

Fundamental Limits in MIMO Broadcast Channels

Babak Hassibi and Masoud Sharif

Abstract—This paper studies the fundamental limits of MIMO broadcast channels from a high level, determining the sum-rate capacity of the system as a function of system parameters, such as the number of transmit antennas, the number of users, the number of receive antennas, and the total transmit power. The crucial role of channel state information at the transmitter is emphasized, as well as the emergence of opportunistic transmission schemes. The effects of channel estimation errors, training, and spatial correlation are studied, as well as issues related to fairness, delay and differentiated rate scheduling.

Index Terms—MIMO broadcast channels, sum-rate capacity, asymptotics, channel state information.

I. INTRODUCTION

A COMMUNICATION scenario where a *single* transmitter sends *independent* information through a *shared* medium to *uncoordinated* receivers is referred to as a broadcast channel [1]. A major motivation for the study of such channels is that they provide a model for the downlink of a cellular system or wireless LAN as shown Fig. 1. In an information-theoretic context, the broadcast channel was first formally introduced by Cover in 1972 [1]. Since then it has attracted a great deal of interest in the research community (for an overview of some of the many results see [2]–[4], [6], [7], [16], [17], [20], [28], [38] and the references therein). Information-theoretic results are of interest since they determine the fundamental limits of the achievable rates of the different users in the broadcast channel (e.g., the maximum achievable rates for the users in the downlink of a cellular system). Since the broadcast channel is a multi-user system, its communication limit is given by a *capacity region*, i.e., by the region of all user rates that are simultaneously achievable. To illustrate this, consider the two-receiver broadcast channel depicted in Fig. 2 where the independent messages for both receivers are encoded at the transmitter and each receiver is responsible for decoding its own message. A rate vector (R_1, R_2) is achievable if there exists a coding scheme for which the error probability of both users goes to zero as the block length of the code increases. The capacity region is the union of all the achievable rate vectors as shown in Fig. 3.

An important point on the boundary of the capacity region is the sum-rate point, which corresponds to the maximum of the sum of the rates that can be conveyed by the transmitter to the receivers. In many single-antenna broadcast channels,

Manuscript received May 30, 2006; revised January 15, 2007. This work was supported in parts by the National Science Foundation under grants no. CCR-0133818 and CCR-0326554, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

Babak Hassibi is with California Institute of Technology, Pasadena, CA (e-mail: hassibi@systems.caltech.edu)

Masoud Sharif is with Boston University, Boston, MA (e-mail: sharif@bu.edu).

Digital Object Identifier 10.1109/JSAC.2007.070907.

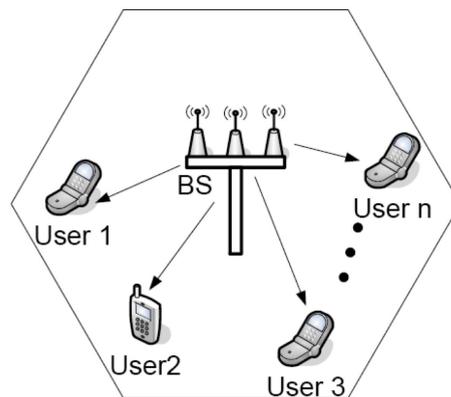


Fig. 1. A multi-antenna broadcast channel, e.g., downlink of a cellular system

the sum-rate point can be achieved by transmitting only to the “strongest” receiver, a scheme that is referred to as “opportunistic transmission” [19]—more on this later.

Since cellular systems and wireless LANs are quite prevalent, there are a variety of different transmission schemes for broadcast channels that are currently in heavy use. These include time-division multiplexing (TDMA), frequency-division-multiplexing (FDMA), and code-division multiplexing (CDMA). One of the surprising results of Cover [1] was that time-division-multiplexing, where at each time slot the transmitter sends information to only one user, is generally not optimal, and one can often do strictly better by a scheme referred to as *superposition coding* [3].

Determining the capacity region for a general broadcast channel is one of the most important open problems in multi-user information theory. Roughly speaking, when the different users in the broadcast channel can be ordered from the strongest to the weakest in a natural way, the broadcast channel is referred to as *degraded* (for example, the Gaussian broadcast channel with single transmit and receive antennas is degraded since the users can be naturally ordered according to their respective SNRs). For degraded broadcast channels the capacity region has been determined and is achieved by the aforementioned superposition coding scheme [3]. In superposition coding, potentially all receivers are transmitted to simultaneously and the different receivers' informations are superimposed on top of one another. The receivers decode the messages from the message of the weakest user to that of the strongest user, with the weakest user decoding only its own message and the strongest user decoding all the messages.

Over the last decade, multiple-antenna point-to-point communication systems have generated a great deal of interest,

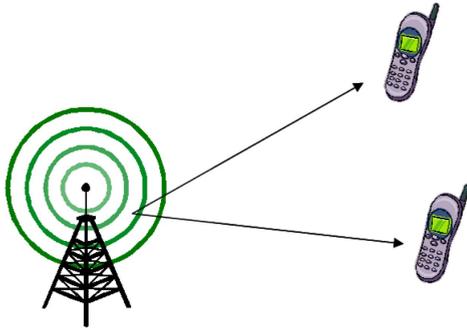


Fig. 2. A two-receiver single antenna broadcast channel

since it has been shown that they are capable of considerably increasing the rate and reliability of a wireless link (see, e.g., [12], [14], [26] and the references therein). Such systems are characterized by a *matrix* channel, where each element of the matrix describes how the signal transmitted from a particular transmit antenna is mapped to the signal received by a corresponding receive antenna. In fact, it has been long known that for such point-to-point systems, if the channel matrix is known to both the transmitter and receiver, then they can jointly diagonalize the channel using unitary operations (essentially through the svd—singular value decomposition—of the channel) thereby creating as many parallel channels as the minimum number of transmit/receive antennas, and thus increase the capacity of the system by the same factor. What was much more surprising was that the same gains could be essentially realized if only the receiver knows the channel matrix [12], [26] and, in fact, if neither the transmitter nor the receiver knows the channel matrix [27], [29] (provided the channel is block-fading).

With all the proven promise of multiple antenna communication systems in point-to-point channels, it is quite natural to ask what role they can play in multi-user communication problems—one of the most important of which is the broadcast channel. In fact, perhaps the first systematic use of multiple transmit antennas for communications was proposed in the context of the broadcast channel [31]. The idea was to use methods from array signal processing whereby a transmitter with multiple antennas could simultaneously transmit multiple beams each carrying independent information for one of the users, thereby increasing the capacity of the system by the number of beams that could be simultaneously transmitted. In any event, there has been a great deal of interest in both the academic and industrial communities on the use of multiple antennas to increase the capacity of cellular systems.

The information-theoretic study of multiple-input/multiple-output (MIMO) broadcast channels, where the transmitter and (possibly) the receivers have multiple antennas, is only very recent. In fact, this study has proven to be quite challenging since the MIMO broadcast channel with Gaussian noise falls under the category of non-degraded broadcast channels¹ for which capacity is not generally known. Determining the capacity region for the MIMO broadcast channel has therefore

¹Roughly speaking, the MIMO broadcast channel is non-degraded since the channels for each user are described by matrices and there is no natural ordering for matrices.

