Sum Capacity of Multiuser MIMO Broadcast Channels with Block Diagonalization

Zukang Shen, Runhua Chen, Jeffrey G. Andrews, Robert W. Heath, Jr., and Brian L. Evans

The University of Texas at Austin, Austin, Texas 78712

Email: {shen, rhchen, jandrews, rheath, bevans}@ece.utexas.edu

Abstract— The sum capacity of a Gaussian broadcast MIMO channel can be achieved with Dirty Paper Coding (DPC). Deploying DPC in real-time systems is, however, impractical. Block Diagonalization (BD) is an alternative precoding technique for downlink multiuser MIMO systems, which can eliminate inter-user interference at each receiver, at the expense of suboptimal sum capacity vs. DPC. In this paper, we study the sum capacity loss of BD for a fixed channel. We show that 1) if the user channels are orthogonal to each other, then BD achieves the complete sum capacity; and 2) if the user channels lie in a common row vector space, then the gain of DPC over BD can be bounded by the minimum of the number of transmit and receive antennas and the number of users. We also compare the ergodic sum capacity of DPC with that of BD in a Rayleigh fading channel. Simulations show that BD can achieve a significant part of the total throughput of DPC. An upper bound on the ergodic sum capacity gain of DPC over BD is derived, which can be evaluated with a few numerical integrations. With this bound, we can easily estimate how far away BD is from being optimal in terms of ergodic sum capacity, which is useful in directing practical system designs.

I. INTRODUCTION

Multiple-input-multiple-output (MIMO) systems can significantly increase the spectral efficiency by exploiting the spatial degree of freedom created by multiple antennas. It has been shown that the point-to-point MIMO channel capacity scales linearly with the minimum number of transmit and receive antennas in Rayleigh fading channels [1]. For Gaussian MIMO broadcast channels (BC), it was conjectured in [2][3] and recently proven in [4] that Dirty Paper Coding (DPC) [5] can achieve the capacity region. A duality relationship for the MIMO broadcast DPC capacity region to the MIMO multiple access channel (MAC) capacity region is shown in [3][6].

The sum capacity in a multiuser MIMO broadcast channel is defined as the maximum aggregation of all the users' data rates. Although the sum capacity of a Gaussian MIMO BC channel can be achieved with DPC [3], deploying DPC in real systems is very complicated and impractical. An alternative low-complexity precoding technique is Block Diagonalization (BD) [10]–[13], which is an extension of the zero-forcing precoding technique for downlink multiuser MIMO systems. With BD, each user's precoding matrix lies in the null space of all other users' channels. Hence if the channel matrices of all users are perfectly known at the transmitter, then there is no interference at every receiver, rendering a simple receiver structure. Furthermore, the transmitter's complexity is much lower for BD than DPC. On the other hand, BD is inferior

in terms of sum capacity to DPC, since the users' signal covariance matrices are generally not optimal for sum capacity.

The sum capacity gain of DPC vs. TDMA has been studied in [15]. For a fixed channel, it has been proven that the gain of DPC over TDMA is bounded by the minimum of the number of users and the number of transmit antennas, for different user number, antenna setting and SNR. Furthermore, it has been shown in [16] that the ergodic sum capacity of BD scales the same as DPC in the number of users for Rayleigh fading channels. In this paper, we focus on the sum capacity gain of DPC over BD. We define BD's sum capacity to be the maximum total throughput over all possible user sets. Hence the TDMA sum capacity is automatically incorporated in BD's sum capacity definition. Therefore, the general bound on the gain of DPC vs. TDMA applies to the gain of DPC vs. BD. We show that for a fixed channel 1) if user channels are orthogonal to each other, then BD achieves the same sum capacity as DPC; 2) if user channels lie in the same subspace, then the gain of DPC over BD can be reduced to the minimum of the number of transmit and receive antennas and the number of users. Furthermore, the ergodic sum capacity of DPC is compared to that of BD in a Rayleigh fading channel. We show that BD achieves a significant part of the DPC sum capacity for low and high SNR regimes, or when the number of transmit antennas is much larger than the sum of all users' receive antennas. An upper bound on the ergodic sum capacity gain of DPC over BD is derived. The proposed upper bound on the gain can be evaluated with a few numerical integrations, hence providing an easy way to compare the performance of BD vs. DPC without performing the time-consuming Monte Carlo simulations.

