Correspondence

Dirty-Paper Coding Versus TDMA for MIMO Broadcast Channels

Nihar Jindal, *Member, IEEE,* and Andrea Goldsmith, *Fellow, IEEE*

***Abstract*—We compare the capacity of dirty-paper coding (DPC) to that of time-division multiple access (TDMA) for a multiple-antenna (multiple-input multiple-output (MIMO)) Gaussian broadcast channel (BC). We find that the sum-rate capacity (achievable using DPC) of the multiple-antenna BC is at most $\min(M, K)$ times the largest single-user capacity (i.e., the TDMA sum-rate) in the system, where M is the number of transmit antennas and K is the number of receivers. This result is independent of the number of receive antennas and the channel gain matrix, and is valid at all signal-to-noise ratios (SNRs). We investigate the tightness of this bound in a time-varying channel (assuming perfect channel knowledge at receivers and transmitters) where the channel experiences uncorrelated Rayleigh fading and in some situations we find that the dirty paper gain is upper-bounded by the ratio of transmit-to-receive antennas. We also show that $\min(M, K)$ upper-bounds the sum-rate gain of successive decoding over TDMA for the uplink channel, where M is the number of receive antennas at the base station and K is the number of transmitters.**

***Index Terms*—Broadcast channel (BC), channel capacity, dirty-paper coding (DPC), multiple-input multiple-output (MIMO) systems, time-division multiple access (TDMA).**

I. Introduction

In this correspondence, we consider a broadcast channel (downlink or BC) in which there are multiple antennas at the transmitter (base station) and possibly multiple antennas at each receiver (mobile). Dirty-paper coding (DPC) [1], [2] is a promising new transmission technique that allows a base station to efficiently transmit data to multiple users at the same time. DPC has been shown to achieve the sum-rate capacity (maximum throughput) of the multiple-antenna broadcast channel [1], [3]–[5], and furthermore, it was recently shown that DPC in fact achieves the full capacity region of the Gaussian multiple-input multiple-output (MIMO) broadcast channel [6]. However, DPC is a rather new and complicated scheme which has yet to be implemented in practical systems. Current wireless systems such as Qualcomm's High Date Rate (HDR) system [7] use the much simpler technique of time-division multiple-access (TDMA) in which the base transmits to only one user at a time.

Considering the difficulty in implementing DPC, a relevant question to ask is the following: How large of a performance boost does DPC provide over TDMA in terms of sum-rate? If DPC is used in a K-user broadcast channel, any rate vector in the K-dimensional DPC achievable region can be achieved. Similarly, if TDMA is used, any rate vector in the K-dimensional TDMA rate region can be achieved. It is easy to see that the DPC achievable region is larger than the TDMA rate region. However, defining a meaningful metric that quantifies the difference between two K-dimensional regions for $K \geq 2$ is quite difficult. Viswanathan, Venkatesan, and Huang [8] first investigated the above question by considering different operating points (i.e., rate vectors in the DPC and TDMA rate regions) that are reasonable for cellular systems, and numerically comparing the rates achievable with DPC and TDMA.

In this correspondence, we focus exclusively on the sum-rate capacity, or maximum throughput, achievable using DPC and TDMA. This operating point is quite reasonable when users have channels with roughly equivalent quality (i.e., no large signal-to-noise ratio (SNR) imbalances), but may not be as fair for asymmetric channels because users with higher SNRs receive a disproportionate fraction of the total data rate. However, the sum capacity is in general an important figure of merit because it quantifies how much total data flow is possible in a broadcast channel. Furthermore, comparing the maximum throughput achievable with DPC and TDMA gives a reasonable estimate of how much "larger" the DPC rate region is relative to the TDMA rate region. It should be noted that we only consider systems where both the transmitter and receiver have perfect and instantaneous channel state information (CSI).

By establishing upper and lower bounds to the DPC sum-rate capacity and the maximum TDMA sum-rate, respectively, we are able to analytically upper-bound the ratio of sum-rate capacity to the maximum TDMA sum-rate. Furthermore, we characterize the DPC gain at asymptotically high and low SNR. We also investigate the DPC gain in a time-varying, Rayleigh-fading channel in which the transmitter and receiver have perfect channel knowledge. Using the same techniques as for the downlink, we also upper-bound the sum-rate gain that successive decoding provides over TDMA on the uplink (multiple-access) channel.

The remainder of this correspondence is organized as follows. In Section II, we define our system model, and in Section III, we give definitions of the sum-rate capacity. In Section IV, we develop an analytical bound on the DPC gain, and we investigate the asymptotic behavior of the DPC gain at low and high SNR in Section V. We study the behavior of the DPC gain in Rayleigh-fading channels in Section VI. In Section VIII, we consider the DPC gain in a frequency-selective broadcast channel and in Section IX, we briefly compare DPC to transmitter beamforming, another suboptimal transmission strategy for the broadcast channel. We end by applying our analytical bounds to the multiple-antenna multiple-access channel in Section X and by stating some conclusions in Section XI.

II. System Model

We consider a broadcast channel with K receivers, $M > 1$ transmit antennas, and $N \geq 1$ receive antennas at each receiver. The transmitter sends independent information to each of the K receivers. The system is pictured in the left half of Fig. 1.

Let $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$ be the transmitted vector signal and let $\boldsymbol{H}_k \in \mathbb{C}^{N\times M}$ be the channel matrix of receiver k where $\boldsymbol{H}_k(i,j)$ represents the channel gain from transmit antenna j to antenna i of receiver k. The circularly symmetric complex Gaussian noise at receiver k is represented by $\boldsymbol{n}_k \in \mathbb{C}^{N\times 1}$ where $\boldsymbol{n}_k \sim N(0, \boldsymbol{I})$. Let $\boldsymbol{y}_k \in \mathbb{C}^{N\times 1}$ be

Manuscript received June 10, 2004; revised January 18, 2005. This work was supported in part by the Office of Naval Research under Grant N00014-02-1-0003. The material in this correspondence was presented at the IEEE International Conference on Communications, Paris, France, June, 2004.

N. Jindal is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nihar@ece.umn.edu).

A. Goldsmith is with the Electrical Engineering Department, Stanford University, Stanford, CA 94305-9515 USA (e-mail: andrea@systems.stanford.edu).

Communicated by R. Müller, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2005.846425

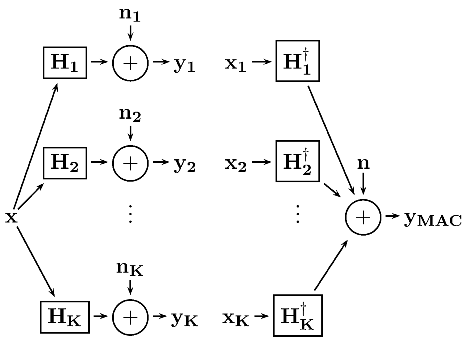


Fig. 1. System models of the MIMO BC (left) and the dual MIMO MAC (right).

the received signal at receiver k. The received signal is mathematically represented as

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x} + \boldsymbol{n}_k, \qquad k = 1, \ldots, K. \tag{1}$$

The covariance matrix of the input signal is $\boldsymbol{\Sigma}_x \triangleq \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]$. The transmitter is subject to an average power constraint P, which implies $\text{Tr}(\boldsymbol{\Sigma}_x) \leq P$. Throughout the correspondence, we refer to P as the SNR of the system, though the true received SNR is a function of the channel matrix and the transmission strategy. We let $\boldsymbol{H} \in \mathbb{C}^{NK \times M}$ refer to the concatenation of all channels, i.e., $\boldsymbol{H} \triangleq [\boldsymbol{H}_1^T \cdots \boldsymbol{H}_K^T]^T$. In the first half of this correspondence, we assume $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K$ are fixed and known perfectly at the transmitter and at all receivers. We explain the time-varying channel model in Section VI.

In terms of notation, we use $\boldsymbol{H}^\dagger$ to indicate the conjugate transpose of matrix $\boldsymbol{H}$ and $\|\boldsymbol{H}\|$ to denote the matrix norm of $\boldsymbol{H}$, defined by

$$\|\boldsymbol{H}\| = \sqrt{\lambda_{\max}(\boldsymbol{H}^\dagger \boldsymbol{H})}.$$

We also use boldface to indicate vector and matrix quantities.

III. Sum-Rate Capacity

For the single-antenna broadcast channel, the sum-rate capacity is equal to the largest single-user capacity in the system [9]. Equivalently, throughput is maximized by transmitting only to the user with the largest channel gain. In fact, the single-antenna Gaussian broadcast channel falls into the class of *degraded* broadcast channels [10], for which the sum-rate capacity is always equal to the largest single-user capacity in the system. A broadcast channel is degraded if users can be ordered in terms of the quality of received signal. In a degraded broadcast channel, there is always one user who has a stronger channel than every other user. For a single-antenna Gaussian broadcast channel, this is the user with the largest channel norm. Since this user has a stronger channel than any other user, he can decode any codeword intended for any other user, and thus sum-rate is maximized by only transmitting to this user in a degraded broadcast channel.

The multiple transmit antenna broadcast channel, however, is in general (but not always) a nondegraded broadcast channel. The fact that matrices can only be partially ordered, as opposed to the full ordering possible on scalars, is an intuitive explanation of the nondegraded nature of the multiple-antenna broadcast channel. For example, in a $K = 2, M = 2, N = 1$ channel with $\boldsymbol{H}_1 = [1\ 0.5]$ and $\boldsymbol{H}_2 = [0.5\ 1]$, it should be apparent that neither channel is absolutely stronger, since $\|\boldsymbol{H}_1\| = \|\boldsymbol{H}_2\|$ but clearly $\boldsymbol{H}_1 \neq \boldsymbol{H}_2$. It is also not *a priori* clear how sum capacity should be achieved. It turns out that the sum capacity of multiple-antenna broadcast channels is generally strictly larger than the single-user capacity of any of the users in the system, and the sum capacity is achieved by using DPC to simultaneously transmit to several users [1], [3]–[5].