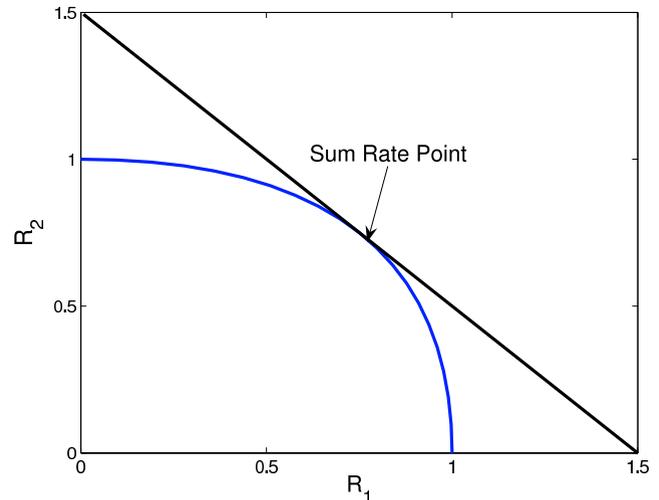


Fig. 3. The capacity region of a two receiver broadcast channel. The sum-rate capacity—often called throughput—corresponds to the point on the boundary that maximizes the sum of the rates to all the receivers.

been one of the major achievements in information theory in recent years [4]–[7], [34]. Interestingly, it turns out that beamforming which has been traditionally used in MIMO broadcast channels is demonstrably sub-optimal. Furthermore, the capacity region is achieved by using a channel coding technique referred to as *dirty paper coding*—a scheme not unrelated to superposition coding that pre-subtracts the interference from other users and which has its own rich history [8].

The goal of this survey paper is not to describe the developments that led to the characterization of the capacity region of the MIMO broadcast channel, nor to describe dirty paper coding, the duality between broadcast channels and multiple access channels (MAC), etc., but rather to use these results to glean some insight into the fundamental limits of performance of MIMO broadcast channels at a rather high level. We hope that this type of analysis will lead to a more clear understanding of the promises and limitations of using multiple antennas in broadcast channels. Some of the performance limits we are interested in are information-theoretic; others are not. In particular, we will be interested in the following questions:

- 1) How does the sum-rate capacity of the MIMO broadcast channel scale with the number of transmit antennas M ?²
- 2) How does the sum-rate capacity scale with respect to other system parameters, such as the total transmit power P , and the number of receive antennas at the users, say N ?
- 3) What is the effect of user diversity? In other words, how does the sum-rate scale as a function of the number of users in the system n ?
- 4) What is the effect of channel state information, in particular at the transmitter, on the above results? How

²We would like at least a linear scaling in M , so that doubling the number of transmit antennas, say, (which is essentially a doubling of the cost and complexity of the base station) leads to a doubling of the system throughput.

much channel state information is necessary to obtain the gains of the system? What is the effect of using training-based schemes?

- 5) What is the effect of channel correlations at the transmitter and how robust is the system compared to idealized models?
- 6) What can one say about more network-theoretic notions of the system such as fairness and delay?
- 7) How best to perform scheduling when users have different rate requests and what is the loss incurred by allowing for differentiated rates³?

II. THE BROADCAST CHANNEL: MODEL AND CAPACITY RESULTS

The standard model for the fading Gaussian noise broadcast channel we will consider in this paper is one where the transmitter is equipped with M antennas and where there are n users, each with N receive antennas.⁴ 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 independent values.⁵ Furthermore, a very reasonable assumption is that, over different users, the fading is assumed to be independent. Although wideband channels are of significant practical interest in current and future standards, in this paper, we only consider narrowband fading channels.

Thus, during any coherence interval, the signal to the i -th user, $i = 1, 2, \dots, n$, can be written as

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

where $x_i(t) \in \mathcal{C}^{N \times 1}$ is the vector of received signals at time t and $H_i \in \mathcal{C}^{N \times M}$ is the channel matrix that is constant during the coherence interval. $w_i(t)$ is spatially and temporally white additive noise whose entries are circularly-symmetric zero-mean unit-variance complex Gaussian random variables $\mathcal{CN}(0, 1)$, $s(t) \in \mathcal{C}^{M \times 1}$ is the transmit symbol satisfying $E\|s(t)\|^2 = 1$ and P is the total transmit power. Throughout the paper, we consider a short-term power constraint implying that the total transmit power per coherence interval is bounded by P . The distribution of the channel matrix H_i depends on the fading environment. In the Rayleigh fading case, which is what we shall consider, it is a zero-mean Gaussian random matrix. We shall also often assume that H_i has iid $\mathcal{CN}(0, 1)$ entries.

We will further make the important assumption that each fading matrix H_i is known to user i and that *all* fading matrices H_i are known to the transmitter. This is referred to as having *full channel state information (CSI)* at the transmitter.

³Giving differentiated rates to users implies operating at non-symmetric boundary points of the capacity region.

⁴It is possible to extend all the results mentioned to the case where each user i has N_i receive antennas. However, for simplicity, we shall not do so here. Moreover, from a practical point of view we shall also most often consider the case $N = 1$, i.e., the users are equipped only with a single receive antenna.

⁵We should remark that, although the assumption of a constant channel for T channel uses is often critical, the requirement that the channels vary independently from one coherence interval to the next is not.

A. Degraded Broadcast Channels and Superposition Coding

When there is only a single transmit antenna ($M = 1$), during each coherence interval the users can be ordered according to their SNRs, $|H_i|^2$, say, $|H_1|^2 \geq |H_2|^2 \geq \dots \geq |H_n|^2$ where $|H_i|$ is the 2-norm of the vector H_i . Thus, the system is degraded and the capacity region can be achieved by superposition coding and interference cancellation [1], [3]. The main idea is to divide the transmit signal into n independent components $\sqrt{P}s(t) = \sum_{i=1}^n \sqrt{P_i}s_i(t)$, such that $\sum_{i=1}^n P_i = P$, and where each $s_i(t)$ is constructed from an independent Gaussian codebook. Assuming no fading and therefore H_i 's are all fixed, we arrange the powers and rates such that the weakest user can only decode its own message, whereas stronger users can decode their own and all weaker messages. This implies that the following set of rates is achievable

$$\begin{aligned} R_1 &< \log(1 + P_1|H_1|^2) \\ R_2 &< \log\left(1 + \frac{P_2|H_2|^2}{1 + P_1|H_2|^2}\right) \\ &\vdots \\ R_n &< \log\left(1 + \frac{P_n|H_n|^2}{1 + \sum_{i=1}^{n-1} P_i|H_n|^2}\right) \end{aligned}$$

Of course, with fading the problem is a bit more complicated since fading makes the ordering random. In this case, using coding over arbitrary large block lengths and adapting to the fading states on a block by block basis, one can obtain the ergodic capacity region averaged over all the fading states [16]. It is possible to give a simple expression for the (ergodic) sum-rate capacity as follows.

$$\begin{aligned} \sum_{i=1}^n R_i \leq C_{sum} &= E \max_{P_i \geq 0, \sum_{i=1}^n P_i = P} \log(1 + P_i|H_i|^2), \\ &= E \log(1 + P \max_{1 \leq i \leq n} |H_i|^2) \end{aligned} \quad (2)$$

where the expectation is over the fading channels H_i . It is thus clear that the scheme that maximizes the sum rate is the one that, during each coherence interval, transmits only to the user with the best SNR (i.e., only one of the P_i 's is nonzero). This is referred to as opportunistic transmission.

B. Non-degraded Broadcast Channels and Dirty Paper Coding

As mentioned earlier, MIMO broadcast channels are not degraded and so superposition coding does not apply. There is, however, a dual coding scheme that can be used. Note that the interference at each receiver, due to the signals intended for the other receivers, is known at the transmitter (since the transmitter is generating all the signals and knows all the channel matrices). Therefore the transmitter can potentially pre-subtract all the interference. The problem is that it may not be possible to do so without violating the power constraint.

In a surprising result in 1983, Costa [8] showed that when the noise and interference are Gaussian, a scheme called *dirty paper coding* achieves a capacity which is the same as if the interference does not exist. In particular, the interference can be pre-subtracted at the transmitter without any average

transmit power penalty. In fact, it is not hard to convince oneself that dirty paper coding provides an alternative to superposition for achieving the rate region of the degraded broadcast channel mentioned earlier.

The idea of pre-canceling the interference at the transmitter can also be applied to non-degraded broadcast channels such as the MIMO Gaussian broadcast channel. In fact, dirty paper coding has been extended to the vector case encountered in MIMO broadcast channels to subtract the multiuser interference without incurring a power penalty. The question is whether the capacity region is achieved with this scheme.