II. SYSTEM MODEL AND BACKGROUND

In a K-user downlink multiuser MIMO system, we denote the number of transmit antennas at the base station as N_t and the number of receive antennas for the jth user as $N_{r,j}$. The transmitted symbol of user j is denoted as a $N_j(\leq N_{r,j})$-dimensional vector $\mathbf{x}_j$, which is multiplied by a $N_t \times N_j$ precoding matrix $\mathbf{T}_j$ and sent to the basestation antenna array.

The received signal $\mathbf{y}_j$ for user j can be represented as

$$\mathbf{y}_j = \mathbf{H}_j \mathbf{T}_j \mathbf{x}_j + \sum_{k=1, k \neq j}^{K} \mathbf{H}_j \mathbf{T}_k \mathbf{x}_k + \mathbf{v}_j \qquad (1)$$

where $\mathbf{v}_j$ denotes the Additive Gaussian White Noise (AWGN)

vector for user j with variance $E[\mathbf{v}_j\mathbf{v}_j^*] = \sigma^2\mathbf{I}$, where $()^*$ denotes the matrix conjugate transpose. Matrix $\mathbf{H}_j \in \mathbb{C}^{N_{r,j} \times N_t}$ denotes the channel transfer matrix from the basestation to the jth user, with each entry following an i.i.d. complex Gaussian distribution $\mathcal{CN}(0,1)$. For analytical simplicity, we assume that $\mathrm{rank}(\mathbf{H}_j) = \min(N_{r,j}, N_t)$ for all users. It is also assumed that the channels $\mathbf{H}_j$ experienced by different users are independent. The key idea of block diagonalization is to precode each user's data $\mathbf{x}_j$ with the precoding matrix $\mathbf{T}_j \in \mathbb{U}(N_t, N_j)$, such that

$$\mathbf{H}_i\mathbf{T}_j = 0 \quad \text{for all } i \neq j \text{ and } 1 \leq i, j \leq K, \quad (2)$$

where $\mathbb{U}(n, k)$ represents the class of $n \times k$ unitary matrices, i.e. the collection of vectors $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ where $\mathbf{u}_i \in \mathbb{C}^n$ for all i, and the k-tuple $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is orthonormal.

Hence with precoding matrices $\mathbf{T}_j$, the received signal for user j can be simplified to

$$\mathbf{y}_j = \mathbf{H}_j\mathbf{T}_j\mathbf{x}_j + \mathbf{v}_j. \quad (3)$$

Let $\widetilde{\mathbf{H}}_j = [\mathbf{H}_1^T \cdots \mathbf{H}_{j-1}^T \ \mathbf{H}_{j+1}^T \cdots \mathbf{H}_K^T]^T$, where $()^T$ denotes the matrix transpose. To satisfy the constraint in (2), $\mathbf{T}_j$ shall be in the null space of $\widetilde{\mathbf{H}}_j$. Let $\widetilde{N}_j$ denote the rank of $\widetilde{\mathbf{H}}_j$. Let the singular value decomposition of $\widetilde{\mathbf{H}}_j$ be $\widetilde{\mathbf{H}}_j = \widetilde{\mathbf{U}}_j\widetilde{\boldsymbol{\Lambda}}_j[\widetilde{\mathbf{V}}_j^1 \ \widetilde{\mathbf{V}}_j^0]^*$, where $\widetilde{\mathbf{V}}_j^1$ contains the first $\widetilde{N}_j$ right singular vectors and $\widetilde{\mathbf{V}}_j^0$ contains the last $(N_t - \widetilde{N}_j)$ right singular vectors of $\widetilde{\mathbf{H}}_j$. The columns in $\widetilde{\mathbf{V}}_j^0$ form a basis set in the null space of $\widetilde{\mathbf{H}}_j$, and hence $\mathbf{T}_j$ can be any rotated verison of $\widetilde{\mathbf{V}}_j^0$. Note that $\mathbf{T}_j$ satisfying (2) does not always exist. The sufficient condition for the existence of such matrices is $N_t \geq \sum_{j=1}^K N_{r,j}$, as shown in [10].

In the rest of the paper, we assume that every user has the same number of receive antennas, i.e. $\{N_{r,k}\}_{k=1}^K = N_r$.

