

The Optimality of Beamforming in Uplink Multiuser Wireless Systems

Wonjong Rhee, *Member, IEEE*, Wei Yu, *Member, IEEE*, and John M. Cioffi, *Fellow, IEEE*

Abstract—This paper considers the optimal uplink transmission strategy that achieves the sum-capacity in a multiuser multi-antenna wireless system. Assuming an independent identically distributed block-fading model with transmitter channel side information, *beamforming* for each remote user is shown to be necessary for achieving sum-capacity when there is a large number of users in the system. This result stands even in the case where each user is equipped with a large number of transmit antennas, and it can be readily extended to channels with intersymbol interference if an orthogonal frequency division multiplexing modulation is assumed. This result is obtained by deriving a rank bound on the transmit covariance matrices, and it suggests that all users should cooperate by each user using only a small portion of available dimensions. Based on the result, a suboptimal transmit scheme is proposed for the situation where only partial channel side information is available at each transmitter. Simulations show that the suboptimal scheme is not only able to achieve a sum rate very close to the capacity, but also insensitive to channel estimation error.

Index Terms—Beamforming, multiple antennas, multiple-access channels, orthogonal frequency division multiplexing (OFDM), sum capacity.

I. INTRODUCTION

ONE of the most important progress in recent wireless system research is the realization of the profound benefit of multiple antennas for achieving high spectral efficiencies. This is apparent both from theoretic studies such as [1], in which the ergodic capacity of an independent identically distributed (i.i.d.) fading channel is shown to increase linearly with the number of transmit and receive antennas, and from system proposals such as [2], in which a linear-complexity decoder is shown to be capable of achieving data rates close to the capacity. The use of multiple antennas, especially at the transmitter, also opens up the area of optimal transmit signal design. Depending on the availability of channel knowledge at the transmitter, many different schemes of utilizing multiple antennas at the transmitter have been investigated in the

Manuscript received December 18, 2001; revised June 17, 2002; accepted October 31, 2002. The editor coordinating the review of this paper and approving it for publication is A. Yener. This work was supported in part by Stanford Graduate Fellowships and in part by National Semiconductor and Samsung. This work was presented in part at the IEEE Globecom, San Antonio, TX, November 2001.

W. Rhee was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA. He is now with ArrayComm, Inc., San Jose, CA 95131-1014 USA (e-mail: wonjong@dsl.stanford.edu).

W. Yu was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA. He is now with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4 Canada (e-mail: weiyu@comm.utoronto.ca).

J. M. Cioffi is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: cioffi@dsl.stanford.edu).

Digital Object Identifier 10.1109/TWC.2003.819017

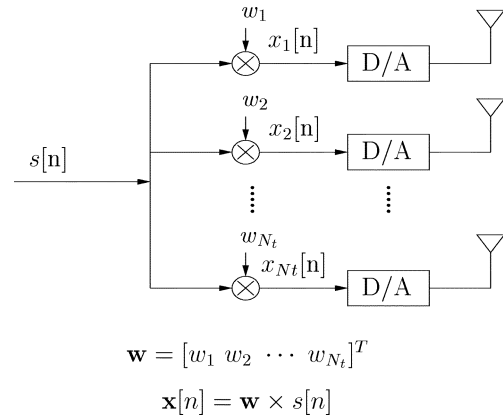
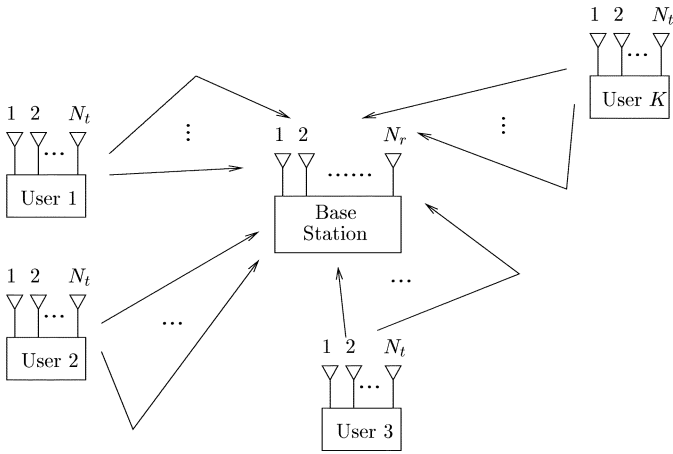


Fig. 1. Beamforming.