DPC is a technique that can be used to pre-subtract interference at the transmitter. In a broadcast channel, the transmitted signal $\boldsymbol{x}$ generally contains information for multiple receivers. While receiver 1 is trying to decode his intended message, the portion of $\boldsymbol{x}$ not intended for receiver 1 is a form of interference. However, by using DPC, this interference can be reduced or completely eliminated. Furthermore, it turns out that this transmission strategy is in fact sum-rate capacity-achieving.

From the results in [1], [3]–[5], the sum-rate capacity of the BC, denoted by $\mathcal{C}_{\text{BC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$, can be expressed in terms of the following maximization:

$$\mathcal{C}_{\text{BC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) = \max_{\boldsymbol{\Sigma}_i \geq 0,\ \sum_{i=1}^K \text{Tr}(\boldsymbol{\Sigma}_i) \leq P} \sum_{i=1}^K \log \frac{|\boldsymbol{I} + \boldsymbol{H}_i (\sum_{j \leq i} \boldsymbol{\Sigma}_j) \boldsymbol{H}_i^\dagger|}{|\boldsymbol{I} + \boldsymbol{H}_i (\sum_{j < i} \boldsymbol{\Sigma}_j) \boldsymbol{H}_i^\dagger|} \tag{2}$$

where the maximization is over the set of $M \times M$ positive semidefinite covariance matrices $(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K)$. The objective function of the maximization is not a concave function of the covariance matrices. This makes the expression in (2) both numerically and analytically difficult to deal with. Fortunately, we are able to work with an alternative expression for the sum-rate capacity that comes from the dual multiple-access channel, as explained below.

In [3], the dirty paper rate region is shown to be equal to the capacity region of the dual Gaussian MIMO multiple-access channel (MAC or uplink) with sum power constraint P. The dual uplink is formed by reversing the roles of transmitters and receivers, i.e., converting the transmitter into an M-antenna receiver and converting each receiver into an N-antenna transmitter.[1] The dual MAC is shown in the right-hand side of Fig. 1. The received signal in the dual MAC is given by

$$\boldsymbol{y}_{\text{MAC}} = \sum_{i=1}^K \boldsymbol{H}_i^\dagger \boldsymbol{x}_i + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{x}_i \in \mathbb{C}^{N \times 1}$ is the ith transmitter signal, $\boldsymbol{H}_i^\dagger$ is the channel of the ith transmitter, and the noise is the same as in the downlink (i.e., each component is a circularly symmetric unit variance complex Gaussian). Notice that the dual uplink channel matrix of each user is equal to the conjugate transpose of the downlink channel matrix.

Due to the MAC-BC duality, the sum-rate capacity of the MIMO BC is equal to the sum-rate capacity of the dual MAC with sum power constraint P

$$\begin{aligned}\mathcal{C}_{\text{BC}}&(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) \\ &= \mathcal{C}_{\text{MAC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) \\ &= \max_{\{\boldsymbol{Q}_i : \boldsymbol{Q}_i \geq 0,\ \sum_{i=1}^K \text{Tr}(\boldsymbol{Q}_i) \leq P\}} \log \left| \boldsymbol{I} + \sum_{i=1}^K \boldsymbol{H}_i^\dagger \boldsymbol{Q}_i \boldsymbol{H}_i \right|\end{aligned} \tag{4}$$

where each of the matrices $\boldsymbol{Q}_i$ is an $N \times N$ positive semidefinite covariance matrix. The expression in (4) is the sum-rate capacity of the dual uplink subject to sum power constraint P. Note that (4) is a maximization of a concave function of the covariance matrices, for which efficient numerical algorithms exist. In this correspondence, we use the specialized algorithm developed in [11] to calculate $\mathcal{C}_{\text{BC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$.

[1] The dual MAC channel can be thought of as the uplink channel in a time-division duplex system, in which the channel coefficients are the same in both uplink and downlink mode. However, it is crucial to note that duality is only used as a mathematical tool in this work and it is not necessary for the actual system to be time-division duplexed for duality to be used.

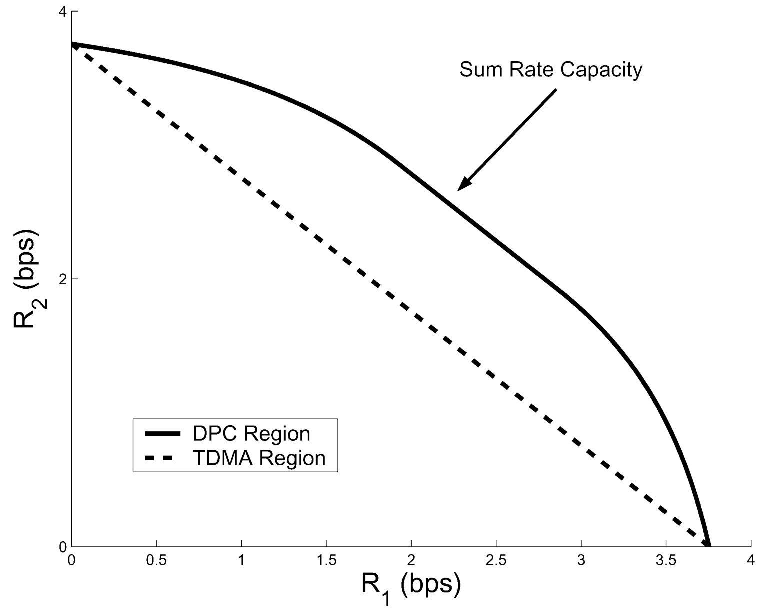


Fig. 2. DPC and TDMA rate regions for a two-user system with two transmit antennas.

The time-division rate region $\mathcal{R}_{\mathrm{TDMA}}$ is defined as the set of average rates that can be achieved by time sharing between single-user transmissions using constant power P

$$\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\triangleq\left\{(R_1,\ldots,R_K):\sum_{i=1}^{K}\frac{R_i}{C(\boldsymbol{H}_i,P)}\leq 1\right\} \tag{5}$$

where $C(\boldsymbol{H}_i,P)$ denotes the single-user capacity of the ith user subject to power constraint P. The single-user capacity of a MIMO channel is given by the following expression:

$$C(\boldsymbol{H}_i,P)=\max_{\{\boldsymbol{Q}_i:\boldsymbol{Q}_i\geq 0,\mathrm{Tr}(\boldsymbol{Q}_i)\leq P\}}\log\left|\boldsymbol{I}+\boldsymbol{H}_i\boldsymbol{Q}_i\boldsymbol{H}_i^\dagger\right|. \tag{6}$$

The maximum is achieved by choosing the covariance matrix $\boldsymbol{Q}_i$ to be along the eigenvectors of the channel matrix $\boldsymbol{H}_i^\dagger\boldsymbol{H}_i$ and by choosing the eigenvalues according to the water-filling procedure [12].

It is easy to see that the maximum sum-rate in $\mathcal{R}_{\mathrm{TDMA}}$, denoted as $C_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$, is the largest single-user capacity of the K users

$$\begin{aligned}\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)&\triangleq\max_{\boldsymbol{R}\in\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}\sum_{i=1}^{K}R_i\\&=\max_{i=1,\ldots,K}C(\boldsymbol{H}_i,P)\end{aligned} \tag{7}$$

and is achieved by transmitting only to the user with the largest capacity. We will refer to this quantity as the TDMA sum-rate.

In this correspondence, we are interested in quantifying the advantage that DPC gives over TDMA in terms of total throughput. Thus, the performance metric analyzed in this correspondence is the DPC gain $G(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$, which we define to be the ratio of sum-rate capacity to TDMA sum-rate

$$G(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\triangleq\frac{\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}. \tag{8}$$

Since $\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\geq\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$ by definition, the DPC gain is always greater than or equal to one. Notice that the DPC gain is a function of the channels $\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K$ and the SNR P.

In Fig. 2, the DPC and TDMA rate regions are shown for a two-user broadcast channel with two transmit antennas and single receive antenna. In this symmetric channel, the TDMA sum-rate is equal to the single-user capacity of either user (3.75 b/s), while the sum-rate capacity is equal to 4.79 b/s. Thus, the DPC gain is equal to 1.28. In Section IV, we develop an analytical upper bound to the DPC gain.

IV. Bounds on Sum-Rate Capacity and DPC Gain

In this section, we develop a precise analytical upper bound to the DPC gain defined in (8). In order to do so, we upper-bound the sum-rate capacity of the MIMO BC and lower-bound the TDMA sum-rate.

Theorem 1: The sum-rate capacity of the multiple-antenna downlink is upper-bounded by

$$\boxed{\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\leq M\log\left(1+\tfrac{P}{M}\|\boldsymbol{H}\|_{\max}^2\right)} \tag{9}$$

where $\|\boldsymbol{H}\|_{\max}=\max_{i=1,\ldots,K}\|\boldsymbol{H}_i\|$.

Proof: We prove this result using the fact that the BC sum-rate capacity is equal to the dual MAC sum-rate capacity with power constraint P. The received signal in the dual MAC is

$$\boldsymbol{y}_{\mathrm{MAC}}=\sum_{i=1}^{K}\boldsymbol{H}_i^\dagger\boldsymbol{x}_i+\boldsymbol{n}.$$

The received covariance is given by

$$\begin{aligned}\boldsymbol{\Sigma}_y&=E[\boldsymbol{y}\boldsymbol{y}^\dagger]=E[\boldsymbol{n}\boldsymbol{n}^\dagger]+\sum_{i=1}^{K}\boldsymbol{H}_i^\dagger E\left[\boldsymbol{x}_i\boldsymbol{x}_i^\dagger\right]\boldsymbol{H}_i\\&=\boldsymbol{I}+\sum_{i=1}^{K}\boldsymbol{H}_i^\dagger\boldsymbol{Q}_i\boldsymbol{H}_i.\end{aligned}$$

Notice that the argument of the maximization in the expression of the sum-rate capacity of the dual MAC in (4) is $\log|\boldsymbol{\Sigma}_y|$.