Caire and Shamai [4] showed that dirty paper coding does indeed achieve the sum-rate capacity of the MIMO broadcast channel with two users. Subsequent researchers then proved that dirty paper coding achieves the sum-rate capacity of MIMO Gaussian broadcast channels with any number of users [5] [7] [6]. In particular, it has been shown that the sum-rate capacity of a system with n users and with an average power constraint P , is given by (3) where the P_i are $M \times M$ positive semi-definite matrices that satisfy the power constraint $\sum_{i=1}^n \text{tr} P_i = P$ and the expectation is over the fading channel matrices H_i . Expressions for the capacity region achieved by dirty paper coding can also be given; however, since they are a bit more involved we omit them here and refer the interested reader to [20] for the details. Finally, it was recently shown by Weingarten et al. [34] that dirty paper coding does indeed achieve the entire capacity region, thus bringing closure to this issue.

C. Computational Issues

It has been shown by Jindal et al. [20] that the sum-rate capacity of the broadcast channel is equal to the sum-rate capacity of a so-called multiple access channel (MAC) with the same total average power constraint. More generally, the capacity region of the MIMO broadcast is the dual of that of the multiple access channel. This duality can greatly simplify the power allocation computation for the MIMO broadcast channel as the capacity region of the multiple access channel (MAC) is much easier to compute [41] (it involves simple convex optimization and some elements of matroid theory). For instance, in order to obtain the power allocation corresponding to any point on the boundary of the capacity region of the broadcast channel, we can alternatively solve its dual MAC problem and use a simple linear transformation to relate its solution to that of the broadcast channel (see [20] [49] [25]). In fact, the sum rate capacity formula (3) can be found in this way.

The optimization step in (3) can be numerically performed since it is a convex optimization problem [24]; however, it can become cumbersome if the problem dimension, especially n , gets large. Of course, an explicit solution is out of the question. For this reason, explicit computation of the expectation cannot be done and the expression for C_{sum} needs to be estimated using some form of a Monte Carlo method. Therefore, as it stands, (3) gives little insight into how the throughput scales with the various system parameters, and how it depends on things like channel state information, antenna correlation, etc. This is what we now turn to.

III. SCALING LAWS FOR THE SUM-RATE CAPACITY

In order to get some perspective on scaling laws for the MIMO broadcast channel, it will be useful to begin by reviewing the scaling laws that are obtained for point-to-point MIMO channels.

A. The Point-to-Point Case

It turns out that in a point-to-point link with M transmit antennas and N receive antennas, assuming that the environment has rich scattering, the channel capacity is not very sensitive to channel knowledge at the transmitter and/or receiver. In fact,

- 1) If both the transmitter and receiver know the channel:

$$C = \min(M, N) \log P + O(1), \quad (4)$$

where $O(1)$ represents terms that do not increase with increasing the power P . This result is perhaps not surprising given that, with perfect CSI, the transmitter and receiver can jointly diagonalize the channel (using the singular-value-decomposition) thereby opening up $\min(M, N)$ parallel channels.

- 2) If the receiver only knows the channel [12], [26]:

$$C = \min(M, N) \log P + O(1). \quad (5)$$

This result is more surprising since, even though the channel cannot be explicitly diagonalized, the capacity scales as the capacity of $\min(M, N)$ parallel channels. So-called coherent *space-time codes* are often used to approach capacity in such systems.

- 3) If neither the transmitter nor receiver knows the channel [27], [29]

$$C = \min(M, N) \left(1 - \frac{\min(M, N)}{T} \right) \log P + O(1), \quad (6)$$

where, as before, T is the coherence interval of the channel. Thus, the capacity scales “almost” linearly in $\min(M, N)$. The loss factor $\frac{\min(M, N)}{T}$ can be interpreted as the loss incurred by having to “learn” the channel at the receiver (see e.g., [30]). In such cases, capacity can be approached using a combination of training symbols and coherent space-time codes, or using so-called *non-coherent* space-time codes [13].

B. The Broadcast Case

In point-to-point multi-antenna systems the throughput scaling is equivalent to the *multiplexing gain* defined as $\lim_{P \rightarrow \infty} \frac{C}{\log P}$. However, in the multi-user setting two different throughput scaling laws can be envisioned.

1) *Large Power Regime*: Consider the downlink cellular system of Section II and assume that the channels to all n users are known at the transmitter. Then for fixed M and n , we have

$$\lim_{P \rightarrow \infty} \frac{C_{sum}}{\log P} = \min(M, \max(N, n)), \quad (7)$$

where, as before, C_{sum} refers to the maximum possible sum of the rates to all n users [10] [53] [39] [42] [11].

This is essentially an encouraging result since it says that, with full CSI at the transmitter, the high SNR capacity of the

$$\sum_{i=1}^n R_i \leq C_{sum} = E_{P_i \geq 0, \sum_{i=1}^n \text{tr}(P_i) \leq P} \log \det \left(I_M + \sum_{i=1}^n H_i^* P_i H_i \right) \quad (3)$$

system increases linearly in the number of transmit antennas at the base station (provided $n > M$, which is a very reasonable assumption). Note, further that when $n > M$, the number of receive antennas per user N has no role in the scaling law.

2) *Large Number of Users Regime:* Here now we let M and P be constant and let the number of users n grow. Again, assuming full CSI at the transmitters and receiver, for fixed M and P we have [18]

$$\lim_{n \rightarrow \infty} \frac{C_{sum}}{\log \log n} = M. \quad (8)$$

The multiplexing gain obtained in (8) can be attributed to the *multi-user diversity* of the system. The fact that the scaling with respect to the number of users is doubly-logarithmic (as opposed to logarithmic for the power P) is due to the tail of the Rayleigh distribution. It is worth mentioning that a channel distribution with a different behavior in the tail, such as more practical models with bounded support, would lead to a different scaling law. This multiuser gain will be better understood when we discuss opportunistic transmission below.

3) *Discussion:* The above are clearly two very different regimes. We argue that, from a practical perspective, the latter regime may be more interesting. There are three reasons that come to mind.

- 1) Many practical systems operate with a large number of per-cell users (n could be in the hundreds, whereas M may be no more than two, three or four).
- 2) Significant rates can be obtained even at low to moderate transmit powers P .
- 3) The first gain requires channel knowledge with very high fidelity at the transmitter (indeed a fidelity that grows with the transmit power) [50], whereas—as we shall see below—the latter requires very little CSI (channel state information).

Another interesting fact is that the number of receive antennas N plays very little role in the sum-rate capacity of the downlink of the cellular system (a result which is in stark contrast to the point-to-point case). In fact, we can give a tighter result on the sum-rate, which explicitly involves N , as follows: for fixed P and M

$$C_{sum} = M \log \log n + M \log \frac{P \log N}{M} + o(1), \quad (9)$$

where $o(1)$ represents terms that vanish as n grows [51] (see also [42] [39]).

The above expression shows also the constant (with respect to n) term in the capacity. It correctly identifies the growth rate in P as $M \log P$. However, the capacity grows only doubly logarithmic in N . Thus, as far as the sum-rate is concerned, adding receive antennas at the users is not really beneficial.

Finally, we should note that an expansion similar to (9) which correctly identifies the next highest term of the sum-rate in terms of the power (rather than the number of users) has not yet been obtained. However, some progress towards this end has been reported in [53] (see also [54]).

4) *Capacity Scaling with No CSI:* At the other extreme, we may assume that the transmitter has no channel state information of the channel matrices H_i . Therefore the users are indistinguishable to the transmitter implying that each user should be able to decode all messages. The capacity region of this channel, in its most generality, has yet to be solved [15], [46].