III. BD vs. DPC: Sum Capacity for Fixed Channels

Consider a set of fixed channels for a multiuser MIMO system, where $\mathcal{K} = \{1, 2, \cdots, K\}$ denote the set of all users, and $\mathcal{A}_i$ be a subset of $\mathcal{K}$. Let $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ denote the effective channel after precoding for user $j \in \mathcal{A}_i$. The total throughput achieved with BD applied to the user set $\mathcal{A}_i$ with total power P can be expressed as

$$C_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2) =$$
$$\max_{\{\mathbf{Q}_j: \mathbf{Q}_j \geq 0, \sum_{j \in \mathcal{A}_i} \mathrm{Tr}(\mathbf{Q}_j) \leq P\}} \sum_{j \in \mathcal{A}_i} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\overline{\mathbf{H}}_j\mathbf{Q}_j\overline{\mathbf{H}}_j^*\right| (4)$$

where $\mathbf{Q}_j = E[\mathbf{x}_j\mathbf{x}_j^*]$ is user j's input covariance matrix of size $N_j \times N_j$ and $\mathrm{Tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. Let $\mathcal{A}$ be the set containing all $\mathcal{A}_i$, i.e. $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots\}$. The sum capacity of BD is defined as the maximum total throughput of BD over all possible user sets, i.e.

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = \max_{\mathcal{A}_i \in \mathcal{A}} C_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2). \quad (5)$$

It has been proven that the sum capacity of a multiuser Gaussian broadcast channel is achieved with dirty paper coding [3]. With the duality results in [3], the DPC sum capacity can be expressed as

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) =$$
$$\max_{\{\mathbf{S}_j: \mathbf{S}_j \geq 0, \sum_{j=1}^K \mathrm{Tr}(\mathbf{S}_j) \leq P\}} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\sum_{j=1}^K \mathbf{H}_j^*\mathbf{S}_j\mathbf{H}_j\right| \quad (6)$$

where $\mathbf{S}_j$ of size $N_r \times N_r$ is the signal covariance matrix for user j in the dual multiple access channel.

In this section, we are interested in the gain of DPC over BD in terms of sum capacity. Analogous to [15], we define the ratio of DPC to BD as

$$G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \triangleq \frac{C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}. \quad (7)$$

The gain is obviously dependent on the channel realizations $\{\mathbf{H}_k\}_{k=1}^K$, the total power, and noise variance. In the next theorem, we give a bound on $G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)$ that is valid for any $\{\mathbf{H}_k\}_{k=1}^K$, P, and σ^2.

Theorem 1: The sum capacity gain of DPC over BD is lower bounded by 1 and upper bounded by the minimum of N_t and K, i.e.

$$1 \leq G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq \min\{N_t, K\} \quad (8)$$

Proof: Theorem 3 in [15] states that

$$\frac{C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)} \leq \min\{N_t, K\} \quad (9)$$

where

$$C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) =$$
$$\max_k \max_{\{\mathbf{Q}_k: \mathbf{Q}_k \geq 0, \mathrm{Tr}(\mathbf{Q}_k) \leq P\}} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^*\right| (10)$$

is the maximum single user capacity among all users. The definition of the sum capacity for BD indicates that

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \geq C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2). \quad (11)$$

Further, since DPC is optimal for sum capacity, we have

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \geq C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2). \quad (12)$$

Combining (9), (11), and (12) completes the proof. $\square$

The above bound can be tightened in two special cases.

Lemma 1: Assume $N_r \leq N_t$ and $K \leq \lfloor\frac{N_t}{N_r}\rfloor$. If $\{\mathbf{H}_k\}_{k=1}^K$ are mutually orthogonal $\mathbf{H}_1 \perp \mathbf{H}_2 \perp \cdots \perp \mathbf{H}_K$, i.e. $\mathbf{H}_i\mathbf{H}_j^* = 0$ for $i \neq j$, then

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2).$$

Proof: Please see appendix I. $\square$

Lemma 1 shows when the user channels are mutually orthogonal, user cooperation is not necessary to achieve the sum capacity because all users do not interfere with each other. Interestingly, BD can also achieve the same capacity in this case. This is different from the TDMA scheme in [15]

where, even if the users are mutually orthogonal, it is not possible to achieve the same sum capacity as DPC. Actually, the gain of DPC over TDMA can still be at the maximum, i.e. $\min\{N_t, K\}$, when the users are mutually orthogonal.

The next Lemma shows a bound on the gain of DPC over BD when all user channels are in the same vector subspace.