The received signal power is given by

$$E[\boldsymbol{y}^\dagger\boldsymbol{y}]=\sum_{i=1}^{K}E\left[\boldsymbol{x}_i^\dagger\boldsymbol{H}_i\boldsymbol{H}_i^\dagger\boldsymbol{x}_i\right]+E[\boldsymbol{n}^\dagger\boldsymbol{n}].$$

Since $\boldsymbol{x}_i^\dagger \boldsymbol{H}_i \boldsymbol{H}_i^\dagger \boldsymbol{x}_i \leq \|\boldsymbol{H}_i^\dagger\|^2 \|\boldsymbol{x}_i\|^2 = \|\boldsymbol{H}_i\|^2 \|\boldsymbol{x}_i\|^2$ by the definition of matrix norm, we have

$$E[\boldsymbol{y}^\dagger \boldsymbol{y}] \leq \sum_{i=1}^{K} \|\boldsymbol{H}_i\|^2 E\left[\boldsymbol{x}_i^\dagger \boldsymbol{x}_i\right] + E[\boldsymbol{n}^\dagger \boldsymbol{n}] \tag{10}$$

$$\leq \|\boldsymbol{H}\|_{\max}^2 \sum_{i=1}^{K} E\left[\boldsymbol{x}_i^\dagger \boldsymbol{x}_i\right] + M \tag{11}$$

$$\leq \|\boldsymbol{H}\|_{\max}^2 P + M \tag{12}$$

where (11) follows from the definition of $\|\boldsymbol{H}\|_{\max}$ and the fact that $E[\boldsymbol{n}^\dagger \boldsymbol{n}] = M$ and (12) follows from the sum power constraint on the transmitters in the dual MAC (i.e., $\sum_{i=1}^{K} E\left[\boldsymbol{x}_i^\dagger \boldsymbol{x}_i\right] \leq P$). Since

$$E[\boldsymbol{y}^\dagger \boldsymbol{y}] = Tr(E[\boldsymbol{y}\boldsymbol{y}^\dagger]) = Tr(\boldsymbol{\Sigma}_y)$$

this implies that $\mathrm{Tr}(\boldsymbol{\Sigma}_y) \leq P\|\boldsymbol{H}\|_{\max}^2 + M$. By [10, Theorem 16.8.4], for any positive definite $M \times M$ matrix $\boldsymbol{K}$,

$$|\boldsymbol{K}| \leq \left(\frac{\mathrm{Tr}(\boldsymbol{K})}{M}\right)^M.$$

Therefore,

$$|\boldsymbol{\Sigma}_y| \leq \left(1 + \frac{P}{M}\|\boldsymbol{H}\|_{\max}^2\right)^M$$

from which we get

$$\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) = \max_{\boldsymbol{\Sigma}_y} \log|\boldsymbol{\Sigma}_y| \leq M \log\left(1 + \frac{P}{M}\|\boldsymbol{H}\|_{\max}^2\right)$$

□

The upper bound is equal to the sum-rate capacity of a system with M spatially orthogonal eigenmodes (distributed in any manner between the K users), each with norm equal to $\|\boldsymbol{H}\|_{\max}$. Interestingly, users need not be spatially orthogonal for the bound to be achieved with equality. If $N = 1$ and there are more receivers than transmit antennas ($K > M$), then if the users' channels are Welch-bound equality sequences [13] (i.e., $\|\boldsymbol{H}_i\| = 1$ for all i and $\boldsymbol{H}^\dagger \boldsymbol{H} = \frac{K}{M}\boldsymbol{I}$), then the bound is also met with equality by allocating equal power (choosing $\boldsymbol{Q}_i = \frac{P}{K}$ in (4)) for each user in the dual MAC. Since $K > M$, it is not possible for the K channels (which are the $1 \times M$ rows of the matrix $\boldsymbol{H}$) to be mutually orthogonal. However, the Welch-bound condition requires the M *columns* of $\boldsymbol{H}$ to be orthogonal. The ith column of $\boldsymbol{H}$ refers to the channel gains from the ith base station antenna to each of the K mobiles.[2]

We now proceed by lower-bounding the TDMA sum-rate.

Theorem 2: The TDMA sum-rate is lower-bounded by the rate achieved by transmitting all power in the direction of the largest eigenmode

$$\boxed{\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) \geq \log\left(1 + P\|\boldsymbol{H}\|_{\max}^2\right).} \tag{13}$$

Proof: For each user, $C(\boldsymbol{H}_i, P) \geq \log(1 + P\|\boldsymbol{H}_i\|^2)$ because single-user capacity is achieved by water-filling over *all* eigenmodes instead of allocating all power to the best eigenmode. Since the TDMA sum-rate is the maximum of the single-user capacities, the result follows. □

This bound is tight when $N = 1$, but is generally not tight for $N > 1$ because each user has $\min(M, N)$ eigenmodes to water-fill over.

By combining Theorems 1 and 2, we can upper-bound the DPC gain

$$\frac{\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)}{\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)} \leq \frac{M \log\left(1 + \frac{P}{M}\|\boldsymbol{H}\|_{\max}^2\right)}{\log\left(1 + P\|\boldsymbol{H}\|_{\max}^2\right)} \tag{14}$$

$$\leq M \tag{15}$$

where we used Theorems 1 and 2 to get (14). Furthermore, since each user's rate in a broadcast channel can be no larger than his respective single-user capacity

$$\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) \leq \sum_{i=1}^{K} C(\boldsymbol{H}_i, P)$$
$$\leq K \cdot \mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P). \tag{16}$$

Combining the upper bounds in (15) and (16) gives the following result.

Theorem 3: The DPC gain is upper-bounded by M, the number of transmit antennas, and K, the number of users

$$\boxed{G(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) \leq \min(M, K).} \tag{17}$$

This bound is valid for any set of channels $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K$, any number of receive antennas N, any number of users K, and any SNR P. When we consider DPC and TDMA from the perspective of signal space dimensions, the upper bound is very intuitive. In the lower bound on TDMA in Theorem 2, only one spatial dimension (corresponding to the largest eigenmode among all users) is used. DPC, on the other hand, can utilize up to M dimensions[3] (Theorem 1). Since the TDMA lower bound uses the strongest eigenmode, the quality of each of these M spatial dimensions can be no better than the quality of the dimension used in the TDMA lower bound. Thus, the rate on each of the M dimensions can be no larger than the TDMA lower bound, which implies that DPC gives a sum rate no larger than M times the TDMA capacity.

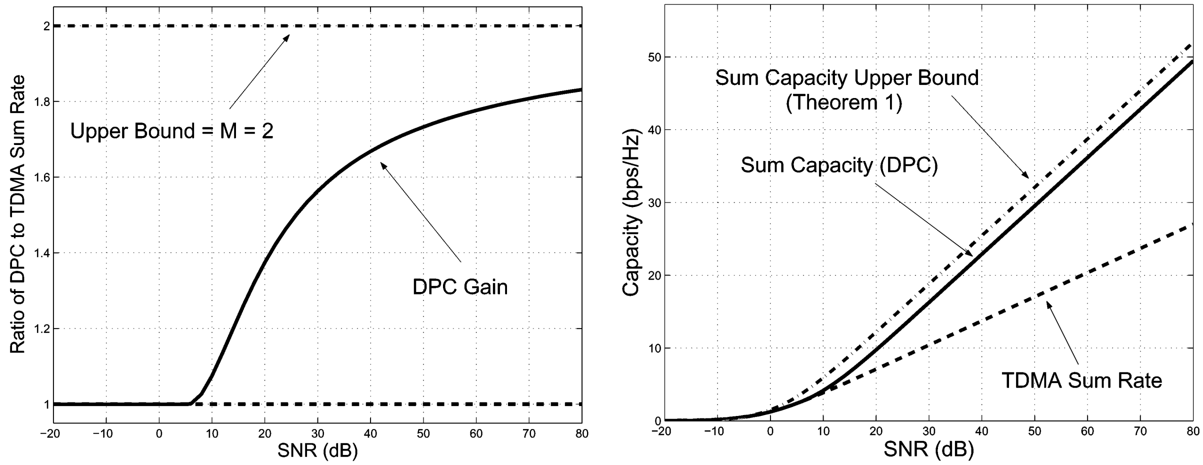
Note: A bound similar to Theorem 1 for the single receive antenna (($N = 1$) downlink when users have the same channel norm and are mutually orthogonal was independently derived in an earlier paper by Viswanathan and Kumaran [14, Proposition 2].

V. Asymptotic DPC Gain

In the regimes of high and low SNR, we are able to show tight results regarding convergence of the DPC gain. At asymptotically high SNR, the DPC gain converges to $\max\left(\min\left(\frac{M}{N}, K\right), 1\right)$, while at low SNR, the DPC gain converges to unity. We present these results in the following theorems (proofs are given in Appendices I and II).

[2]Note that if $N > 1$ and $KN > M$, the upper bound in Theorem 1 is also achieved without having spatially orthogonal channels (i.e., eigenvectors) if each row $\boldsymbol{H}$ has unity norm and $\boldsymbol{H}^\dagger \boldsymbol{H} = \frac{KN}{M}\boldsymbol{I}$ by choosing $\boldsymbol{Q}_i = \frac{P}{K}\boldsymbol{I}_N$ in (4)) for each user in the dual MAC.

[3]The term "dimensions" is used rather loosely when applied to DPC. In a single-user non-time-varying MIMO channel, spatial dimensions correspond to purely orthogonal signaling directions (generally corresponding to the right eigenvectors of the channel gain matrix). When using DPC, signaling dimensions are generally not orthogonal because orthogonal signaling generally leads to much lower data rates. Thus, the number of spatial dimensions is more accurately interpreted as the rank of the transmit covariance matrix when using DPC.



(a) (b)

Fig. 3. Plots of DPC gain and sum rate for a two-user, two transmit antenna, one receive antenna channel. (a) DPC gain. (b) Sum-rate plots.