However, when the channels are all identically and independently Rayleigh distributed, then the channel is ergodically degraded and hence a time-division transmission is optimal [15]. Furthermore, it can be shown

- 1) For fixed M and n :

$$\lim_{P \rightarrow \infty} \frac{C_{sum}}{\log P} = \min(M, N). \quad (10)$$

- 2) For fixed P and M :

$$\lim_{n \rightarrow \infty} \frac{C_{sum}}{\log \log n} = 0. \quad (11)$$

Thus, with no CSI at the transmitter, the high SNR performance of the system is essentially that of a point-to-point multi-antenna system. However, the multi-user gain of the system is zero since the transmitter has no knowledge of the user fading channels to exploit them. In fact, the optimal scheme here is to transmit to one user at a time.

In conclusion, unlike the point-to-point case, lack of channel knowledge at the transmitter significantly reduces the sum rate of the system.

IV. PARTIAL CHANNEL STATE INFORMATION

We see a large gap between the achievable rates when full CSI is assumed at the transmitter compared to when no CSI is available. In practice, the assumption of full CSI at the transmitter may be problematic, especially when all or either of M , N and n are large. CSI at the transmitter generally requires a reverse channel for the users to feed back their estimates of the channel matrices H_i after they have estimated them from the transmission of pilot symbols. When the M , N , and, especially, n are large this is a lot of channel information and so the resulting system overhead to achieve CSI at the transmitter may be unacceptably high. This is further exacerbated when the users are highly mobile so that the channel conditions change rapidly and training symbols need to be sent more frequently.

Another issue has to do with the implementation aspects of the optimal dirty paper coding scheme. In its exact form, DPC requires performing vector quantizations of large dimensions at both the transmitter and receiver, a task that may not be computationally feasible in practice. More practical, though suboptimal, implementations have been reported in [35] [36] [37], where implementation aspects of dirty paper coding has been considered. In principle all these schemes require perfect CSI at the transmitter and it is not clear how sensitive they are to the inevitable imperfections of the channel knowledge.

It is also worthwhile mentioning that beamforming has been traditionally used for the downlink scheduling in MIMO broadcast systems as a heuristic method to minimize the multiuser interference in the system. In this scheme the information for the different users are modulated onto different beams where the beams and their corresponding powers are optimized to minimize the interference [47] [44]. Such beamforming schemes also require full CSI at the transmitter, although there is evidence that they may be less sensitive to channel estimation errors. Although sub-optimal, it can be shown that the sum-rate of certain types of beamforming is quite close to that of dirty paper coding in some asymptotic regimes—see also below.

Since lack of CSI at the transmitter results in large rate hits, whereas obtaining perfect CSI may be infeasible, it is important to identify the amount of partial CSI that allows one to achieve most of the promised capacity gains of the MIMO broadcast system. In fact, several such schemes have been proposed, e.g., [52] [40] [45]. In what follows, we shall describe one such scheme.

A. Opportunistic Transmission

It will be worthwhile to first consider the SISO (single-input/single-output) broadcast channel where $M = N = 1$. In this case, the sum-rate capacity formula of (3) specializes to

$$\sum_{i=1}^n R_i \leq C_{sum} = E \max_{p_i \geq 0, \sum_{i=1}^n p_i \leq P} \log \left(1 + \sum_{i=1}^n p_i |h_i|^2 \right), \quad (12)$$

where we have used lower case p_i and h_i to emphasize that the quantities are scalar. It is quite clear from (12) that the sum-rate is achieved when $p_i = P$ for the strongest user (the one for which $|h_i|^2$ is largest) and $p_i = 0$ otherwise. In other words, the sum-rate is achieved by transmitting only to the strongest user in each coherence interval. Such a scheme is referred to as *opportunistic* transmission, since the channel variations among the users are exploited and viewed as opportunities. In this case, it is straightforward to see that

$$\sum_{i=1}^n R_i \leq C_{sum} = E \log \left(1 + P \max_{i=1, \dots, n} |h_i|^2 \right). \quad (13)$$

Exploiting the random channel gains to each user so that the system operates as the “best” of all n channels is often referred to as exploiting *multi-user diversity*. In the case under consideration, where the channels to all the users are iid Rayleigh fading, one can actually quantify the multi-user gain. In this case, the $|h_i|^2$ are all iid exponential random variables with unit mean and simple extremal theory shows that for large n the maximum of n such random variables behaves as $\log n$ with high probability. Thus, for large n , we have

$$C_{sum} = \log P \log n + o(1), \quad (14)$$

which explains the $\log \log n$ nature of the multi-user gain encountered earlier—in effect, transmitting to the strongest user results in a $\log n$ -fold increase in the SNR and therefore a $\log \log n$ -fold increase in the rate.

B. Opportunistic Random Beamforming

One may attempt the above opportunistic transmission scheme for the MIMO broadcast channel. The problem is that if we transmit only to the “best” user we cannot hope to obtain an M -fold increase in the capacity. To do so, one needs to simultaneously transmit to M users. While in the MIMO case it is not clear what “best” means, one possibility is to perform *zero-forcing beamforming* in which the transmitter simultaneously sends information to M users without generating any interference. It is not difficult to show that such a scheme is order optimal, in the sense that it achieves the full multiplexing gain of the MIMO broadcast channel [47] [44].

The drawback is that zero-forcing beamforming also requires perfect CSI. We now briefly describe a simple scheme that achieves most of the broadcast sum-rate in the large n regime, yet requires very little CSI at the transmitter. The idea is based on transmitting M random beams and exploiting multi-user diversity. The details are described in [18] (a similar construction, albeit with little analysis, appears in the appendix of [38]).

Basically, during any coherence interval the transmitter chooses M random orthonormal vectors $\phi_m \in \mathcal{C}^{M \times 1}$ according to an isotropic distribution and then transmits the vector

$$s(t) = \sum_{m=1}^M \phi_m s_m(t), \quad (15)$$

where each $s_m(t)$ is a scalar signal 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

$$\text{SINR}_{m,i} = \frac{|H_i \phi_m|^2}{\frac{M}{P} + \sum_{l \neq m} |H_i \phi_l|^2}. \quad (16)$$

If each user (or, in fact, only those users who have favorable SINRs) 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. In this way, the multi-user diversity of the system can be exploited. (This is the jist of the idea—for more details see [18].)

If one refers to the sum-rate of this scheme by C_{ob} then it can be shown that [51]

$$\lim_{n \rightarrow \infty} (C_{ob} - C_{sum}) = 0. \quad (17)$$

Thus, asymptotically, opportunistic random beamforming has no loss compared to a scheme that has full CSI at the transmitter. The main reason is that when the number of users is large, there exists almost surely a user well-aligned to each beam, yet, with very little interference from other beams, so that the resulting SINR would be quite favorable. However, this fails if the number of users is not too large. One can give a more precise characterization of this, in the sense that the number of antennas should be no more than $O(\log n)$. In other

words,

$$\text{if } \frac{M}{\log n} = c_1 > 0 \quad \text{then} \quad \frac{C_{sum}}{M} = c_2 > 0 \quad (18)$$

where c_1 and c_2 are two positive constants independent of n . Whereas,

$$\text{if } \frac{M}{\log n} = \infty \quad \text{then} \quad \frac{C_{sum}}{M} = 0. \quad (19)$$

On the other hand,

$$\lim_{P \rightarrow \infty} \frac{C_{ob}}{\log P} = 0. \quad (20)$$

In other words, opportunistic random beamforming is highly sub-optimal 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. (In fact, to obtain the multiplexing gain of M at high power requires essentially eliminating the interference, such as is done by a zero-forcing solution [44].)

C. Channel Estimation Error

As mentioned earlier, while dirty paper coding is the optimal transmission scheme for the Gaussian MIMO broadcast channel, it requires perfect knowledge of the channel state information for all the users. In practice, channel estimation errors are inevitable and so it is interesting to study their effect on the achievable rate region. An inner bound on the achievable rate region can be found by treating the channel estimation error as noise [23], [43], [50]. (This is similar to the approach taken for point-to-point MIMO channels in [30].)