Lemma 2: Assume $N_r \leq N_t$. If the row vector spaces of all user channels are the same, i.e. $\text{span}(\mathbf{H}_1) = \text{span}(\mathbf{H}_2) = \cdots = \text{span}(\mathbf{H}_K)$, which is denoted as $\mathcal{W}$, then

$$G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq \min\{N_r, K\}.$$

Proof: Please see appendix II. $\square$

IV. BD vs. DPC: Ergodic Sum Capacity in Rayleigh Fading Channels

In this section, we analyze the ergodic capacity of a multiuser MIMO system with block diagonalization vs. DPC. Let $\overline{\mathbf{H}}_j = \mathbf{H}_j \mathbf{T}_j$ (of size $N_r \times N_j$) be the effective channel for user j after precoding. Assuming that $\mathbf{H}_j$ are independent for different j and the elements in $\mathbf{H}_j$ are i.i.d. complex Gaussian random variables, we have the following theorem on the probability density function of $\overline{\mathbf{H}}_j$.

Theorem 2: In a downlink MIMO system with Block Diagonalization applied to a fixed set of users, if the MIMO channel for each user is modeled as i.i.d. complex Gaussian, then the effective channel after precoding is also an i.i.d. complex Gaussian matrix.

Proof: Since $\overline{\mathbf{H}}_j = \mathbf{H}_j \mathbf{T}_j$ and $\mathbf{H}_j$ is i.i.d. complex Gaussian, then $\overline{\mathbf{H}}_j$ conditioned on $\mathbf{T}_j$ is also complex Gaussian and independent of $\mathbf{T}_j$. Hence $\overline{\mathbf{H}}_j$ is independent of $\mathbf{T}_j$. $\square$

Theorem 2 indicates that if BD is applied to a fixed set of users, the ergodic capacity of user j can be easily evaluated with the eigenvalue distribution of $\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*$ [1] [18].

A. A Lower Bound on Ergodic Sum Capacity with BD

Let $\mathcal{A}_i = \{1, \cdots, i\}$ be a subset of users, i.e. $\mathcal{A}_i \in \mathcal{K}$, for $i = 1, \cdots, I$ where $I = \min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\}$. With $i \leq \min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\}$ and the elements in $\{\mathbf{H}\}_{k=1}^K$ are generated according to an i.i.d. complex Gaussian distribution, we have

$$
\begin{aligned}
E\left[C_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2)\right] &\overset{(a)}{\geq} E\left[\sum_{j=1}^{i} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\overline{\mathbf{H}}_j \mathbf{Q}_j \overline{\mathbf{H}}_j^*\right|\right] \\
&= \sum_{j=1}^{i} E\left[\sum_{n=1}^{N_r} \log\left|1 + \frac{P_{j,n}}{\sigma^2}\overline{\lambda}_{j,n}^2\right|\right] \\
&\overset{(b)}{\geq} \sum_{j=1}^{i} E\left[\sum_{n=1}^{N_r} \log\left|1 + \frac{P}{iN_r\sigma^2}\overline{\lambda}_{j,n}^2\right|\right] \\
&= \sum_{j=1}^{i} N_r E\left[\log\left|1 + \frac{P}{iN_r\sigma^2}\overline{\lambda}_{j,1}^2\right|\right] \\
&\overset{(c)}{=} i N_r E\left[\log\left|1 + \frac{P}{iN_r\sigma^2}\overline{\lambda}_{i,1}^2\right|\right] \quad (13) \\
&\triangleq \overline{C}_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2)
\end{aligned}
$$

where $\overline{\lambda}_{j,n}^2$ are nth unordered eigenvalues of $\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*$ and $\overline{\mathbf{H}}_j$ is of size $N_r \times (N_t - (i-1)N_r)$. Inequality (a) holds because the RHS assumes all i users are simultaneously transmitting for all channel realizations. Inequality (b) holds because the RHS assumes equal power is allocated to every non-zero eigenmodes. Equality (c) holds because $\overline{\lambda}_{j,1}$ has the same distribution for $j = 1, 2, \cdots, i$.

For notational simplicity, we denote $\alpha_i = \overline{\lambda}_{i,1}^2$ and $\overline{N}_i = N_t - (i-1)N_r$. With Theorem 2 and [1], the distribution of α_i can be expressed as

$$p_{\overline{N}_i, N_r}(\alpha_i) = \frac{1}{N_r}\sum_{m=1}^{N_r} \varphi_m(\alpha_i)^2 \alpha_i^{\overline{N}_i - N_r} e^{-\alpha_i} \quad (14)$$

where

$$\varphi_{k+1}(\alpha_i) = \left[\frac{k!}{(k + \overline{N}_i - N_r)!}\right]^{1/2} L_k^{\overline{N}_i - N_r}(\alpha_i) \quad (15)$$

for $k = 0, 1, \cdots, m - 1$, and

$$L_k^{n-m}(x) = \frac{1}{k!}e^x x^{m-n}\frac{d^k}{dx^k}\left(e^{-x}x^{n-m+k}\right). \quad (16)$$

Hence (13) can be evaluated with a numerical integration.

