_{l_1}^{l_2} \alpha_2 n \frac{e^{-y}}{1 - e^{-y}} (1 - e^{-y})^{\alpha_2 n} \right. \\ &\quad \cdot \left. \int_{l_3}^{l_4} \log(y - x) \alpha_1 n \frac{e^{-x}}{1 - e^{-x}} (1 - e^{-x})^{\alpha_1 n} dx dy \right), \end{aligned}$$

where $l_1 = \log n - 4 \log \log n$, $l_2 = \log n + 4 \log \log n$, $l_3(y) =$

$\frac{y(1 - \frac{1}{(\log n)^\gamma})}{1 + \frac{\rho y}{(\log n)^\gamma}}$, and $l_4(y) = y(1 - \frac{1}{(\log n)^\gamma})$. Defining $y - x = z$, $w = y - \log n$ and using Integration by part, the series of equality given in (24) can be shown for B^* . where (a) follows by integrating over w first. Equality (b) follows because in the interval $[1, (\log n)^{2-\gamma}]$, the integrand is upper bounded by e^{-z} and therefore the contribution of the integral in this interval is of order constant. Also for $[(\log n)^{1-\gamma}, 1]$ the numerator of the integrand is $1 - o(1)$. Finally, (c) follows by verifying that the integral is increasing in α_2 and in both extremes, i.e. $\alpha_2 = 1$ and $\alpha_2 = 0$ the integral is $(\gamma - 1) \log \log n + O(1)$.

Putting the evaluated values of A^* and B^* together we get the following rate for a user in group \mathcal{G}_1

$$R^1 = (1 + (2 - \gamma) \frac{\alpha_2}{\alpha_1}) \frac{\log \log n}{n} (1 - O(\frac{\log \log n}{n \log n})).$$

Similarly for a user in group \mathcal{G}_2 we have

$$R^2 = (\gamma - 1) \frac{\log \log n}{n} (1 - O(\frac{\log \log n}{n \log n})).$$

Therefore, to meet the ratio constraints between the rates of users in different groups, we should have $\frac{(\gamma-1)}{(1+(2-\gamma)\frac{\alpha_2}{\alpha_1})} = \frac{1}{\beta}$, or accordingly $\gamma = 1 + \frac{1}{\alpha_1\beta + \alpha_2}$. ■

It is clear that the convergence rate for the superposition coding should be faster than or equal to TO and WO beamforming. In the next lemma, we provide a bound on the difference between the sum-capacity and the sum-rate obtained by the scheduling discussed in this section.

Lemma 8. Suppose $\beta > 1$ is fixed and μ_1, μ_2 are chosen as in Theorem 7. Then

$$C_{sum} - \sum_{i=1}^n R_i = O \left(\frac{1}{(\log n)^{2\gamma-1}} \right),$$

where R_i 's are the rates given to different users in Theorem 7.

Proof: Using (20), we can write the throughput under constraints of (2) as⁷,

$$\begin{aligned} &E \log(1 + \rho x \mathbb{1}(x \geq \mu y) + \rho y \mathbb{1}(x < \mu y)) \\ &+ E \log \left(\frac{(1 + \rho x)y}{(1 + \rho y)x} | (x, y) \in \mathcal{R} \right), \end{aligned}$$

where x and y are the maximum norm channel in group one and two respectively. Therefore, $\Delta(n)$, the difference of the sum-capacity and the throughput given above can be written as

$$\begin{aligned} 0 \leq \Delta(n) &= E \log \left(\frac{1 + \rho x \mathbb{1}(x \geq y) + \rho y \mathbb{1}(x < y)}{1 + \rho x \mathbb{1}(x \geq \mu y) + \rho y \mathbb{1}(x < \mu y)} \right) \\ &\quad - E \log \left(\frac{(1 + \rho x)y}{(1 + \rho y)x} | (x, y) \in \mathcal{R} \right). \end{aligned}$$

We can simplify the right hand side of the above equation to get

$$\begin{aligned} \Delta(n) &= E \log \left(\frac{1 + \rho y}{1 + \rho x} | y \geq x \geq \mu y \right) \\ &\quad + E \log \left(\frac{(1 + \rho y)x}{(1 + \rho x)y} | (x, y) \in \mathcal{R} \right). \end{aligned}$$

It can be easily checked that the second term is positive over

⁷In this proof, we refer to μ_2 as μ .