We model the channel matrix of the i 'th user by H_i where each entry of H_i is an independent circularly symmetric complex Gaussian random variable with mean zero and variance one. Each user estimates the channel as \hat{H}_i which consists of independent columns, each of which is a circularly symmetric complex Gaussian random vector with covariance \mathcal{A} . Note that we will assume that $H_i - \hat{H}_i$ is uncorrelated with the estimate \hat{H}_i , which means that we are using MMSE (minimum mean-square-error) estimation.

For a large number of users, we can determine a lower bound on the sum-rate with estimation errors as in (21), where $o(1)$ goes to zero as n grows [43]. Eq. (21) suggests that as long as the estimation error covariance is fixed (does not scale with n), one gets the same multi-user multiplexing gain as if there is no estimation error; however, there is a constant rate penalty which is a function of the covariance matrix of the estimation error.

It is perhaps more important to note that if \mathcal{A} is fixed, in the high power regime the achievable sum-rate will not scale with P (see also [50]). In this regime, the performance of opportunistic random beamforming is also sensitive to the errors in the SINR due to channel estimation, channel feedback delay, and quantization errors [23], [45].

D. Training-Based Schemes

The result mentioned so far is based on a given estimation error covariance. To estimate the channel, a training phase is often required in which some portion of the transmission

time and some fraction of the power is devoted to sending training signals. There is a clear trade off between the achievable sum-rate and the duration of the training interval. The longer the training interval, the more accurate the channel estimates and the higher the achievable rate; however, the longer the training interval, the less time we have to transmit the actual data.

One can optimize the parameters in the training-based schemes (i.e., the training interval length and power devoted to training) to optimize the resulting achievable sum-rate. A lower bound on the sum-rate using training can be obtained as [43]

$$C_{sum,train} \geq \min(M, \max(N, n)) \left(1 - \frac{M}{T}\right) \log P + O(1), \quad (22)$$

for large P and where T is the coherence interval of the channel. This result shows that, if the coherence interval is long enough, training-based schemes can achieve almost all of the sum-rate of the MIMO broadcast channel. It should be also mentioned that the result of (22) requires the channel remain unchanged for T transmissions. However, an actual fading channel may change continuously and could result in different asymptotic behaviors in high SNR [56].

V. SPATIAL CHANNEL CORRELATION

So far we have assumed that the entries of the channel matrix are iid. In practice, due to local scatterers around the transmitter and/or user, if the antennas are not spaced far enough, the entries of the channel matrix will be correlated. In what follows, for simplicity, we shall assume that $N = 1$ and so will consider only correlation at the transmitter antennas.

Since the local scatterers at the transmitter cause the correlation, it is very reasonable to assume that the channel vectors H_i all have the same nonsingular covariance matrix \mathcal{R} such that $\text{tr}(\mathcal{R}) = M$. In this case, the sum-rate can be shown to be [48]

$$C_{sum} = M \log \log n + M \log \frac{P}{M} + M \log \sqrt[M]{\det \mathcal{R}} + o(1), \quad (23)$$

for large n . Comparing (9) with (23), we observe that spatial correlation does degrade the sum-rate capacity. The sum-rate loss is the logarithm of the Geometric mean of the eigenvalues of the covariance matrix.

In the presence of spatial fading several types of random opportunistic beamforming are possible. One is to use channel pre-whitening followed by opportunistic random beamforming. In this case, instead of using the $M \times M$ random unitary beamforming matrix $\Phi = [\phi_1 \dots \phi_M]$, we use $\sqrt{\alpha} \mathcal{R}^{-1/2} \Phi$ where α is a constant to make sure that the average transmit power constraint is satisfied and $\mathcal{R}^{-1/2}$ is the inverse of the square-root of \mathcal{R} . It is then quite straightforward to show that the sum-rate of this scheme scales as,

$$E(C_{ob-w}) = M \log \log n + M \log \frac{P}{M} - M \log \frac{\text{tr}(\mathcal{R}^{-1})}{M} + o(1), \quad (24)$$

for large n . This implies that the sum-rate loss is the logarithm of harmonic mean of the eigenvalues of the covariance matrix \mathcal{R} . Thus, this scheme is suboptimal.

$$C_{sum,err} \geq M \log \log nN + M \log \frac{P}{M} + \log \det(I - \mathcal{A}) - \log \det(I + P\mathcal{A}) + o(1) \quad (21)$$

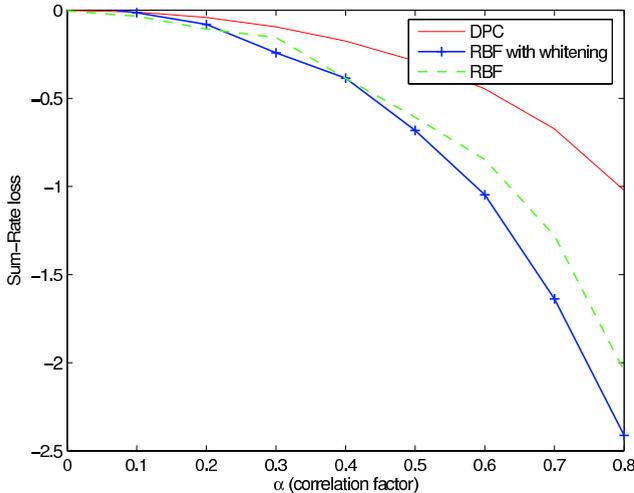


Fig. 4. Sum-rate loss versus the correlation factor for a system using random beamforming (RBF) with two transmit antennas and $n = 100$.

A generalization of channel pre-whitening is to use a general precoding matrix K instead of $\mathcal{R}^{-1/2}$ prior to beamforming. Judicious choice of K generally outperforms channel pre-whitening. The scaling law of the sum-rate is complicated and given in [48].

Fig. 4 compares the sum-rate loss of different schemes as a function of the correlation strength of the channel. We consider a broadcast channel with two transmit antennas, i.e., $M = 2$, and $n = 100$ users with spatial correlation matrix of

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \quad (25)$$

where α is the correlation factor. Fig. 5 also shows the sum-rate versus the number of users in a system with $\alpha = 0.5$ and $P = 10$ and compares it with sum-rate without correlation. It is clear that the sum-rate loss at $\alpha = 0.5$ is about 5 percent.

Determining the sum-rate loss in the presence of spatial correlation in the high power regime is an interesting open issue.

VI. FAIRNESS AND DELAY

The sum rate capacities computed in the previous sections were often obtained by opportunistic transmission, i.e., the transmitter transmits to the user(s) with the best channel condition(s). While this can be optimal for throughput, it will inevitably lead to unfairness in the system as users with poor channel conditions may incur large delays. See for example [32], [38], [45] for the analysis of fairness in broadcast channels

A. Long Term Fairness

The issue is much more pronounced when the system is heterogenous, in the sense that the users have different average

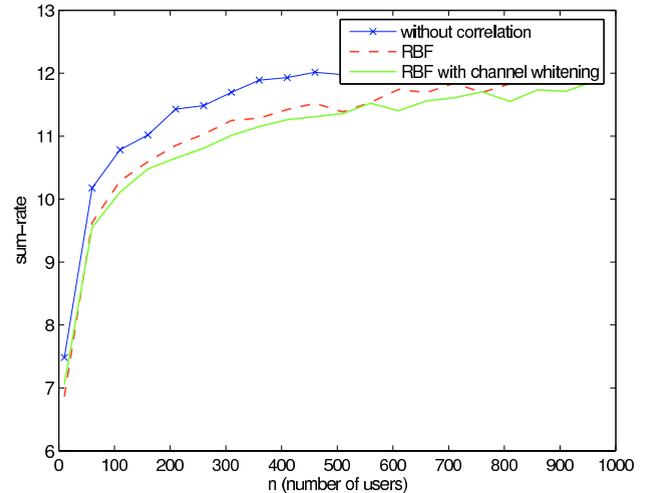


Fig. 5. Sum-rate versus the number of users in a system using random beamforming (RBF) with $M = 2$ and $\alpha = 0.5$