Now we can lower bound the ergodic sum capacity with BD by

$$E\left[C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right] \geq \max_{1 \leq i \leq I} \overline{C}_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2). \quad (17)$$

It is important to note that in order to evaluate the lower bound, up to $I = \min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\}$ numerical integrations need to be carried out because of the maximization in the RHS of (17).

B. An Upper Bound on the Ergodic Sum Capacity of DPC

It is well known that the sum capacity of a K-user broadcast channel with DPC is upper bounded if the receivers are allowed to cooperate [2][3]. Let $\mathbf{H} = [\mathbf{H}_1^T \ \mathbf{H}_2^T \ \cdots \ \mathbf{H}_K^T]^T$, and $N = \max\{N_t, KN_r\}$ and $M = \min\{N_t, KN_R\}$, then

$$
\begin{aligned}
E\left[C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right] &\leq E\left[\log\left|\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^*\right|\right] \\
&= \sum_{m=1}^{M} E\left[\log\left(1 + \frac{P_m}{\sigma^2}\lambda_m^2\right)\right] = ME\left[\log\left(1 + \frac{P_1}{\sigma^2}\alpha_1\right)\right] \\
&\leq M \int_{\sigma^2/\Gamma_0}^{\infty} \log\left(\frac{\Gamma_0\alpha_1}{\sigma^2}\right) p_{N,M}(\alpha_1)d\alpha_1 \quad (18) \\
&\triangleq \overline{C}_{coop}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \quad (19)
\end{aligned}
$$

where λ_m^2 is mth unordered eigenvalue of $\mathbf{H}^*\mathbf{H}$ and $\alpha_1 \triangleq \lambda_1^2$; $p_{N,M}(\alpha_1)$ is the distribution for α_1, which is given by (14) with N_r and $\overline{N}_i$ replaced by M and N respectively. The parameter Γ_0 is optimized so that the ergodic sum capacity is maximized with the average power constraint, i.e. $M \int_{\sigma^2/\Gamma_0}^{\infty}\left(\Gamma_0 - \frac{\sigma^2}{\alpha}\right) p_{N,M}(\alpha)d\alpha = P$. Details on the inequality (18) can be found in [17].
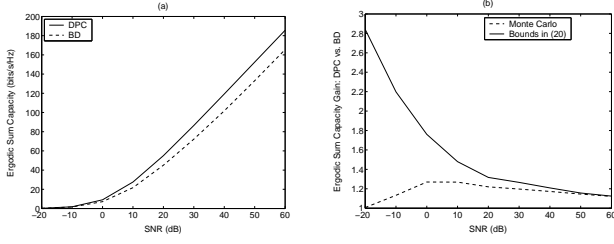
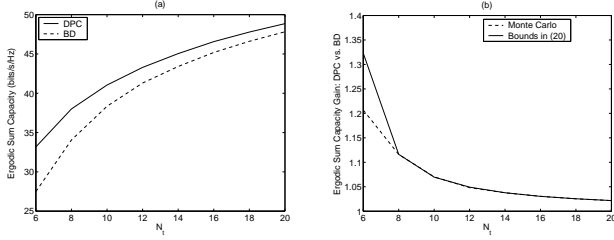Fig. 1. Ergodic sum capacity of DPC vs. BD in Rayleigh fading channels. $N_t = 10$, $N_r = 2$, $K = 5$.

Fig. 2. Ergodic sum capacity of DPC vs. BD in Rayleigh fading channels. $N_r = 2$, $K = 3$, SNR $= 20$ dB.

Fig. 3. Ergodic sum capacity of DPC vs. BD in Rayleigh fading channels. $N_t = 10$, $N_r = 2$, SNR $= 20$ dB.

C. An Upper Bound on the Ergodic Capacity of DPC vs. BD

From the above two sections, we can upper bound the ergodic sum capacity gain of DPC over BD as

$$\frac{E\left[C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right]}{E\left[C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right]} \leq \frac{\overline{C}_{coop}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{\max_{1 \leq i \leq I} \overline{C}_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2)}. \quad (20)$$

Notice that the upper bound in (20) is a function of N_t, N_r, K, P, and σ^2. Furthermore, $\min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\} + 1$ numerical integrations are necessary to evaluate the bound in (20).