Theorem 4: If the concatenated channel matrix $\boldsymbol{H}$ is full-rank and at least one of the channel matrices $\boldsymbol{H}_i$ is full-rank, as $P \to \infty$ we have

$$\lim_{P\to\infty} G(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P) = \max\left(\min\left(\frac{M}{N},K\right),1\right). \quad (18)$$

Theorem 5: For any channel matrix $\boldsymbol{H}$, DPC and TDMA are equivalent at asymptotically low SNR

$$\lim_{P\to 0} G(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P) = 1. \quad (19)$$

If there are more transmit antennas than receive antennas ($M \geq N$), then the high-SNR DPC gain is $\min\left(\frac{M}{N},K\right)$. If there are more receive antennas than transmit antennas ($N \geq M$), however, the DPC gain converges to unity at high SNR. The high-SNR DPC gain can be intuitively explained from a dimension counting argument as follows. At high SNR, it is easy to show that the capacity of a MIMO channel grows as $L\log(P)$ (to first order), where L is the number of spatial dimensions available in the channel, which is equal to the rank of the channel matrix. Therefore, only the number of spatial dimensions is important at high SNR and the quality of these spatial dimensions (i.e., the channel gain) is unimportant. This is expected since as P becomes large, $\log(1+\alpha P) \approx \log(P)+\log(\alpha) \approx \log(P)$. When using TDMA, there are $\min(M,N)$ dimensions available. This is equal to the number of dimensions available in an M-transmit, N-receive antenna MIMO channel. When using DPC, each linearly independent row (i.e., each received signal that is not equal to a linear combination of some other received signals) gives a spatial signaling dimension. If $\boldsymbol{H}$ is full-rank, the matrix has $\min(M,NK)$ linearly independent rows and thus DPC can utilize $\min(M,NK)$ spatial dimensions. When $M \geq N$, the ratio of spatial dimensions using DPC versus TDMA is given by

$$\frac{\min(M,NK)}{N} = \min\left(\frac{M}{N},K\right)$$

and thus, DPC gives a sum-rate equal to $\min\left(\frac{M}{N},K\right)$ times the sum rate in TDMA. If $M < N$, then DPC and TDMA can both utilize M spatial dimensions and thus, the DPC gain converges to unity at high SNR.

Since the sum-rate capacity grows as $\min(M,NK)\log(P)$ for large P, it is also easy to show that the sum-rate capacity upper bound given in Theorem 1 is asymptotically tight when $M \leq NK$, or that

$$\lim_{P\to\infty} \frac{\mathcal{C}_{\rm BC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{M\log\left(1+\frac{P}{M}\|\boldsymbol{H}\|^2_{\max}\right)} = 1. \quad (20)$$

Interestingly, the TDMA sum capacity lower bound in Theorem 2 is a factor of N times smaller than the actual TDMA sum capacity in the asymptotic limit

$$\lim_{P\to\infty} \frac{\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{\log\left(1+P\|\boldsymbol{H}\|^2_{\max}\right)} = N. \quad (21)$$

However, this factor of N does not preclude the general upper bound of $\min(M,K)$ on the DPC gain.

We can also compare the high-SNR behavior of the sum-rate capacity with that of the cooperative-receiver channel. Since receiver cooperation can only increase the capacity of the broadcast channel, the sum capacity $\mathcal{C}_{\rm BC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$ is upper-bounded by the capacity of the system in which the K receivers fully cooperate, i.e., the capacity of the M transmit, NK receive antenna MIMO channel, which is given by $C(\boldsymbol{H},P)$ as defined in (6). The cooperative upper-bound MIMO channel also has $\min(M,NK)$ spatial dimensions. Thus, at high SNR, the ratio of sum capacity of the broadcast channel to the capacity of the cooperative channel converges to unity. In fact, an even tighter result shows that the difference between the sum-rate capacity and the cooperative upper bound converges to zero as SNR goes to infinity [1, Theorem 3].

The intuition behind the low-SNR result in Theorem 5 is exactly the opposite of the high-SNR scenario. At low SNR, only the quality of the best signaling dimension is important and the number of available signaling dimensions is unimportant. To see this, note that for P small

$$\begin{aligned}\mathcal{C}_{\rm BC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P) &\leq M\log\left(1+\frac{P}{M}\|\boldsymbol{H}\|^2_{\max}\right)\\ &\approx M\frac{P}{M}\|\boldsymbol{H}\|^2_{\max} = P\|\boldsymbol{H}\|^2_{\max}\\ &\approx \mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P).\end{aligned}$$

DPC allows for simultaneous transmission over all of the different spatial dimensions. Since only the best signaling dimension is of importance at low SNR, this option is of no use and DPC and TDMA are equivalent at low enough SNR.

In Fig. 3(a), the DPC gain $G(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$ is plotted as a function of SNR for a two-user channel with two transmit antennas and

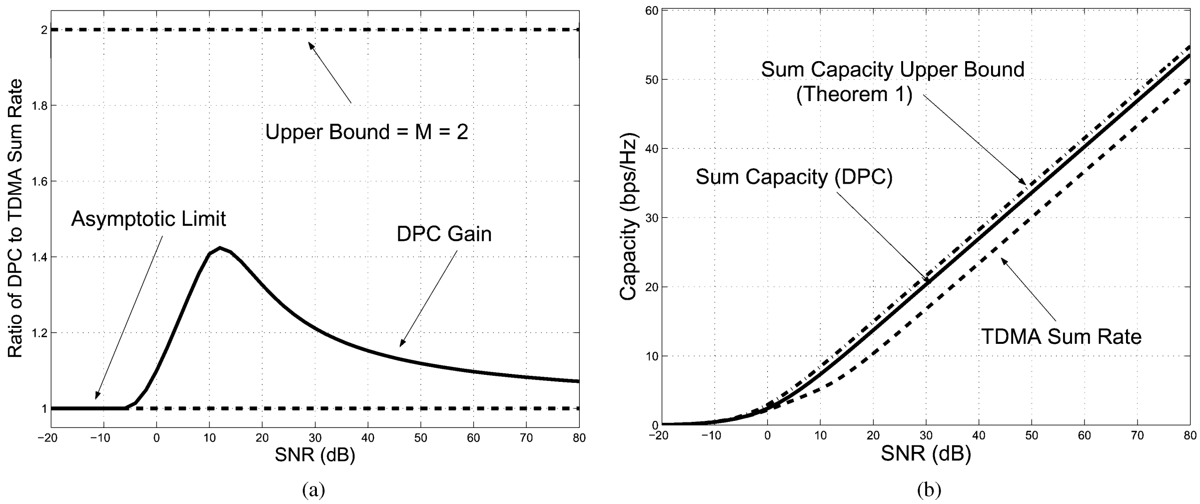


Fig. 4. Plots of DPC gain and sum rate for a two-user, two transmit antenna, two receive antenna channel. (a) DPC gain. (b) Sum-rate plots.

single receive antennas. The upper bound $(= M = 2)$ on the DPC gain is also included. Notice the monotonicity of the DPC gain and convergence of the gain at both low and high SNR. When $N = 1$, we conjecture that $G(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$ is in fact a monotonically nondecreasing function of P. Showing that $\mathcal{C}_{\rm BC}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$ and $\mathcal{C}_{\rm TDMA}(\boldsymbol{H}, P)$ are increasing functions of P is trivial, but it appears difficult to show the monotonicity of the ratio of these quantities for even the two-user, $N = 1$ scenario for which an exact expression for $\mathcal{C}_{\rm BC}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$ is known [1, Theorem 1]. In Fig. 3(b), the DPC and TDMA sum rate ($\mathcal{C}_{\rm BC}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$ and $\mathcal{C}_{\rm TDMA}(\boldsymbol{H}, P)$, respectively) are plotted for the same system, along with the sum-rate capacity upper bound $M \log\left(1 + \frac{P}{M}\|\boldsymbol{H}\|^2_{\max}\right)$ from Theorem 1. Note that the sum-rate capacity upper bound becomes tight in the ratio sense as $P \to \infty$. Furthermore, notice that the slope of the sum-rate capacity curve is approximately twice the slope of the TDMA sum rate curve, leading to the convergence of the DPC gain to two.

In Fig. 4(a), the DPC gain is plotted as a function of SNR for a two-user channel with two transmit antennas and two receive antennas per user. Notice that the DPC gain converges to $\frac{M}{N} = 1$ at both extremes, but takes its maximum at a finite SNR. When $M = N > 1$, $G(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$ is generally not monotonically nondecreasing and actually achieves its maximum at a finite SNR. When $M = N$ as in the figure, both TDMA $(\min(M, N) = M)$ and DPC $(\min(M, NK) = M)$ can use M spatial dimensions. When using TDMA, the transmitter must choose to use one of the K user's M spatial dimensions. When using DPC, the transmitter can choose M spatial dimensions from all NK available dimensions instead of being forced to choose one of the K sets (corresponding to each user) of N dimensions. This is not important at high SNR, where only the number of dimensions is relevant, or at low SNR, where only the strongest spatial dimensions are relevant. However, this improvement in the quality of dimensions leads to a strict DPC gain at finite SNR. In Fig. 4(b), the DPC and TDMA sum-rate are plotted for the same system, along with the sum-rate capacity upper bound from Theorem 1. In this channel, it is easy to see that all three curves have the same growth rate, i.e., are asymptotically equivalent in the ratio sense.

VI. Tightness of Bound in Rayleigh Fading

In this section, we consider the downlink sum-rate capacity in uncorrelated Rayleigh fading, i.e., where each entry of $\boldsymbol{H}_k$ is independently and identically distributed as a complex circularly symmetric Gaussian with unit variance. Here we consider a time-varying system, but we assume the transmitter and receiver have perfect and instantaneous CSI, and thus can adapt to the channel in each fading state. We also assume that the transmitter (the base station) is subject to a short-term power constraint, so that the base station must satisfy power constraint P in every fading state. This implies that there can be no adaptive power allocation over time.[4] Assuming that the fading process is ergodic, the sum-rate is equal to the expected value of the sum-rate in each fading state. Therefore, a reasonable performance metric is the ratio of the average sum rate using DPC to the average sum rate using TDMA, i.e.,

$$\frac{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\rm BC}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)\right]}{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)\right]}.$$

Note that this is not the same as the quantity $\mathbb{E}_{\boldsymbol{H}}[G(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)]$, which is not as meaningful when considering average rates achievable in a fading channel.