$$\begin{aligned}
B^* &= \Theta \left(\int_{-4 \log \log n}^{4 \log \log n} \alpha_2 e^{-w} \left(1 - \frac{e^{-w}}{n}\right)^{\alpha_2 n - 1} \int_{(\log n)^{1-\gamma}}^{(\log n)^{2-\gamma}} \log(z) \alpha_1 e^z e^{-w} \left(1 - \frac{e^z e^{-w}}{n}\right)^{\alpha_1 n - 1} dz dw \right) \\
&= \Theta \left(\int_{-4 \log \log n}^{4 \log \log n} \alpha_2 e^{-w} \exp(-\alpha_2 e^{-w}) \int_{(\log n)^{1-\gamma}}^{(\log n)^{2-\gamma}} \log(z) \alpha_1 e^z e^{-w} \exp(-\alpha_1 e^z e^{-w}) dz dw \right) \\
&= \Theta \left(\alpha_2 (\gamma - 1) \log \log n - \int_{-4 \log \log n}^{4 \log \log n} \alpha_2 \frac{\exp(-\alpha_2 e^{-w})}{e^w} \int_{(\log n)^{1-\gamma}}^{(\log n)^{2-\gamma}} \frac{\exp(-\alpha_1 e^z e^{-w})}{z} dz dw \right) \\
&\stackrel{(a)}{=} \Theta \left(\alpha_2 (\gamma - 1) \log \log n - \alpha_2 \int_{(\log n)^{1-\gamma}}^{(\log n)^{2-\gamma}} \frac{\exp(-\alpha_1 e^z \frac{1}{(\log n)^4})}{(\alpha_2 + \alpha_1 e^z) z} dz \right) \\
&\stackrel{(b)}{=} \Theta \left(\alpha_2 (\gamma - 1) \log \log n - \alpha_2 \int_{(\log n)^{1-\gamma}}^1 \frac{1}{(\alpha_2 + (1 - \alpha_2) e^z) z} dz \right) \\
&\stackrel{(c)}{=} \Theta \left(\alpha_2 (\gamma - 1) \log \log n - \alpha_2 \int_{(\log n)^{1-\gamma}}^1 \frac{1}{z} dz \right) = O(1). \tag{24}
\end{aligned}$$

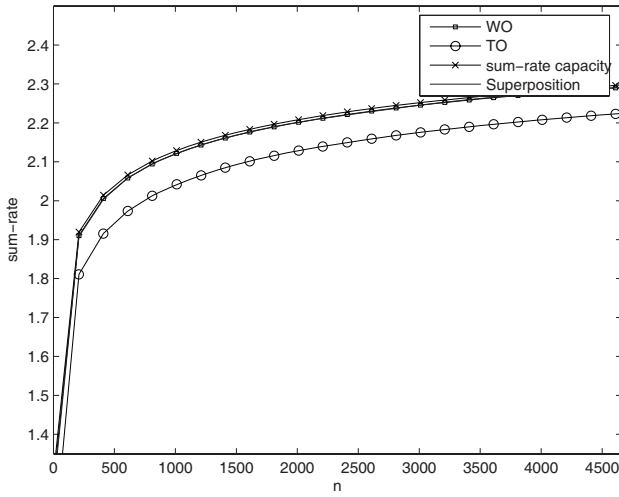


Fig. 2. The sum of the transmitted rates for WO, TO, and SC, as well as the sum-rate capacity of the single antenna broadcast channel as a function of the number of users for a system with $K = 2$ and $\beta = 2$.

region \mathcal{R} defined in (22). Furthermore, $1 \leq \frac{1+\rho y}{1+\rho x} \leq \frac{1}{\mu}$ for $y \geq x \geq \mu x$. Therefore, we have

$$\begin{aligned}
\Delta(n) &\leq \mathbb{E} \log \left(\frac{1+\rho y}{1+\rho x} \middle| y \geq x \geq \mu y \right) \\
&\leq -\log \mu \cdot \Pr(y \geq x \geq \mu y). \tag{25}
\end{aligned}$$

Using the techniques developed so far, it can be shown that $\Pr(y \geq x \geq \mu y) = O\left(\frac{1}{(\log n)^{\gamma-1}}\right)$. Therefore, substituting μ with its value in (25), we have $\Delta(n) = O\left(\frac{1}{(\log n)^{2\gamma-1}}\right)$. ■

VII. SIMULATION RESULTS

In this section we present simulation results for the three scheduling schemes studied in this paper.