V. NUMERICAL RESULTS

In this section, we provide some numerical demonstrations of the gain of DPC over BD.

Fig.1 (a) shows the ergodic sum capacity of DPC vs. BD under different SNRs, with $N_t = 10$, $N_r = 2$, and $K = 5$. In the low SNR regime, BD achieves almost the same sum capacity as DPC because beamforming to the user with the best channel eigenvalue is asymptotically optimal for sum capacity in low SNRs. As SNR goes to infinity, the sum capacity of both DPC and BD increase with the same slope because both BD and DPC exploit the maximum number of eigenmodes among the users. Fig. 1 (b) shows the gain of DPC over BD from the curves in Fig. 1 (a), as well as the bound on the gain evaluated from (20). As SNR increases, the bound from (20) gets tighter to the results from Monte Carlo simulations. For low SNR, the bound in (20) is loose mainly because 1) the lower bound on BD assume equal power allocation to all non-zero eigenvalues; 2) the cooperative upper bound on DPC is also loose in low SNR.

Fig. 2 (a) shows the ergodic sum capacity of DPC vs. BD for different N_t, with $N_r = 2$, $K = 3$, and SNR $= 20$ dB. As the number of transmit antenna increases, the sum capacity of BD gets closer to the sum capacity of DPC. Fig. 2 (b) shows
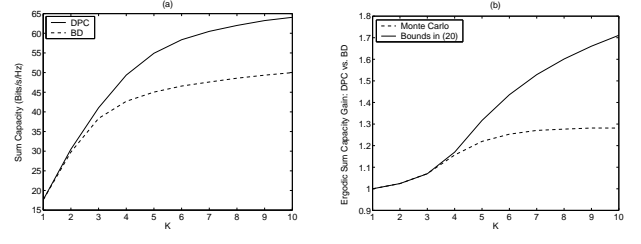
the gain of DPC over BD from the curves in Fig. 2 (a). It is observed that the bound from (20) is very tight for $N_t \geq 8$, with the specified N_r, K, and SNR.

Fig. 3 (a) shows the ergodic sum capacity of DPC vs. BD for different numbers of users, with $N_t = 10$, $N_r = 2$, and SNR $= 20$ dB. For small numbers of users, BD achieves almost the same sum capacity as DPC. As the number of users increases, DPC exhibits higher performance than BD. Fig. 3 (b) show the gain of DPC over BD from the curves on Fig. 3 (a). For small numbers of users, the bound from (20) is very tight compared to the Monte Carlo simulations. For larger numbers of users, the bound from (20) loosens. The main reason is that the lower bound on the sum capacity of BD in (17) only considers users $1 - 5$ and hence the effect of multiuser diversity is not reflected in (17).

In summary, BD achieves a significant part of the sum capacity for low and high regimes, or when $N_t \gg KN_r$. The bound in (20) is tight for median to high SNRs or when $K \leq \lfloor \frac{N_t}{N_r} \rfloor$.

APPENDIX I
PROOF OF LEMMA 1

Proof: Let the SVD of $\mathbf{H}_i$ be

$$\mathbf{H}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^* \quad (21)$$

where $\mathbf{U}_i$ is of size $N_r \times N_r$ and $\mathbf{U}_i \mathbf{U}_i^* = \mathbf{I}$; $\mathbf{\Lambda}_i = \mathrm{diag}\{\lambda_{i,1}, \lambda_{i,2}, \cdots, \lambda_{i,N_r}\}$ is a diagonal matrix of size $N_r \times N_r$; and $\mathbf{V}_i$ is of size $N_t \times N_r$ and $\mathbf{V}_i^* \mathbf{V}_i = \mathbf{I}$. Furthermore, $\mathbf{H}_i^* \mathbf{H}_i \mathbf{V}_i = \mathbf{V}_i \mathbf{\Lambda}_i^2$. For $i \neq j$, $\mathbf{V}_j^* \mathbf{V}_i = (\mathbf{\Lambda}_j^2)^{-1} \mathbf{V}_j^* \mathbf{H}_j^* \mathbf{H}_j \mathbf{H}_i^* \mathbf{H}_i \mathbf{V}_i (\mathbf{\Lambda}_i^2)^{-1} = 0$ because $\mathbf{H}_j \mathbf{H}_i^* = 0$.