In the previous sections, we were able to establish bounds and asymptotic limits for the DPC gain for a fixed channel $\boldsymbol{H}$. In this section, we attempt to gain some intuition about the "average" DPC gain, where we average rates over Rayleigh-fading channels and then calculate the ratio. By Theorem 3, we have

$$\mathcal{C}_{\rm BC}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P) \le \min(M, K) \cdot \mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)$$

for each instantiation of $\boldsymbol{H}$. By taking the expectation of both sides, we get

$$\frac{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\rm BC}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)\right]}{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K, P)\right]} \le \min(M, K). \tag{22}$$

In this section, we show that this bound can be tightened to $\min\left(\frac{M}{N}, K\right)$ in the limit of high SNR, in the limit of a large number of transmit antennas, and in the limit of a large number of users. Note that the same limiting behavior occurs for the DPC gain of each channel instantiation in the limit of high SNR (Theorem 4). We also provide numerical results that show the DPC gain for nonasymptotic systems. To compute the sum-rate capacity for each channel instantiation, we use the algorithm provided in [11]. Furthermore, we use the standard Monte Carlo method to approximate the expected value of sum-rate over the distribution of $\boldsymbol{H}$.

[4]If the transmitter is subject to an average power constraint instead of a peak power constraint, the fading channel is theoretically equivalent to the frequency-selective broadcast channel model discussed in Section VIII. In the frequency-selective scenario, we show that the DPC gain is also upper-bounded by $\min(M, K)$.

A. High SNR

We first consider the scenario where M, N, and K are fixed, but the SNR P is taken to infinity. Furthermore, we assume $M \geq N$ and $M \leq KN$, which is quite reasonable for practical systems. In this scenario, the DPC gain is shown to asymptotically equal $\frac{M}{N}$

$$\lim_{P\to\infty} \frac{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]}{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]} = \frac{M}{N}. \tag{23}$$

We show convergence by establishing upper and lower bounds on TDMA and DPC sum-rate.[5]

Using Theorem 1 on the single-user ($K=1$) broadcast channel $\boldsymbol{H}_i^\dagger$, we can upper-bound the single-user capacity $C(\boldsymbol{H}_i,P)$ by $N\log\left(1+\frac{P}{N}\|\boldsymbol{H}_i\|^2\right)$. Then, using Jensen's inequality, the TDMA sum-rate can be bounded as

$$\begin{aligned}\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right] &\leq N\mathbb{E}_{\boldsymbol{H}}\left[\log\left(1+\frac{P}{N}\|\boldsymbol{H}\|^2_{\max}\right)\right]\\ &\leq N\log\left(1+\frac{P}{N}\mathbb{E}_{\boldsymbol{H}}\left[\|\boldsymbol{H}\|^2_{\max}\right]\right).\end{aligned}$$

Since $\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P) \geq \mathcal{C}(\boldsymbol{H}_1,P)$, we can lower-bound the TDMA capacity as

$$\begin{aligned}\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)] &\geq \mathbb{E}_{\boldsymbol{H}_1}[C(\boldsymbol{H}_1,P)]\\ &\geq \mathbb{E}_{\boldsymbol{H}}\left[\log\left|\boldsymbol{I}+\frac{P}{N}\boldsymbol{H}_1^\dagger\boldsymbol{H}_1\right|\right]\\ &= N\mathbb{E}_{\boldsymbol{H}}\left[\log\left(1+\frac{P}{N}\lambda_i\right)\right]\\ &\geq N\left(\log\left(\frac{P}{N}\right)+\mathbb{E}_{\boldsymbol{H}}\left[\log\left(\lambda_i\right)\right]\right)\end{aligned}$$

where λ_i is an unordered eigenvalue of the Wishart matrix $\boldsymbol{H}_1\boldsymbol{H}_1^\dagger$ and the single-user capacity is lower-bounded by transmitting equal power (as opposed to the optimal water-filling power allocation) on each of the N eigenmodes of User 1.

Using Theorem 1 and Jensen's inequality, we can upper-bound the sum-rate capacity as

$$\begin{aligned}\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right] &\leq M\mathbb{E}_{\boldsymbol{H}}\left[\log\left(1+\frac{P}{M}\|\boldsymbol{H}\|^2_{\max}\right)\right]\\ &\leq M\log\left(1+\frac{P}{M}\mathbb{E}_{\boldsymbol{H}}\left[\|\boldsymbol{H}\|^2_{\max}\right]\right).\end{aligned}$$

We can also lower-bound the sum-rate capacity by choosing $\boldsymbol{Q}_i = \frac{P}{KN}\boldsymbol{I}$ in (4) for each user

$$\begin{aligned}\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)] &\geq \mathbb{E}_{\boldsymbol{H}}\left[\log\left|\boldsymbol{I}+\frac{P}{KN}\boldsymbol{H}^\dagger\boldsymbol{H}\right|\right]\\ &= M\mathbb{E}_{\boldsymbol{H}}\left[\log\left(1+\frac{P}{KN}\lambda_i\right)\right]\\ &\geq M\left(\log\left(\frac{P}{KN}\right)+\mathbb{E}_{\boldsymbol{H}}\left[\log(\lambda_1)\right]\right)\end{aligned}$$

where λ_i is distributed as an unordered eigenvalue of the $M \times M$ Wishart matrix $\boldsymbol{H}^\dagger\boldsymbol{H}$. Using these bounds, as P becomes large, we can upper- and lower-bound the ratio

$$\frac{\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]}{\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]}$$

by $\frac{M}{N}$. It then follows that

$$\frac{\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]}{\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]}$$

converges to $\frac{M}{N}$ in the limit of high SNR. This result is intuitively closely related to Theorem 4.

In Fig. 5, the ratio of sum-rate capacity to the TDMA sum-rate (i.e., $\frac{\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]}{\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]}$) is plotted for a system with 10 users. The ratio is plotted for $M=4$ and $N=1$, $N=2$, and $N=4$. In each case the DPC gain converges to $\frac{M}{N}$, though it does so quite slowly for the $N=1$ case.

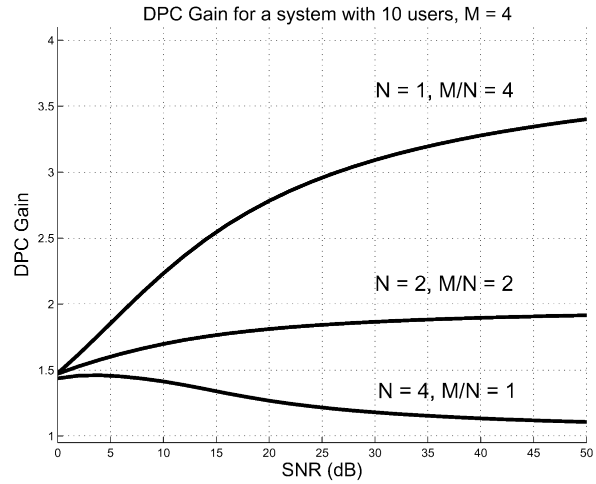


Fig. 5. DPC gain as a function of SNR for a system with 10 users.

B. Large M

In this subsection, we examine the scenario where the number of users (K), number of receive antennas (N), and SNR (P) are fixed but the number of transmit antennas (M) is taken to be large. We will show that the DPC gain converges to K in this case, i.e.,

$$\lim_{M\to\infty} \frac{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]}{\mathbb{E}_{H}\left[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]} = K. \tag{24}$$

As in the previous section, we lower-bound the sum-rate capacity by choosing $\boldsymbol{Q}_i = \frac{P}{KN}\boldsymbol{I}$ in (4) for each user and in each fading state. Since the identity covariance is optimal for point-to-point MIMO channels in Rayleigh fading [12], with the above choice of $\boldsymbol{Q}_i$ the lower bound is equal to the point-to-point capacity of a KN transmit, M receive MIMO channel where only the receiver has channel knowledge. If the number of receive antennas in this point-to-point link is allowed to become large (i.e., $M\to\infty$) but the number of transmit antennas in this point-to-point model (KN) is kept fixed, then the capacity of the point-to-point system grows as $KN\log\left(1+\frac{MP}{KN}\right)$ [16].

As in the previous section, the TDMA sum-rate is upper-bounded as

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)] \leq N\mathbb{E}_{\boldsymbol{H}}\left[\log\left(1+\frac{P}{N}\|\boldsymbol{H}\|^2_{\max}\right)\right].$$

Using standard probability arguments, we can upper-bound

$$\mathbb{E}_{\boldsymbol{H}}[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)]$$

by $N\log\left(1+\frac{P}{N}M(1+\alpha)\right)$, where α is any strictly positive number.

[5]Note that the high-SNR result in Theorem 4 cannot be extended in a straightforward manner here because the theorem only gives convergence for each instantiation of $\boldsymbol{H}$ and we require some uniformity of convergence for the result to hold in an expected value sense across instantiations.

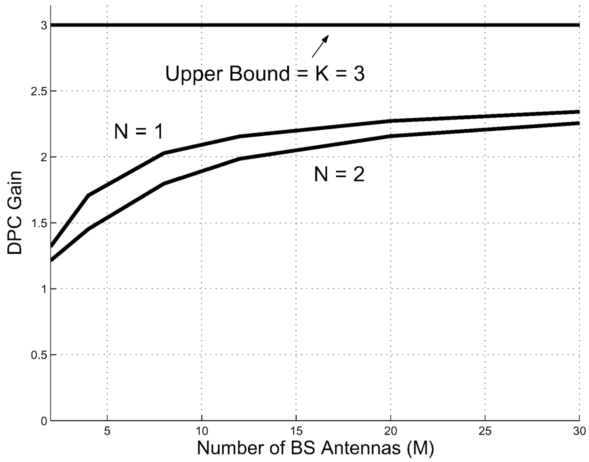


Fig. 6. DPC gain as a function of M for a system with three users at 10 dB.

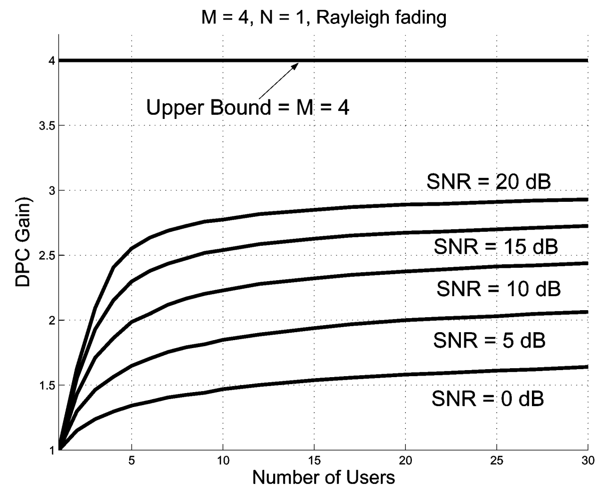


Fig. 7. DPC gain as a function of # of users for a system with four transmit antennas and one receive antenna.