The first set of simulations are for $M = 1$, $K = 2$, and $\beta = 2$, i.e., one group requires twice the rate of the second group. We consider the groups to be of equal size. Figure 2 shows the sum of the transmitted rate for WO, TO, and SC as a function of the number of users. As expected, all show a $\log \log n$ growth rate. In fact, the sum of the transmitted rates

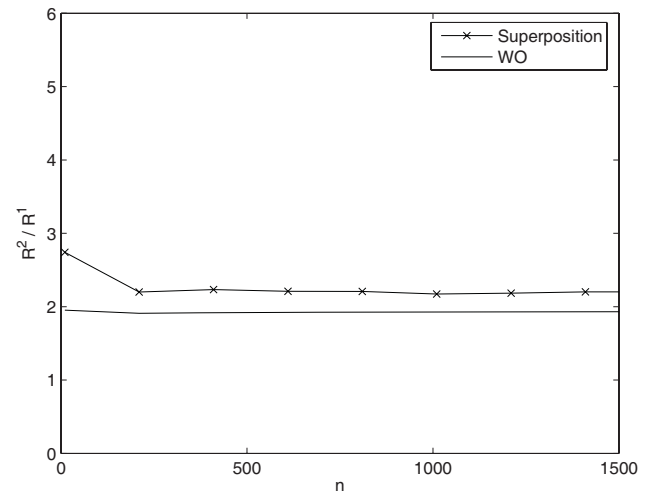


Fig. 3. The ratio of the rates transmitted to the two groups of users as a function of the number of users for WO and SC for a system with $\beta = 2$.

of WO and SC are quite close to the actual sum-rate capacity, signifying that the rate constraints do not lead to much of a rate hit on the throughput. Figure 3 shows the ratio of the rates transmitted to the two groups as a function of the number of users for the WO and SC schedules. As we see the rate of convergence of the SC to the desired ratio is slower than the WO beamforming. TO is not shown as it clearly gives the correct ratio of $\beta = 2$.

Next we consider the performance for multiple antenna broadcast channels. For this set of simulations we consider a broadcast channel with two antennas at the transmitter, $M = 2$. We consider two groups of different size with $\alpha_2 = 2\alpha_1$ and different rate requirements as $\beta_2 = 2\beta_1$. Figure 4 shows the achievable throughput for TO and WO and compares it with the sum rate of opportunistic beamforming with no rate constraints. We have shown the ratio between the rates of users in different groups in Figure 5. As we see WO schedule converges fast to the desired ratio. In Figure 6, we have shown the ratio of the rate for smaller size networks in more detail. As we see, the ratio of the rates converges to the desired ratio even for moderate size networks. For instance,

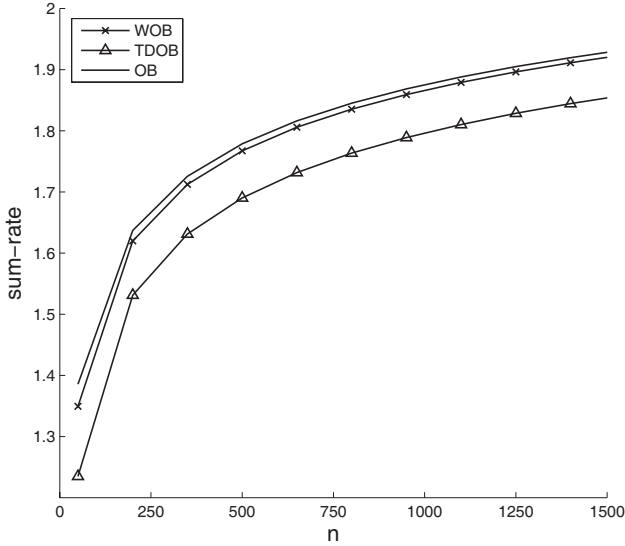


Fig. 4. The sum of the transmitted rates for WO, TO, as well as the opportunistic for a broadcast channel with $M = 2$, $K = 2$, $\frac{\alpha_2}{\alpha_1} = 2$, and $\frac{\beta_1}{\beta_2} = \frac{1}{2}$ as a function of the number of users.