Let $\mathbf{H} = [\mathbf{H}_1^* \ \mathbf{H}_2^* \ \cdots \ \mathbf{H}_K^*]^*$, then the SVD of $\mathbf{H}$ can be expressed as $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$, where $\mathbf{U} = \mathrm{bdiag}\{\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_K\}$ is an unitary block diagonal matrix of size $KN_r \times KN_r$; $\mathbf{\Lambda} = \mathrm{bdiag}\{\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \cdots, \mathbf{\Lambda}_K\}$ is a diagonal matrix of size $KN_r \times KN_r$; and $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2 \ \cdots \ \mathbf{V}_K]$ is of size $N_t \times KN_r$ and $\mathbf{V}^* \mathbf{V} = \mathbf{I}$.

The capacity of the point-to-point MIMO channel $\mathbf{H}$ can be regarded as an upper bound on the sum capacity of the broadcast channel because user cooperation is allowed with $\mathbf{H}$. Hence

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq C_{coop}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \quad (22)$$
$$= \sum_{i=1}^{K} \sum_{n=1}^{N_r} \log\left(1 + \frac{P_{i,n}}{\sigma^2} \lambda_{i,n}^2\right) \quad (23)$$

where $P_{i,n}$ is the power allocated to user i's nth eigenmode and $P_{i,n}$ is obtained by the water-filling algorithm with total power constraint $\sum_{i=1}^{K} \sum_{n=1}^{N_r} P_{i,n} = P$.

On the other hand, since $\mathbf{V}_j^* \mathbf{V}_i = 0$ for $i \neq j$, we have $\mathbf{H}_j^* \mathbf{V}_i = 0$ for $i \neq j$. Thus we can set $\mathbf{T}_j = \mathbf{V}_j$ to satisfy the null constraint in (2). Notice the effective channel $\overline{\mathbf{H}}_j = \mathbf{H}_j \mathbf{V}_j$ has the same singular values as $\mathbf{H}_j$. Hence

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \geq \max_{\{\mathbf{Q}_j : \mathbf{Q}_j \geq 0, \sum_{j \in \mathcal{K}} \text{Tr}(\mathbf{Q}_j) \leq P\}}$$

$$\sum_{j \in \mathcal{K}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \overline{\mathbf{H}}_j \mathbf{Q}_j \overline{\mathbf{H}}_j^* \right| \quad (24)$$

$$= \sum_{i=1}^{K} \sum_{n=1}^{N_r} \log \left(1 + \frac{P_{i,n}}{\sigma^2} \lambda_{i,n}^2 \right) \quad (25)$$

With (23), (25), and the fact that $C_{DPC} \geq C_{BD}$, we have $C_{DPC} = C_{BD}$ as the conditions in Lemma 3 are satisfied. $\square$

APPENDIX II
PROOF OF LEMMA 2

Proof: Let $\mathbf{E} = [\mathbf{e}_1^* \, \mathbf{e}_2^* \, \cdots \mathbf{e}_{N_r}^*]^*$ be a basis in $\mathcal{W}$, which is the row vector space spanned by $\{\mathbf{H}_i\}_{k=1}^{K}$, where $\mathbf{e}_i$ is of size $1 \times N_t$. Hence $\mathbf{E}\mathbf{E}^* = \mathbf{I}$. Let the SVD of $\mathbf{H}_i$ be $\mathbf{H}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^*$. There exists a unitary matrix $\mathbf{R}_i$ of size $N_r \times N_r$ such that $\mathbf{V}_i^* = \mathbf{R}_i \mathbf{E}$. Then $\mathbf{H}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{R}_i \mathbf{E}$. Denote $\mathbf{H}_i^{(\mathcal{W})} = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{R}_i$, it is easy to see that $\mathbf{H}_i^{(\mathcal{W})}$ has the same singular values of $\mathbf{H}_i$. Hence

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)$$

$$= \max_{\{\mathbf{S}_j : \mathbf{S}_j \geq 0, \sum_{j=1}^{K} \text{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{j=1}^{K} \mathbf{H}_j^* \mathbf{S}_j \mathbf{H}_j \right|$$

$$= \max_{\{\mathbf{S}_j : \mathbf{S}_j \geq 0, \sum_{j=1}^{K} \text{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{\sum_{j=1}^{K} \mathbf{E}^* (\mathbf{H}_i^{(\mathcal{W})})^* \mathbf{S}_j \mathbf{H}_i^{(\mathcal{W})} \mathbf{E}}{\sigma^2} \right|$$

$$= \max_{\{\mathbf{S}_j : \mathbf{S}_j \geq 0, \sum_{j=1}^{K} \text{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{j=1}^{K} (\mathbf{H}_i^{(\mathcal{W})})^* \mathbf{S}_j \mathbf{H}_i^{(\mathcal{W})} \right|$$

$$= C_{DPC}(\mathbf{H}_{1,\cdots,K}^{(\mathcal{W})}, P, \sigma^2). \quad (26)$$

Since the size of $\mathbf{H}_i^{(\mathcal{W})}$ (for $i = 1, 2, \cdots, K$) is $N_r \times N_r$, analogous to Theorem 1 in [15], we can obtain