If we now take the ratio of DPC sum-rate capacity to TDMA sum-rate as $M \to \infty$, we get

$$\begin{aligned}\lim_{M\to\infty} & \frac{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]}{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]} \\ &\geq \lim_{M\to\infty} \frac{KN\log\left(1+\frac{MP}{KN}\right)}{N\log\left(1+\frac{P}{N}M(1+\alpha)\right)} \\ &= K. \end{aligned} \tag{25}$$

By Theorem 3, this ratio is also upper-bounded by K for all M, Thus, in the limit of many transmit antennas and with a fixed number of receivers, the DPC gain goes to K. Intuitively, as M becomes large, the NK rows of $\boldsymbol{H}$ become mutually orthogonal because each row is a random vector in $\mathbb{C}^M$. Using TDMA, signaling can be done over N roughly orthogonal dimensions, whereas DPC allows signaling over NK dimensions. Thus, DPC can use K times as many signaling dimensions. Furthermore, the received SNR increases linearly with M because the transmitter has perfect knowledge of the channel and the gain of each antenna element is assumed to have a variance that is independent of M. Thus, we are effectively in the high-SNR regime (in terms of received SNR, but not in terms of P, which is fixed) when considering asymptotically large M, which implies that the factor of K increase in the number of spatial dimensions gained by using DPC translates to a factor of K increase in rate.

In Fig. 6, the DPC gain is plotted as a function of the number of transmit antennas for a system with three users, each with 10-dB average SNR. Notice that for both $N = 1$ and $N = 2$, slow convergence to $K = 3$ is observed as M becomes large.

C. Large K

If the number of antennas and the SNR are kept fixed and the number of users is taken to be large, it is shown in [17] that the dirty paper gain converges to $\frac{M}{N}$, i.e.,

$$\lim_{K\to\infty} \frac{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]}{\mathbb{E}_{\boldsymbol{H}}\left[\mathcal{C}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\right]} = \max\left(\frac{M}{N},1\right). \tag{26}$$

More specifically, the authors show that the sum-rate capacity and the TDMA sum-rate grow as $M\log\log(K)$ and $\min(M,N)\log\log(K)$, respectively. The intuition in this scenario is that as the number of users grows large, you can find a roughly orthogonal set (of size M or N) of channels to transmit over. Furthermore, the quality of these channels (i.e., the channel gain) grows roughly as $\log(K)$ because the maximum

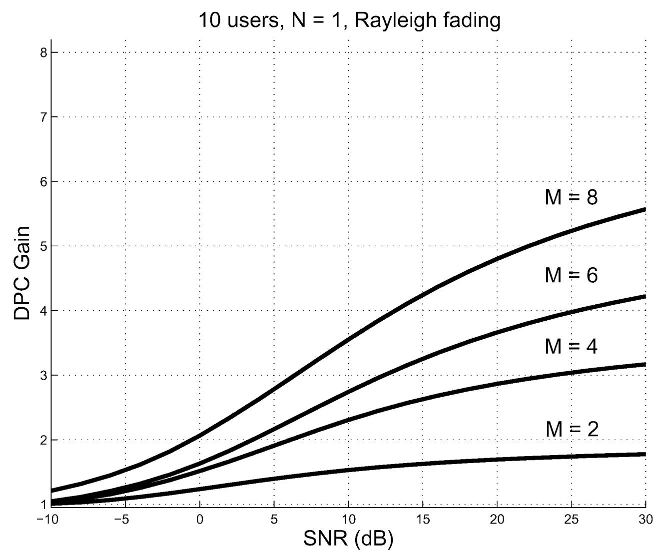


Fig. 8. DPC gain as a function of SNR for a system with 10 users and one receive antenna.

of independent exponential random variables grows logarithmically. Thus, the approximate rate is equal to (# of dimension) $\cdot \log\log(K)$.

D. Numerical Results

In this subsection, we provide plots and analysis of the DPC gain in Rayleigh fading for more realistic system parameters. In Fig. 7, the DPC gain is plotted as a function of the number of users for a system with four transmit antennas and one receive antenna. Plots are provided for different average SNR values. The DPC gain converges to four in the limit of a large number of users (i.e., each SNR curve converges to four), but convergence occurs extremely slowly, particularly for the lower SNR values. However, a factor of 2 to 3 increase in sum rate is possible for systems with 20 users and average SNRs ranging from 5 to 20 dB.

In Fig. 8, the DPC gain is plotted against SNR for systems with single receive antenna and differing numbers of transmit antennas. Each curve converges to M at very high SNR, and convergence occurs relatively quickly, unlike in the previous figure. Notice that at 0 dB, the DPC gain is larger than two only when there are eight transmit antennas. However, at 10 to 20 dB, gains of 2 to 5 are feasible if four or more antennas are used.

In general, note that the largest DPC gain occurs in systems with a large number of users relative to the number of transmit antennas operating at a high average SNR. Cellular systems typically have a large number of users per cell (i.e., 20–30), which is large compared to reasonable base station antenna deployments, but SNRs are typically quite low.

VII. Rate Region Bounds

Given that the sum-rate gain of DPC over TDMA can be elegantly upper-bounded by $\min(M,K)$, a natural question to ask is whether the entire DPC rate region can be upper-bounded by the same factor $\min(M,K)$ times the TDMA rate region, i.e., is

$$\mathcal{R}_{DPC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\subseteq\min(M,K)\cdot\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)?$$

We have thus far investigated the gain DPC provides when all users' rates are weighted equally, but a general rate region upper bound would bound the total gain DPC can provide for *any* weighting of transmission rates.

By single-user capacity bounds, we know that

$$\boldsymbol{R}\in\mathcal{R}_{\mathrm{DPC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$$

implies

$$\boldsymbol{R}_i\le C(\boldsymbol{H}_i,P),\qquad \text{for } i=1,\ldots,K.$$

Thus,

$$\frac{1}{K}\sum_{i=1}^{K}\boldsymbol{R}_i\le\sum_{i=1}^{K}\frac{1}{K}C(\boldsymbol{H}_i,P)$$

which by the definition of $\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$ in (5) implies $\frac{1}{K}\boldsymbol{R}\in\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$. Thus, it follows that

$$\mathcal{R}_{\mathrm{DPC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\subseteq K\cdot\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P).$$

However, it is in fact surprisingly simple to show that it is not always true that

$$\mathcal{R}_{\mathrm{DPC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\subseteq M\cdot\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P).$$

Consider a single transmit and single receive antenna broadcast channel ($M=N=1$). As long as each of the channel gains are not equal, the capacity region (which is equal to the DPC region for the scalar broadcast channel) of the scalar broadcast channel is strictly larger than the TDMA region [10], i.e.,

$$\mathcal{R}_{\mathrm{DPC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\supset\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P).$$

Interestingly, for symmetric channels (i.e., single-user capacities of all users are equal $C(\boldsymbol{H}_1,P)=C(\boldsymbol{H}_2,P)=\cdots=C(\boldsymbol{H}_K,P)$), the $\min(M,K)$ bound on the DPC gain implies that for any $\boldsymbol{R}$ in the DPC region,

$$\sum_{i=1}^{K}R_i\le\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\le\min(M,K)\cdot C(\boldsymbol{H}_1,P).$$

Therefore, it follows that

$$\mathcal{R}_{\mathrm{DPC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\subseteq\min(M,K)\cdot\mathcal{R}_{\mathrm{TDMA}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$$

for symmetric channels. However, this result is not true in general. The implication of this is that DPC can provide a multiplicative gain of larger than $\min(M,K)$ for non-sum-rate points for certain channels.

VIII. Frequency-Selective Broadcast Channels

A frequency-selective multiple-antenna broadcast channel can be decomposed into a set of parallel, independent flat-fading multiple-antenna broadcast channels [18]. In practical systems, orthogonal frequency division multiplexing (OFDM) can be used to reduce a frequency-selective channel into a finite number of parallel, frequency-flat MIMO broadcast channels (corresponding to each frequency tone). Either DPC or TDMA (i.e., transmit to only a single user on each tone) could be used on each tone. By Theorem 3, the gain of DPC over TDMA on each tone is upper-bounded by $\min(M,K)$. Therefore, the aggregate gain of DPC across all tones versus TDMA is also upper-bounded by $\min(M,K)$. This is true if there is a power constraint on each tone, or a total sum power constraint across all tones. In fact, the upper bound also holds if there is time-selective fading in addition to the frequency-selective fading. However, if only one user is selected for transmission across all tones, which implies that this user may not be the best user for each tone, the gain of using DPC on each tone is no longer upper-bounded by M, but is still upper-bounded by K, the number of users.

IX. Transmitter Beamforming

Transmitter beamforming[6] is a suboptimal technique that supports simultaneous transmission to multiple users on a broadcast channel. Each active user is assigned a beamforming direction by the transmitter and multiuser interference is treated as noise. Transmit beamforming is actually quite similar to DPC, but with DPC some multiuser interference is "pre-subtracted" at the transmitter, thus increasing the rates of some users. In [5], it is shown that transmitter beamforming for the broadcast channel without precoding is dual to receiver beamforming in the multiple-access channel without successive interference cancellation. As a result, when $N=1$, the maximum sum rate using beamforming can be written as [19]

$$C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)=\max_{\left\{P_i:\sum_{i=1}^{K}P_i\le P\right\}}\sum_{j=1}^{K}\log\frac{\left|\boldsymbol{I}+\sum_{i=1}^{K}\boldsymbol{H}_i^\dagger P_i\boldsymbol{H}_i\right|}{\left|\boldsymbol{I}+\sum_{i\ne j}\boldsymbol{H}_i^\dagger P_i\boldsymbol{H}_i\right|}.\quad(27)$$

This optimization cannot be cast in a convex form, and does not lend itself to numerical computation. However, we are able to analytically show that beamforming and DPC are equivalent at low and high SNR.