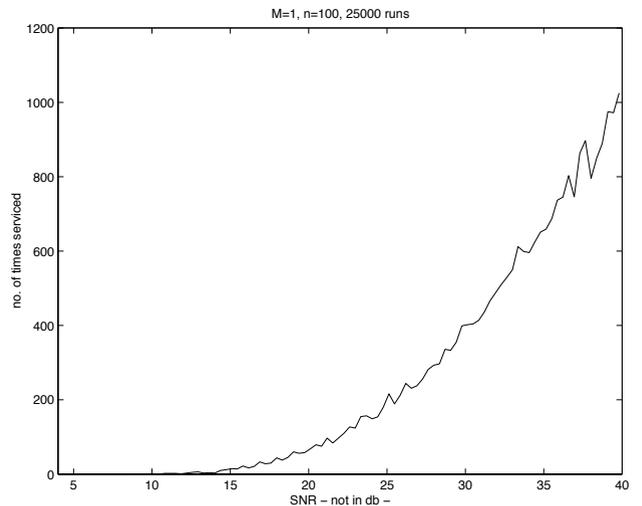


Fig. 6. $M = N = 1$, $n = 100$, SNR 4-40 (6-16db), 25000 runs of opportunistic transmission.

SNRs⁶ as a result of some users being closer to the base station and some users being further away. In fact, this is what is most often encountered in practice. It is very easy to see that in this case opportunistic transmission (though throughput optimal) can be quite unfair. To this end, consider the setting of Fig. 6 which represents a broadcast channel with $M = N = 1$ and $n = 100$ users whose SNRs vary from 4 to 40 (equivalently, 6 to 16 db). As can be seen after 25,000 coherence intervals the opportunistic transmission transmits mostly to the stronger users and very rarely to the weaker ones.

The situation, however, can be quite different in the MIMO case. Here if we use opportunistic random beamforming it can

⁶Averaging here is done over all channel realizations.

$$P(\text{choosing user with } \text{SNR}_{\min}) \geq \frac{M}{n} e^{-\left(\frac{1}{\text{SNR}_{\min}} - \frac{1}{\text{SNR}_{\max}}\right)} \left(e^{\frac{\log n}{M}} - 1\right) \quad (26)$$

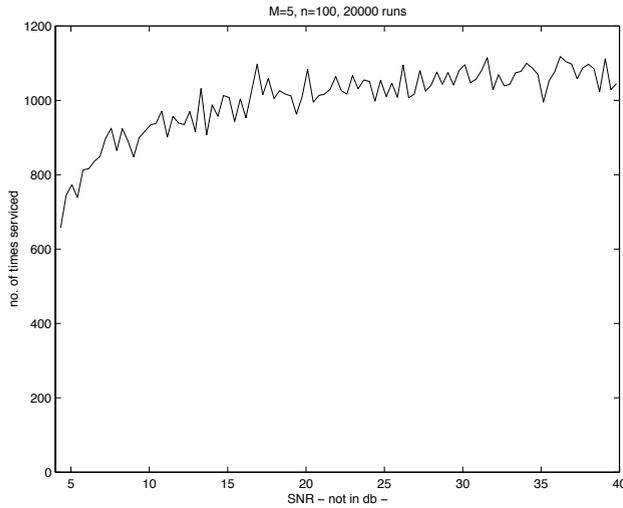


Fig. 7. $M = 5$, $n = 100$, SNR 4-40 (6-16db), 20,000 runs.

be shown that the probability of transmitting to the user with the “worst” channel condition is as in (26), where SNR_{\max} and SNR_{\min} are the maximum and minimum SNRs in the system [55]. Clearly, if $M \gg \log n$, the above probability will approach $\frac{M}{n}$ which means that the system will be fair. The problem, in view of (19) is that the throughput now will be small.

The above analysis shows that a very reasonable operating point for the opportunistic random beamforming that yields high rates and fair transmission is to have $M = \alpha \log n$ transmit antennas, for some fixed and reasonably-sized $\alpha \approx 1$. Thus, consider Fig. 7, which is the same setting as Fig. 6 except that we have $M = 5 \approx \log 100$ transmit antennas. As can be seen, the system is remarkably fair, with the different users being transmitted to with almost equal probability. The total sum-rate of the system is almost 5 times that of the SISO system of Fig. 6.

B. Short Term Fairness and Delay

For homogenous systems (where the distributions of the channels of all the users are identical) opportunistic transmission is *long-term fair*, meaning that eventually all users will be transmitted to equally. However, it may not be *short-term fair*, meaning that it may take a long time until all users are transmitted to equally and the system behaves fairly. This short-term fairness is essentially related to the question of delay.

To formalize the idea and to obtain some explicit results, assume that transmission over each coherence interval is packetized (i.e., consists of a single packet). Thus, during any coherence interval only M packets are transmitted to M users (one packet per user). Further assume that all the channels change to independent values after each coherence interval. We will define the delay of the system as $D_{m,n}$,

the number of channel uses it takes until all n users have successfully received m packets. Clearly, one delay optimal scheme would be round-robin scheduling which clearly incurs a delay of $D_{m,n} = \frac{mn}{M}$. Now the throughput optimal scheme will inevitably incur a larger delay and the following results characterize the resulting delay penalty [55].

- 1) For n fixed and $m \rightarrow \infty$,

$$E(D_{m,n}) = \frac{mn}{M} + O(n \log m), \quad (27)$$

which clearly states that the system is long-term fair, since the round-robin delay is $\frac{mn}{M}$.

- 2) For m fixed and $n \rightarrow \infty$, we have

$$E(D_{m,n}) = \frac{n \log n}{M} + O(n \log \log n). \quad (28)$$

This result shows that when $m = 1$ (we are interested in transmitting a single packet to each user) the delay penalty over the round-robin schemes is only a factor of $\log n$. In fact, this is the worst possible delay penalty.

- 3) For $m = \log n$ and $n \rightarrow \infty$, we have

$$E(D_{m,n}) = 3.126 \frac{mn}{M} + O(n \log \log n). \quad (29)$$

Thus, if m grows as the logarithm of the number of users the delay penalty is roughly a factor of 3.

- 4) For $m = (\log n)^r$ where $r > 1$ is fixed and $n \rightarrow \infty$,

$$E(D_{m,n}) = \frac{mn}{M} + O(n \log n), \quad (30)$$

which demonstrates how long it takes for the system to become fair. Thus, after roughly $n \log n$ transmissions the system behaves fairly.

The above results show that multiple antennas do not significantly reduce the expected delay over a system that operates M times as fast. However, multiple transmit antennas can significantly improve the long-term fairness in a heterogeneous network. For example, in the presence of channel temporal correlations, multiple antennas can significantly reduce the delay by decorrelating in time the effective channel through means such as random beamforming [38].

In summary, the opportunistic scheduling lead to a $\log n$ fold increase in “the worst-case” delay compared to the round-robin type transmission. One can improve this loss by a simple modification of the transmission, namely instead of sending the packet to the best user, the transmitter considers the d best users and transmits to the one that has received the least number of packets. This scheme still exploits the multiuser diversity and so loses very little in terms of rate performance; however it can potentially improve the expected delay by a factor of $\frac{1}{d}$ for large number of users. Fig. 8 compares the expected delay in sending one packet to all n users with the d algorithm.