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = C_{DPC}(\mathbf{H}_{1,\cdots,K}^{(\mathcal{W})}, P, \sigma^2)$$

$$\leq N_r \log \left(1 + \frac{P}{\sigma^2} \lambda_{max}^2 \right) \quad (27)$$

where $\lambda_{max} = \max_{1 \leq i \leq K, 1 \leq n \leq N_r} \lambda_{i,n}$ where $\lambda_{i,n}$ is the ith user's nth singular value.

On the other hand, if $\text{span}(\mathbf{H}_1) = \text{span}(\mathbf{H}_2) = \cdots = \text{span}(\mathbf{H}_K)$ and only one user is supported with BD, we have

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \quad (28)$$

$$\geq \log \left(1 + \frac{P}{\sigma^2} \lambda_{max}^2 \right) \quad (29)$$

Then we can immediately obtain

$$G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq \min\{N_r, K\} \quad (30)$$

by Theorem 3 in [15]. $\square$

REFERENCES

[1] I. E. Telatar, "Capacity of Multi-Antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[2] G. Caire and S. Shamai, "On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel," *IEEE Trans. on Information Theory*, vol. 49, No. 7, pp. 1691–1706, Jul. 2003.

[3] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. on Information Theory*, vol. 49, No. 10, pp. 2658–2668, Oct. 2003.

[4] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The Capacity Region of the Gaussian MIMO Broadcast channel," in *Proc. IEEE Int. Symposium on Information Theory*, pp. 174, Jun. 2004.

[5] M. Costa, "Writing on Dirty Paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[6] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Trans. on Information Theory*, vol. 49, No. 8, pp. 1912–1921 Aug. 2003

[7] W. Yu and J. M. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Trans. on Information Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[8] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative Water-Filling for Gaussian Vector Multiple-Access Channels," *IEEE Trans. on Information Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[9] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum Power Iterative Water-filling for Multi-Antenna Gaussian Broadcast Channels", *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[10] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[11] L. U. Choi and R. D. Murch, "A Transmit Preprocessing Technique for Multiuser MIMO Systems Using a Decomposition Approach," *IEEE Trans. on Wireless Communications*, vol. 3, no. 1, pp. 20–24, Jan. 2004.

[12] K. K. Wong, R. D. Murch, K. B. Letaief, "A Joint-Channel Diagonalization for Multiuser MIMO Antenna Systems," *IEEE Trans. on Wireless Communications*, vol. 2, no. 4, pp. 773–786, Jul. 2003.

[13] R. Chen, J. G. Andrews, and R. W. Heath. Jr., "Multiuser Space-Time Block Coded MIMO System with Downlink Precoding" in *Proc. IEEE Int. Conf. on Communications*, vol. 5, pp. 2689–2693, Jun. 2004.

[14] R. Chen, J. G. Andrews, and R. W. Heath. Jr., "Transmit Selection Diversity for Multiuser Spatial Multiplexing Systems", in *Proc. IEEE Global Communications Conf.*, vol. 4, pp. 2625–2629, Dec. 2004.

[15] N. Jindal and A. Goldsmith, "Dirty Paper Coding vs. TDMA for MIMO Broadcast Channels," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[16] T. Yoo and A. J. Goldsmith, "Optimality of Zero-Forcing Beamforming with Multiuser Diversity," in *IEEE Int. Conf. on Communications*, vol. 1, pp. 542–546, May 2005.

[17] Z. Shen, R. W. Heath. Jr., J. G. Andrews, and B. L. Evans, "Comparison of Space-Time Water-filling and Spatial Water-filling for MIMO Fading Channels," in *Proc. IEEE Global Communications Conf.*, vol. 1, pp. 431–435, Dec. 2004.

[18] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc. 1982.

[19] A. Edelman, *Eigenvalue and Condition Numbers of Random Matrices*, Ph.D. thesis, MIT, May 1989.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.