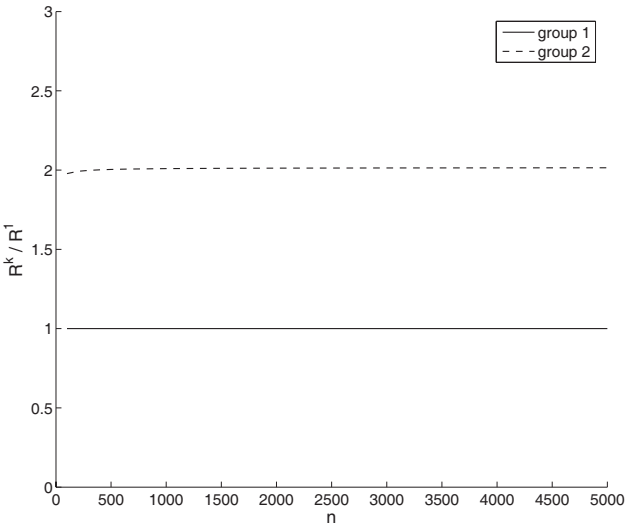


Fig. 5. The ratio of the rates transmitted to the two groups of users as a function of the number of users for WO for a system with $M = 2$, $K = 2$, $\frac{\alpha_2}{\alpha_1} = 2$, and $\frac{\beta_1}{\beta_2} = \frac{1}{2}$.

for a network of size $n = 50$, the ratio of the rates is 1.95 (only 0.05 off from the desired value).

Finally, the last set of simulation is done for a broadcast channel with two antennas at transmitter and three groups of users, i.e., $K = 3$ with equal size. The desired ratio is given as $\beta_1 = 1, \beta_2 = 2, \beta_3 = 3$. In Figure 7, we have plotted the achievable sum-rate for TO and WO scheduling. In Figure 8, we have shown the achieved ratio of rates for different groups.

VIII. CONCLUSION

We considered a homogeneous fading broadcast channel with n users demanding different rates. In our model we assumed users are divided into K groups each one of which demands a different rate and where the ratio of the rates of the groups are given. Users in each group receive an equal

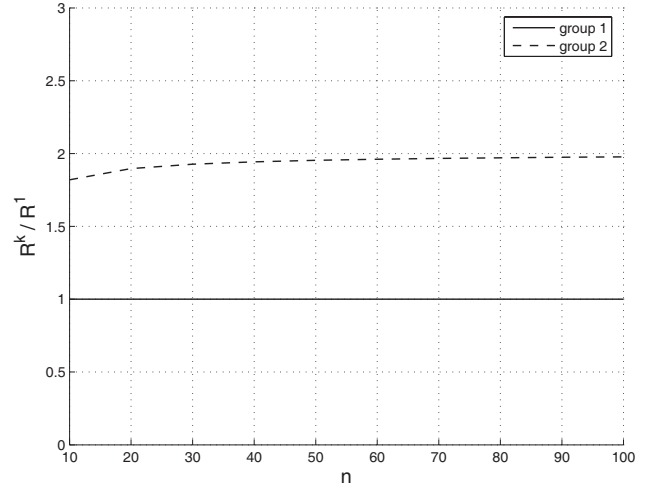


Fig. 6. A closer look at the ratio of the rates transmitted to users in different groups for the example in Figure 5.

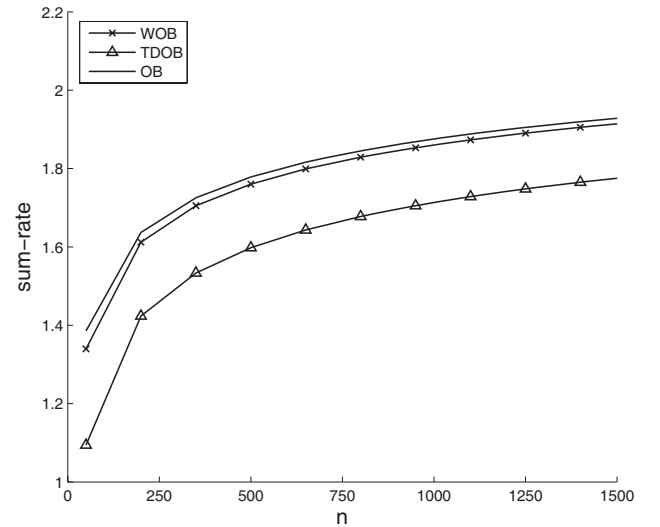


Fig. 7. The sum of the transmitted rates for WO, TO, as well as the opportunistic for a broadcast channel with $M = 2$, $K = 3$, and $\beta_1 = 1, \beta_2 = 2, \beta_3 = 3$ as a function of the number of users.

amount of rate. We considered the problem of scheduling to users in order to maximize the throughput of the system while maintaining the rate constraints. While the problem in its generality can be solved, it requires full CSI in the transmitter and high computational complexity. We therefore focused on two simple schemes that requires very little feedback, namely, time division opportunistic (TO) and weighted opportunistic (WO) beamforming. We further showed that the throughput loss due to these constraints tends to zero for both schemes as the number of users increases, with the performance of WO being significantly superior.