Theorem 6: If $\boldsymbol{H}$ has at least M independent rows, beamforming performs as well as DPC in the ratio sense at both asymptotically low and high SNR

$$\lim_{P\to\infty}\frac{C_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}=\lim_{P\to 0}\frac{C_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}=1.\quad(28)$$

Proof: By Theorem 1, we have

$$\mathcal{C}_{\mathrm{BC}}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)\le M\log\left(1+\frac{P}{M}\|\boldsymbol{H}\|_{\max}^2\right).$$

[6]Transmitter beamforming is also referred to as SDMA, or space-division multiple access.

$$\mathcal{R}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P}) \triangleq \bigcup_{\alpha_i\geq 0,\ \sum_{i=1}^K \alpha_i=1} \left(\alpha_1 C\left(\boldsymbol{H}_1, \frac{P_1}{\alpha_1}\right),\ldots,\alpha_K C\left(\boldsymbol{H}_K,\frac{P_K}{\alpha_K}\right)\right).$$

For simplicity, assume that the first M rows of $\boldsymbol{H}$ are linearly independent. In the proof of Theorem 4, we show that the sum-rate achievable using channel inversion is at least as large as

$$\sum_{i=1}^{M} \log\left(1+\alpha_i^2 \frac{P}{M}\right).$$

Channel inversion is a particular method of transmitter beamforming (zero-forcing beamforming), and thus,

$$C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P) \geq \sum_{i=1}^{M}\log\left(1+\alpha_i^2\frac{P}{M}\right).$$

Therefore, we have

$$\lim_{P\to\infty}\frac{C_{\rm BC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)} \leq \lim_{P\to\infty}\frac{M\log\left(1+\frac{P}{M}\|\boldsymbol{H}\|^2_{\max}\right)}{\sum_{i=1}^{M}\log\left(1+\alpha_i^2\frac{P}{M}\right)} = 1.$$

The low-SNR result follows directly from Theorem 5 and the fact that $C_{BF}(\boldsymbol{H},P)\geq C_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)$. □

The equivalence of transmitter beamforming and DPC at high SNR follows from the fact that both DPC and transmitter beamforming can use $\min(M,NK)$ signaling dimensions. However, the use of DPC reduces interference seen at the receivers and therefore improves the quality of each of the signaling dimensions, leading to an increase in sum-rate at finite SNR. Thus, an interesting open problem is to analytically bound the gain that DPC provides over transmitter beamforming. We conjecture that the ratio $\frac{C_{\rm BC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}$ is bounded by a constant ($< M$) independent of the number of antennas and the channel $\boldsymbol{H}$ for all P, but we have been unable to prove this due to the difficulty of working with the beamforming sum-rate expression in (27). Viswanathan and Venkatesan [20] recently characterized the performance of downlink beamforming and DPC as M and K both grow to infinity at some fixed ratio $\frac{M}{K}=\alpha$. In this asymptotic regime, the ratio $\frac{C_{\rm BC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}{C_{BF}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,P)}$ is bounded by 2 for all values of α and P.

X. Bound on Sum-Rate Gain of Successive Decoding for Uplink

Successive decoding is a capacity-achieving scheme for the MAC (uplink) in which multiple users simultaneously transmit to the base station and the receiver successively decodes and subtracts out the signals of different users. This technique achieves the sum-rate capacity of the MIMO MAC [10, Ch. 14], but is difficult to implement in practice. The sum-rate capacity of the MAC with transmitter power constraints $\boldsymbol{P}=(P_1,\ldots,P_K)$ is given by

$$\mathcal{C}_{\rm MAC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P}) = \max_{\{{\rm Tr}(\boldsymbol{Q}_i)\leq P_i\ \forall i\}}\log\left|\boldsymbol{I}+\sum_{i=1}^{K}\boldsymbol{H}_i^{\dagger}\boldsymbol{Q}_i\boldsymbol{H}_i\right|.$$

Notice that the sum-rate capacity of the MAC is identical to the BC sum-rate capacity expression in (4) except that the MAC expression has *individual* power constraints instead of a sum constraint.

Using the proof technique of Theorem 1 on the dual uplink (K transmitters with N antennas each and a single receiver with M antennas) modified for individual power constraints $\boldsymbol{P}=(P_1,\ldots,P_K)$, it is easy to see that the following holds:

$$\mathcal{C}_{\rm MAC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P}) \leq M\log\left(1+\frac{\sum_{i=1}^{K}P_i\|\boldsymbol{H}_i\|^2}{M}\right). \quad (29)$$

A suboptimal transmission scheme is to allow only one user to transmit at a time. Since each user in the uplink has an individual power constraint, users are allocated orthogonal time slots in which they transmit. Thus, the TDMA rate region is defined asshown in the equation at the top of the page. As before, the TDMA sum-rate is defined as the maximum sum of rates in this region. As used in the proof of Theorem 2, for each user

$$C\left(\boldsymbol{H}_i,\frac{P_i}{\alpha_i}\right)\geq\log\left(1+\frac{P_i}{\alpha_i}\|\boldsymbol{H}_i\|^2\right)$$

for any α_i. Thus,

$$\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P}) \geq \max_{\alpha_i\geq 0,\ \sum_{i=1}^K\alpha_i=1}\sum_{i=1}^{K}\alpha_i\log\left(1+\frac{P_i}{\alpha_i}\|\boldsymbol{H}_i\|^2\right).$$

The right—hand side of this expression corresponds to the TDMA region of a *scalar* MAC with channel gains $\|\boldsymbol{H}_1\|,\ldots,\|\boldsymbol{H}_K\|$. This expression is maximized by choosing

$$\alpha_i = \frac{P_i\|\boldsymbol{H}_i\|^2}{\sum_{j=1}^{K}P_j\|\boldsymbol{H}_j\|^2}$$

because this choice of α_i yields a sum-rate equal to the sum-rate capacity of the scalar MAC [10, Sec. 14.3.6]. With this choice of α_i, we get the following upper bound:

$$\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P}) \geq \log\left(1+\sum_{i=1}^{K}P_i\|\boldsymbol{H}_i\|^2\right). \quad (30)$$

Combining (29) and (30) we get

$$\frac{\mathcal{C}_{\rm MAC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P})}{\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P})}\leq M$$

As before, the single-user capacity of each user also upper bounds this ratio by K. Thus, we finally get

$$\boxed{\frac{\mathcal{C}_{\rm MAC}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P})}{\mathcal{C}_{\rm TDMA}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_K,\boldsymbol{P})}\leq\min(M,K)} \quad (31)$$

or that performing optimal successive decoding at the base station offers a gain of at most $\min(M,K)$ over TDMA.

XI. Conclusion

We have established that the sum-rate capacity of a multiple-antenna broadcast channel, achievable using DPC, is at most $\min(M,K)$ times larger than the maximum achievable sum rate using TDMA, where M is the number of transmit antennas (base station antennas) and K is the number of users. This bound applies at any SNR and for any number of receive antennas, and also generalizes to frequency-selective and time-selective channels. For Rayleigh-fading channels, the bound tightens to $\max\left(\min\left(\frac{M}{N},K\right),1\right)$ at high SNR, for a large number of transmit antennas, or for a large number of users. Using the same techniques for the uplink, we found that the performance gain using successive decoding on the uplink versus TDMA is also upper bounded by $\min(M,K)$, where M is the number of receive antennas (at the base station) and K is the number of mobiles (i.e., transmitters). Thus, it seems that for systems with many users that operate at relatively high SNRs, DPC can provide a significant performance gain if additional antennas are added only at the base station without the extra expense associated with adding additional antennas at each mobile. In contrast, if the number of mobile antennas is the same as the number of base station antennas or if the system is operating at low SNR, the benefit of using DPC on the downlink or successive decoding on the uplink is rather limited.

Appendix I
Proof of Theorem 4

The basic premise is to show that $C_{\rm BC}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P)$ grows as $\min(M,NK)\log(P)$ and $C_{\rm TDMA}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P)$ grows as $\min(M,N)\log(P)$. Let $L \triangleq \min(M,NK)$. Since $\boldsymbol{H}$ is full-rank by assumption, it has L linearly independent rows. We form a matrix (denoted by $\boldsymbol{H}^{L,L}$) consisting of any set of L linearly independent rows of $\boldsymbol{H}$. By the linear independence of the rows, $\boldsymbol{H}^{L,L}$ is invertible. Therefore, we can invert the channel at the transmitter (i.e., perform zero-forcing, whereby the transmitted signal is $\boldsymbol{x} = (\boldsymbol{H}^{L,L})^{-1}\boldsymbol{u}$, and $\boldsymbol{u}$ is the $L \times 1$ data vector) to give L independent and parallel nonzero channels. The result is L independent and parallel channels with strictly nonzero channel gains $\alpha_1,\dots,\alpha_L$. The channel gain α_i is equal to the norm of the projection of the ith row of $\boldsymbol{H}^{L,L}$ onto the other rows of $\boldsymbol{H}^{L,L}$. Due to the invertibility of $\boldsymbol{H}^{L,L}$, each of these channel gains is strictly positive. Thus, we have

$$C_{\rm BC}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P) \geq \sum_{i=1}^{L} \log\left(1+\alpha_i^2 \frac{P}{L}\right)$$

because allocating equal power on the parallel channels is suboptimal. Furthermore, it is easy to see that the sum capacity of the broadcast channel is upper-bounded by the capacity of the point-to-point MIMO channel where all K receivers cooperate, i.e., $C_{\rm BC}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P) \leq C(\boldsymbol{H},P)$. From single-user MIMO theory, it is well known that $C(\boldsymbol{H},P) \leq L\log\left(1+\frac{P}{L}\|\boldsymbol{H}\|^2\right)$. Thus, $C_{\rm BC}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P)$ is upper-bounded above and below by $L\log(P)$ as $P \to \infty$.