For a more comprehensive discussion of the implications of these results see [55]. In general, a thorough understanding of the throughput delay trade-offs in MIMO broadcast channels,

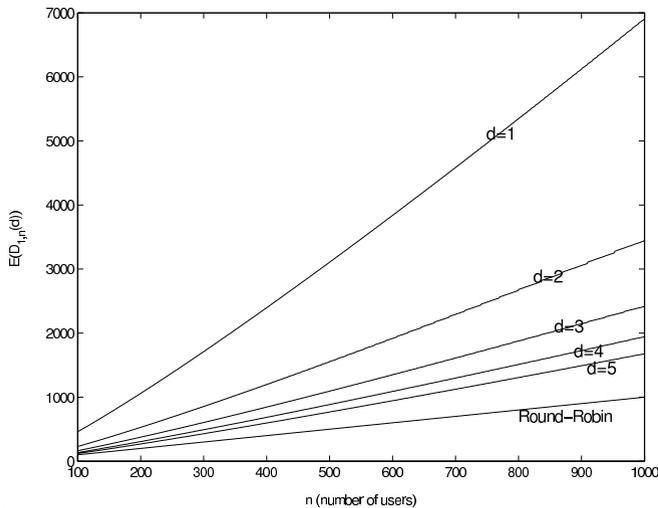


Fig. 8. Sum-rate versus the number of users in a system with $M = 2$ and $\alpha = 0.5$

especially when we have random arrival rates for packets at the transmitter, as well as temporal correlation of the channels is a very important open problem.

VII. DIFFERENTIATED RATE SCHEDULING

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., [22] 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 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, an important goal is to develop *simple* schemes, that require very *little* CSI, give *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.

A. Basic Problem

We are interested in giving different rates to the different users. Thus, assume that the n users are divided into K groups, each with $\alpha_k n$ users ($\sum_{k=1}^K \alpha_k = 1$) and each require a different rate. In particular, $R^k/R^K = \beta_k$, $k = 1, \dots, K-1$, where R^k is the rate required for group k and β_k represents the rational rate requirements. The problem we are interested in is solving

$$\max \sum_{i=1}^n R_i$$

subject to $\frac{R^k}{R^K} = \beta_k$, $k = 1, \dots, K-1$

where R_i denotes the rate of the i 'th user for $i = 1, \dots, n$ and R^k corresponds the rate of any user in group k for $k = 1, \dots, K$.

B. Time-Division Opportunistic (TO) Beamforming

Assume 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. It is not hard to convince oneself that to satisfy the rational rate constraints, we must have

$$\frac{t_k}{T} = \frac{\alpha_k \beta_k}{\sum_{l=1}^K \alpha_l \beta_l}, \quad k = 1, \dots, K \quad (31)$$

We now have the following result [51]

$$C_{tdob} = C_{sum} + \Theta\left(\frac{1}{\log n}\right), \quad (32)$$

where C_{tdob} represents the sum-rate for the time-division opportunistic scheme⁷. In particular, C_{tdob} approaches C_{sum} as n grows, indicating that there is little loss compared to the sum rate point even though we are providing the users with different rates.

C. Weighted Opportunistic (WO) Beamforming

Here we weigh the SINR of each user according to its group by a weight 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.

In the WO beamforming scheme there are two questions to be answered. First, how to determine the weights 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 first question is answered by setting [51]

$$\mu_k = 1 + \frac{\log \beta_k}{\log n - (M-1) \log \log n}. \quad (33)$$

and the second by

$$C_{wob} = C_{sum} + O\left(\frac{1}{\log n}\right), \quad (34)$$

where C_{wob} is the sum-rate for the weighted opportunistic scheme. Note again that there is no asymptotic loss compared to the sum-rate capacity. However, the convergence rate is faster than the TO scheme and the performance is generally better.

To gain some insight into the performance of the schemes described, we present a simple simulation result in this section. We consider the case of $K = 2$ groups of equal size and require that $\frac{R^1}{R^2} = \beta_1 = 2$. Finally, we assume $M = 2$ transmit antennas at the base station, $\frac{P}{M} = 1$ (so that the system operates at 0 db) and vary the number of users from $n = 50$ to $n = 5000$. Figure 9(a) shows the ratio of the rates of

⁷Here $g(x) = \Theta(f(x))$ implies that $\alpha_2 \leq \lim_{x \rightarrow \infty} \left| \frac{g(x)}{f(x)} \right| \leq \alpha_1$ where α_1 and α_2 are two constants independent of n . Similarly $g(x) = O(f(x))$ implies that $\lim_{x \rightarrow \infty} \left| \frac{g(x)}{f(x)} \right| \leq \alpha_1$.

two users in the two different groups when WO beamforming is used with μ_1 and μ_2 as in (33). As n increases the ratio converges to the desired value. In Figure 9(b) the sum rate of the WO and TO schemes are plotted and compared to that of the unconstrained opportunistic scheme in which the users are not divided into groups. For reference, we also plot $M \log \log n + M \log \frac{P}{M} = M \log \log n$ (since $\frac{P}{M} = 1$). The throughputs all converge to $M \log \log n + M \log \frac{P}{M}$, though the convergence rate is quite slow. Finally, we remark that the WO scheme outperforms the TO scheme, and has negligible performance loss compared to the unconstrained opportunistic scheme.

VIII. CONCLUSION

In this paper we studied some of the fundamental limits of Gaussian MIMO broadcast channels. In particular, we focused on the high power regime and on the many user regime and generally observed a throughput that linearly increases with the number of transmit antennas. We also studied the behavior of the throughput as a function of other system parameters such as the number of users, the number of receive antennas per user, and the total transmit power. We also studied the effect of channel state information on the results and mentioned some promising schemes that require only minimal CSI and are based on opportunistic transmission. We looked at issues of channel estimation errors, spatial correlation, and touched upon issues of fairness, delay and differentiated rate scheduling. Although all this was done at a high level, the hope is that it will give some insight into the capabilities and limitations of Gaussian MIMO broadcast channels.

REFERENCES

- [1] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.
- [2] T. Cover, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 9, pp. 2524–2530, September 1998.
- [3] P. Bergman, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 197–207, Mar. 1973.
- [4] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
- [5] P. Viswanath and D. N. Tse, "Sum capacity of the vector Gaussian broadcast channel and downlink-uplink duality," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.
- [6] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channel," *IEEE Trans. Inform.*, vol. 49, no. 10, pp. 2658–2668, Oct. 2002.
- [7] W. Yu and John Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 29, pp.439-441, May 1983.
- [8] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp.439-441, May 1983.
- [9] 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, June 2003.
- [10] B. Hochwald and S. Viswanath, "Space-time multiple access: linear growth in the sum rate," in *Proc. of the 40th Annual Allerton Conf.*, 2002.
- [11] H. Vishwanathan and S. Venkatesan, "Asymptotics of sum rate for dirty paper coding and beamforming in multiple antenna broadcast channels," in *Proc. of the 41st Annual Allerton Conf.*, 2003.
- [12] E. Telatar, "Capacity of multi-antenna Gaussian channel," *European Trans. Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.
- [13] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, January 1991.
- [14] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas. Commu.*, vol. 21, no. 5, pp. 684–702, June 2003.
- [15] A. Amraoui, G. Kramer, and S. Shamai, "Coding for the MIMO broadcast channel," in *Proc. IEEE ISIT*, 2003.
- [16] 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.
- [17] H. Viswanathan, S. Venkatesa, and H. Huang, "Downlink capacity evaluation of cellular networks with known interference cancellation," *IEEE J. Select. Areas in Comm.*, vol. 21, no. 5, pp. 802–811, 2003.
- [18] 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.
- [19] 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.
- [20] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels" *IEEE Trans. Info. Theory*, vol. 50, no. 5, pp. 768–783, 2004.
- [21] N. Jindal and A. Goldsmith, "Dirty paper coding vs. TDMA for MIMO broadcast channel," to appear in *Proc. IEEE Inter. Conf. Commu.*, June 2004.
- [22] 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.
- [23] N. Jindal and A. Goldsmith, "Finite rate feedback MIMO broadcast channels with a large number of users," in *Proc. IEEE ISIT*, 2006.
- [24] S.P. Boyd and L. Vandenberghe, "Convex Optimization", *Cambridge University Press*, 2004.
- [25] N. Jindal, S. Jafar, S. Vishwanath, and A. Goldsmith, "Sum power waterfilling for Gaussian broadcast channels," in *Proc. 36th Asilomar Conf. on Sig. and Syst.*, 2002.
- [26] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998.
- [27] L. Zheng and D. N. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna Channels," *IEEE Trans. Inform. Theory*, vol. 49, May 2003, pp. 1073–96.
- [28] B. Hochwald, T. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.
- [29] B. Hassibi and T.L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received density in closed form," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1473–1484, June. 2002.
- [30] B. Hassibi and B.M. Hochwald, "How much training is needed in a multiple antennas wireless link?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.
- [31] A.J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception", US Patent 5,345,999, 1994.
- [32] M. Andrews, K. Kumaran, K. Ramanan, and A. Stoylar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 30, no. 2, pp. 246-251, February 2001.
- [33] E. Yeh and A. S. Cohen, "Throughput and delay optimal resource allocation in multiaccess fading channels," in *Proc. IEEE ISIT*, 2003, pp. 245–245.
- [34] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian MIMO broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.
- [35] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE Glob. Comm. Conf.*, 2001.
- [36] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1277, June 2002.
- [37] C. B. Peel, B. Hochwald, and A. L. Swindlehurst, "A vector perturbation technique for near capacity multi-antenna multi-user communication-Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
- [38] 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.
- [39] M. Sharif and B. Hassibi, "Scaling laws of sum rate using time-sharing, DPC, and beamforming for MIMO broadcast channels," in *Proc. Inter. Symp. on Information Theory*, 2004.
- [40] J. Chul Roh and B. Rao, "Multiple antenna channels with partial feedback," in *Proc. of IEEE ICC*, 2003.