REFERENCES

- [1] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 2–14, 1972.
- [2] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE Inter. Conf. Comm.*, vol. 1, pp. 331–335, 1995.

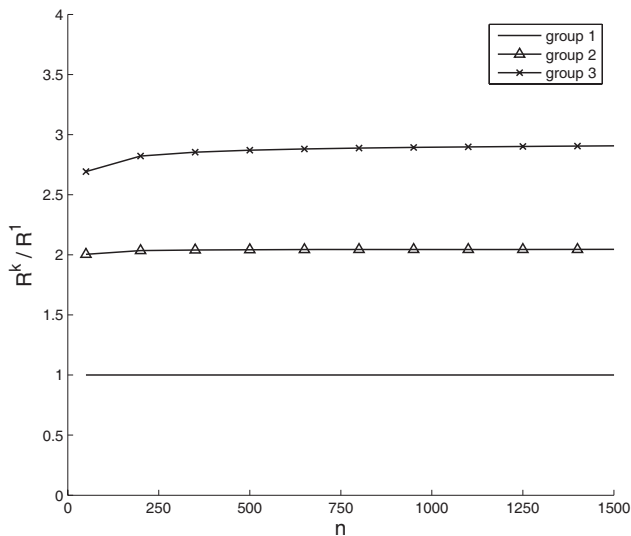


Fig. 8. The ratio of the rates transmitted to the two groups of users as a function of the number of users for WO for a system with $M = 2$, $K = 3$, and $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$.

- [3] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels—I: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2796–2815, 1998.
- [4] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels—I: ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1083–1102, 2001.
- [5] H. Vishwanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known interference cancellation," *IEEE J. Select. Areas. Commun.*, vol. 21, no. 5, pp. 802–811, 2003.
- [6] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [7] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian MIMO broadcast channel," in *Proc. IEEE ISIT*, pp. 174, 2004.
- [8] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, 1983.
- [9] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 768–783, 2004.
- [10] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, 2005.
- [11] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2895–2909, 2003.
- [12] M. Sharif, A. Dana, and B. Hassibi, "Differentiated rate scheduling for Gaussian broadcast channels," in *Proc. IEEE Intern. Symp. on Info. Theory*, 2005.
- [13] P. Bergman, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 197–207, 1973.
- [14] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, 2002.
- [15] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, 2006.
- [16] N. Jindal, "High SNR Analysis of MIMO Broadcast Channels," in *Proc. ISIT 2005*, 2005.
- [17] M. Sharif and B. Hassibi, "Scaling laws of sum rate using time-sharing, DPC, and beamforming for MIMO broadcast channels," in *Proc. ISIT 2004*, pp. 175, 2004.
- [18] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1473–1484, 2002.
- [19] A. Lapidoth, S. Shamai, and M. A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Proc. 43rd Allerton Conf. on Comm., Con. and Comp.*, 2005.

- [20] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.
- [21] H. A. David, *Order Statistics*. New York: Wiley, 1970.
- [22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. London: Academic Press Inc., 1965.
- [23] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge University Press, 2003.



coding for wireless networks, and capacity of broadcast channels.



tems, multi-user information theory, and signal processing techniques for bio-sensors.



Ali Vakili was born in Tehran, Iran, in 1982. He received the B.Sc. degree from Sharif University of Technology, Tehran, Iran, in 2004, and the M.S. degree from the California Institute of Technology, Pasadena, in 2005, both in electrical engineering. He is currently pursuing the Ph.D. degree at the California Institute of Technology. His research interests include MIMO broadcast channels and estimation over communications networks.



Babak Hassibi was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran, Iran, in 1989 and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1993 and 1996, respectively, all in electrical engineering. From October 1996 to October 1998, he was a Research Associate at the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000, he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ. Since January 2001, he has been with the Department of Electrical Engineering at the California Institute of Technology, Pasadena, where he is currently an Associate Professor. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linköping University, Sweden. His research interests include wireless communications, robust estimation and control, adaptive signal processing, and linear algebra. He is the coauthor of the books *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 2000). Dr. Hassibi is a recipient of an Alborz Foundation Fellowship, the 1999 O.Hugo Schuck Best Paper Award of the American Automatic Control Council, the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering, and the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE). He has been a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY Special Issue on Space-Time Transmission, Reception, Coding and Signal Processing.