To show the order growth of $C_{\rm TDMA}$, without loss of generality, we assume $\boldsymbol{H}_1$ is full-rank. Since $\boldsymbol{H}_i\boldsymbol{H}_i^\dagger$ has $\min(M,N)$ nonzero eigenvalues, we have $C(\boldsymbol{H}_i,P) \leq \min(M,N)\log(1+P\|\boldsymbol{H}\|_{\max}^2)$. Therefore,

$$C_{\rm TDMA}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P) \leq \min(M,N)\log(1+P\|\boldsymbol{H}\|_{\max}^2)$$

by the definition of the TDMA capacity. Furthermore, we clearly have $C_{\rm TDMA}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P) \geq C(\boldsymbol{H}_1,P)$. Due to the water-filling structure of the capacity-achieving input for a single-user MIMO channel and because $\boldsymbol{H}_i\boldsymbol{H}_i^\dagger$ has $\min(M,N)$ nonzero eigenvalues, we have

$$C(\boldsymbol{H}_1,P) \geq \sum_{i=1}^{\min(M,N)} \log\left(1+\lambda_i \frac{P}{\min(M,N)}\right)$$

where $(\lambda_1,\dots,\lambda_{\min(M,N)})$ are the nonzero eigenvalues of $\boldsymbol{H}_1\boldsymbol{H}_1^\dagger$. Therefore, we have that $C_{\rm TDMA}$ is bounded above and below by $\min(M,N)\log(P)$ as $P \to \infty$.

Combining these bounds, we can establish that

$$\lim_{P\to\infty} \frac{C_{\rm BC}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P)}{C_{\rm TDMA}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P)}$$

is bounded above and below by $\frac{\min(M,NK)}{\min(M,N)}$. When $M \geq N$, we have

$$\frac{\min(M,NK)}{\min(M,N)} = \min\left(\frac{M}{N},K\right)$$

which is clearly greater than or equal to unity. When $M < N$, the quantity is equal to one. Thus, we have the result.

Appendix II
Proof of Theorem 5

Without loss of generality, assume that $\|\boldsymbol{H}_1\| \geq \|\boldsymbol{H}_2\| \cdots \geq \|\boldsymbol{H}_K\|$. First notice that on the dual MAC we have

$$\begin{aligned}
\log\left|\boldsymbol{I}+\sum_{i=1}^{K}\boldsymbol{H}_i^\dagger\boldsymbol{Q}_i\boldsymbol{H}_i\right| &= I(X_1,\dots,X_K;Y) && (32)\\
&= \sum_{i=1}^{K} I(X_i;Y|X_1,\dots,X_{i-1}) && (33)\\
&\leq \sum_{i=1}^{K} I(X_i;Y|X_1,\dots,X_{i-1},\\
&\qquad X_{i+1},\dots,X_K) && (34)\\
&= \sum_{i=1}^{K}\log\left|\boldsymbol{I}+\boldsymbol{H}_i^\dagger\boldsymbol{Q}_i\boldsymbol{H}_i\right| && (35)
\end{aligned}$$

where we have used the chain rule for mutual information in (33) and the fact that $X_1,\dots,X_K$ are independent to get

$$I(X_i;Y|X_1,\dots,X_{i-1},X_{i+1},\dots,X_K) \geq I(X_i;Y|X_1,\dots,X_{i-1})$$

in (34). Thus, we can upper-bound the sum capacity of the dual MIMO MAC by the sum capacity of a multiple-access channel where each transmitter has an independent channel $\left(\boldsymbol{H}_i^\dagger\right)$ to the receiver. By diagonalizing each $\boldsymbol{H}_i$, we can write the sum capacity of this upper bound as

$$\max_{P_{i,j}\geq 0:\sum_{i,j}P_{i,j}\leq P} \sum_{i,j}\log(1+P_{i,j}\lambda_{i,j}) \qquad (36)$$

where $\lambda_{i,j}$ is the jth eigenvalue of $\boldsymbol{H}_i\boldsymbol{H}_i^\dagger$. Assume that $\lambda_{i,j} \geq \lambda_{i,k}$ for all $k > j$. Notice that the optimal power allocation maximizing (36) is found by the standard water-filling procedure for parallel Gaussian channels [10, Ch. 10.4].

We must separately consider two different cases: $\|\boldsymbol{H}_1\| > \|\boldsymbol{H}_2\|$ and $\|\boldsymbol{H}_1\| = \|\boldsymbol{H}_2\|$. First consider $\|\boldsymbol{H}_1\| > \|\boldsymbol{H}_2\|$, which implies $\lambda_{1,1} > \lambda_{2,1}$. From the water-filling procedure, we know that for $P \leq \frac{1}{\lambda_{1,1}} - \frac{1}{\lambda_{2,1}}$, there is not enough power to fill any of the channels of Users $2,\dots,K$. Therefore, the maximum in (36) is achieved by only allocating power to User 1, i.e., TDMA is optimal, for small enough P. Thus, for $P \leq \frac{1}{\lambda_{1,1}} - \frac{1}{\lambda_{2,1}}$, we have

$$C_{\rm BC}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P) = C_{\rm TDMA}(\boldsymbol{H}_1,\dots,\boldsymbol{H}_K,P).$$

Now consider the scenario where $\|\boldsymbol{H}_1\| = \|\boldsymbol{H}_2\| = \cdots = \|\boldsymbol{H}_L\|$, i.e., where L users have the same largest eigenvalue. Let m_i be the multiplicity of the largest eigenvalue of the ith user, and assume $\sum_{i=1}^{L} m_i = J$. Then for $P \leq J\left(\frac{1}{\lambda_{1,1}} - \frac{1}{\lambda^*}\right)$, where λ^* is the second largest eigenvalue among all users, the maximum in (36) is achieved by allocating equal power to the J eigenmodes with the largest eigenvalue. The corresponding capacity is $J\log\left(1+\|\boldsymbol{H}_1\|^2\frac{P}{J}\right)$. In this case, note that

$$\lim_{P\to 0}\frac{J\log\left(1+\|\boldsymbol{H}_1\|^2\frac{P}{J}\right)}{\log\left(1+\|\boldsymbol{H}_1\|^2 P\right)} = 1. \qquad (37)$$

Thus, in the limit of small P, the ratio of the sum capacity to the TDMA capacity goes to one.

Acknowledgment

The authors wish to thank Jerry Foschini for his assistance with this work.

REFERENCES

[1] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[2] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[3] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[4] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," in *Proc. Int. Symp. Inforation. Theory*, Lausanne, Switzerland, Jun./Ju. 2002, p. 498.

[5] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink–downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[6] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2004.

[7] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR: A high efficiency-high data rate personal communication wireless system," in *Proc. IEEE Vehicular Technology Conf.*, Tokyo, Japan, May 2000, pp. 1854–1858.

[8] H. Viswanathan, S. Venkatesan, and H. C. Huang, "Downlink capacity evaluation of cellular networks with known interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 1, pp. 802–811, Jun. 2003.

[9] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels-Part I: Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] 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. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[12] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[13] M. Rupf and J. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1261–1266, Jul. 1994.

[14] H. Viswanathan and K. Kumaran, "Rate scheduling in multiple antenna downlink wireless systems," Lucent Tech. Memo., ITD-01-41685K.

[15] N. Jindal, S. Jafar, S. Vishwanath, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Asilomar, CA, 2002.

[16] B. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[17] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, submitted for publication.

[18] A. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 211–219, Jan. 2001.

[19] S. Venkatesan and H. Huang, "System capacity evaluation of multiple antenna systems using beamforming and dirty paper coding," unpublished manuscript.

[20] H. Viswanathan and S. Venkatesan, "Asymptotics of sum rate for dirty paper coding and beamforming in multiple-antenna broadcast channels," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.

Computing the Capacity of a MIMO Fading Channel Under PSK Signaling

Wenyan He and Costas N. Georghiades, *Fellow, IEEE*

***Abstract*—We study the constrained capacity of a multiple-input multiple-output (MIMO) fading channel with a phase-shift keying (PSK) input alphabet and show a uniform prior distribution is capacity achieving. An expression for the capacity is derived which requires a single expectation and can be evaluated easily through simulation. The simulations are facilitated by analytical expressions for the eigenvalues and eigenvectors of a required covariance matrix. The derived expression is used to provide good approximations to the capacity at low signal-to-noise ratios (SNRs) as well as to compare the input-constrained MIMO capacity to the unconstrained MIMO capacity.**

***Index Terms*—Capacity-achieving distribution, constrained input, multiple-input multiple-output (MIMO) capacity, phase-shift keying (PSK) signaling.**

I. INTRODUCTION

Since the work on multiple-input multiple-output (MIMO) systems by Telatar [1] and Foschini and Gans [2], MIMO capacity has become a subject of significant research. Foschini [3] pointed out that the MIMO capacity could be substantially higher than that of a single-antenna system. He went further to establish the relationship between the M-dimensional architecture and M one-dimensional architectures. The many currently available results on MIMO capacity are based on three exclusive assumptions: the channel known at both the transmitter and receiver, the channel known only at the receiver, and the channel known at neither the transmitter nor the receiver. In the first category, the water-filling power allocation on the singular values of the channel matrix is shown to be optimal at the transmitter [1], [4]. In the second category, when the channel matrix entries are independent and identically distributed (i.i.d.), uniform power allocation is assumed at the transmitter since it has no knowledge of the channel state. Compared to the first two cases, the third is mathematically more difficult to deal with and interest in it is relatively new. Marzetta and Hochwald [5] shed some light on this case by showing, for instance, that increasing the number of transmit antennas beyond the number of symbol periods in a coherence interval does not increase capacity. Zheng and Tse [6] also did some original work that falls into this category. They focused on the asymptotic capacity at high signal-to-noise ratio (SNR) and tried to give a geometric interpretation to the problem as sphere packing in the Grassmann manifold [7].

A brief overview of MIMO capacity can be found in [8]. Goldsmith *et al.* [9] gave a more detailed overview of recent results on single-user and multiuser MIMO capacity, in which multiple definitions of time-varying channel capacity are listed, i.e., outage capacity, ergodic capacity, and minimum-rate capacity. Since a closed-form expression for the MIMO capacity is unavailable, the asymptotic behavior naturally becomes a primary topic of interest. When the number of transmit and receive antennas is large, the instantaneous MIMO capacity, as a random variable, can be well approximated by Gaussian

Manuscript received April 12, 2004; revised January 30, 2005. This work was supported by a grant from the Telecommunication and Informatics Taskforce (TITF) at Texas A&M University.

The authors are with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: whe@ee.tamu.edu; georghiades@ee.tamu.edu).

Communicated by R. Müller, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2005.846426