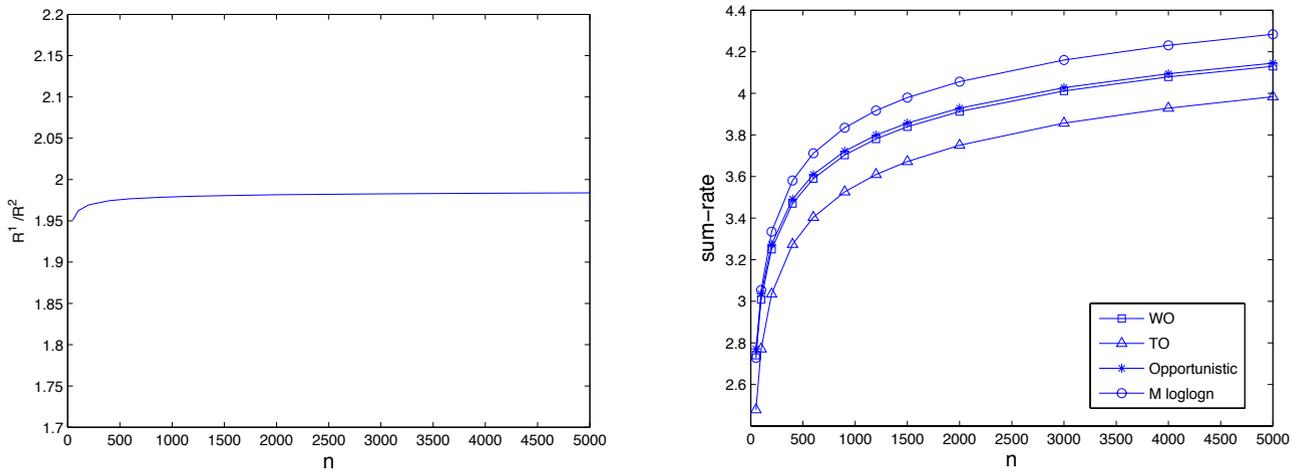


Fig. 9. (a) ratio of the rates and (b) the sum-rate of WO and TO schemes for $M = 2$, $K = 2$, $\alpha_1 = \alpha_2 = \frac{1}{2}$ and $\beta_1 = 2$

- [41] D. N. 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, Nov. 1998.
- [42] Y. Xie and C. Georghiades, "Some results on the sum rate capacity of MIMO fading broadcast channel," in *Proc. Inter. Symp. in Advances in Wireless Comm. 2002*.
- [43] A. F. Dana, M. Sharif, and B. Hassibi, "On the capacity region of multi-antenna Gaussian Broadcast channels with estimation error," in *Proc. IEEE ISIT 2006*, Seattle, WA.
- [44] T. Yoo and A. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *Proc. IEEE ICC 2005*, vol. 1, May 2005, pp. 542–546.
- [45] M. Kobayashi, and G. Caire, "Joint beamforming and scheduling for a MIMO downlink with random arrivals," in *Proc. IEEE ISIT 2006*, Seattle, WA.
- [46] D. Tuninetti and S. Shamai, "Gaussian broadcast channels with state information at the receivers," *DIMACS Center at Rutgers University*, Piscataway, NJ, March 2003.
- [47] M. Schubert and H. Boche, "Solution of multiuser downlink beamforming problem with individual SINR constraint," *IEEE Trans. Veh. Technol.*, vol. 53, January 2004.
- [48] T. Al-Naffouri, M. Sharif, and B. Hassibi, "How much does transmit correlation affect the sum-rate of MIMO downlink channels" *submitted to IEEE ISIT*, 2006.
- [49] M. Sharif, A. F. Dana, and B. Hassibi, "Differentiated rate scheduling for the downlink of cellular systems," in *Proc. IEEE International Symposium on Information Theory*, September 4–9, 2005, Adelaide, Australia.
- [50] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of MIMO fading broadcast channel with imperfect transmitter side information," in *Proc. of Allerton Conference on Comm., Control, and Computation*, September 2005.
- [51] A. Vakili, A.F. Dana, M. Sharif and B. Hassibi, "Differentiated Rate Scheduling for MIMO Broadcast Channels", in *Proc. Allerton Conference on Comm., Control, and Computation*, September 2005.
- [52] M. Kountouris and D. Gesbert, "Memory based opportunistic multi-user beamforming," in *Proc. IEEE International Symposium on Information Theory*, September 4–9, 2005, Adelaide, Australia.
- [53] N. Jindal "High SNR analysis of MIMO Broadcast channels," in *Proc. IEEE International Symposium on Information Theory*, September 4–9, 2005, Adelaide, Australia.
- [54] A. Lozano, A. Tulino, and S. Verdu, "High SNR power offset in multi-antenna communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, December 2005.
- [55] M. Sharif and B. Hassibi, "A delay analysis for opportunistic transmission in fading broadcast channels," in *Proc. IEEE INFOCOM*, March 2005.
- [56] A. Lapidoth and S. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 437–452, February 2006.



Babak Hassibi was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran in 1989, and the M.S. and Ph.D. degrees from Stanford University 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, CA., 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). He 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), and was a participant in the 2004 National Academy of Engineering "Frontiers in Engineering" program.

He has been a Guest Editor for the IEEE Transactions on Information Theory special issue on "space-time transmission, reception, coding and signal processing" was an Associate Editor for Communications of the IEEE Transactions on Information Theory during 2004–2006, and is currently an Editor for the Journal "Foundations and Trends in Information and Communication".



Masoud Sharif was born in 1977. He received the BSc (with honors) and Msc degrees from the Sharif University of Technology, Tehran, Iran, and the PhD degree from the California Institute of Technology, Pasadena, CA, in 1999, 2001, and 2005, respectively, all in electrical engineering. In 2005, he was a post-doctoral scholar at Caltech, and since 2006 he has been an assistant professor in the department of electrical and computer engineering at Boston University.

His research interests are in the areas of information theory, wireless communications, and signal processing. Among others, he has worked on the capacity of MIMO broadcast channels, peak-to-average power ratio of multicarrier signals, and data collection in sensory networks. He was the recipient of the 2006 Charles H. Wilts Prize for the best PhD thesis in electrical engineering at Caltech.