literature. For single-user point-to-point systems, space-time coding [3] has been shown to be capable of providing spatial diversity when no channel information is available at the transmitter. When perfect channel knowledge is available, a transmission scheme based on multicarrier modulation and spatial singular-value decomposition (SVD) has been shown to achieve a rate close to the channel capacity [4]. When the channel is unknown but its statistics (such as the covariance information) is available, the conditions for the optimality of beamforming has been derived in [10]. For multiuser cellular systems, the idea of stream control has been examined in [7], in which the authors conclude that not all space dimensions should be utilized by every user if the users mutually interfere. In [8], the use of multiple transmit antennas to artificially induce large and fast channel fluctuations has been considered for a system with no transmitter channel knowledge, and it has been shown to achieve a high level of multiuser diversity. In this paper, we focus on the uplink of a multiuser wireless system (i.e., multiple-access channel [5]) in which the transmitters have (at least partial) channel side information. The goal is to identify an asymptotically optimal transmission strategy when a large number of users simultaneously transmit to a base-station. Interestingly and perhaps surprisingly, the optimal capacity-achieving transmission scheme is beamforming.

Beamforming is a simple transmission scheme: A single coded data stream $s[n]$ is weighted by w_i for transmission at the i th antenna at the time index n . $s[n]$ is assumed to be i.i.d. with unit power. The vector signal $\mathbf{x}[n] = [w_1 \ w_2 \ \dots \ w_{N_t}]^T \times s[n] = \mathbf{w} \times s[n]$ is transmitted simultaneously from the N_t transmit antennas (Fig. 1). Since the weighting vector \mathbf{w} is fixed, the rank of the transmit signal covariance matrix, $\text{rank}(\mathcal{E}[\mathbf{x}[n]\mathbf{x}^*[n]]) = \text{rank}(\mathbf{w}\mathcal{E}[s[n]s^*[n]]\mathbf{w}^*) = \text{rank}(\mathbf{w}\mathbf{w}^*)$,

Fig. 2. K -user multiple-access channel.

is one. On the other hand, if q (≥ 2) independent data streams $\{s^1[n], s^2[n], \dots, s^q[n]\}$ are simultaneously transmitted with linearly independent weightings $\{\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^q\}$ (i.e., $\mathbf{x}[n] = \sum_{i=1}^q \mathbf{w}^i \times s^i[n]$), the rank of the transmit covariance matrix, $\text{rank}(\mathcal{E}[\mathbf{x}[n]\mathbf{x}^*[n]]) = \text{rank}(\sum_{i=1}^q \mathbf{w}^i \mathbf{w}^{i*})$, is $\min(N_t, q)$, and is larger than one. To distinguish this method from beamforming, we call it *space-multiplexing* [9] in this paper. More precisely, space-multiplexing refers to the situation in which $\text{rank}(\mathcal{E}[\mathbf{x}[n]\mathbf{x}^*[n]]) \geq 2$. Between these two transmission methods, beamforming is preferred in terms of complexity. Since only one data stream $s[n]$ is used, coding and transmission can be simply done as in a single antenna system [6]. In practice, beamforming is widely used in multi-antenna applications due to the easiness of adopting multiple antennas to conventional single antenna transmission systems.

However, for a single-user multi-antenna system, beamforming usually cannot achieve a data rate close to the capacity, because it utilizes only one space dimension. The data rate loss from the capacity can be substantial in a rich-scattering environment, and space-multiplexing must be used to prevent such a large loss [1], [2]. However, the same conclusion does not hold in multiuser systems.

Consider a multiple-access channel, where K users each with N_t transmit antennas try to communicate with a common receiver with N_r receive antennas (Fig. 2). Unlike in a single-user situation, K independent data streams are received at the common receiver even if only one data stream is transmitted from each user. Therefore, when K is large, all the space dimensions can be utilized even with each user using beamforming. This suggests that beamforming at the remote terminals may be sufficient for achieving a high data rate in the uplink environment. The main goal of this paper is to make this intuition precise by showing that beamforming is not only sufficient but also necessary for achieving the sum capacity of a multiple-access channel, provided that the number of users is much larger than the number of receive antennas. This result implies that sophisticated space-multiplexing at the remote terminals, which requires multiple space codes per user, is not needed for achieving a high spectral efficiency in a multiuser uplink system, because combining beamforming with a conventional one-dimensional coding is optimal. This

conclusion is derived under the assumptions that the channels are slowly fading and, thus, can be considered fixed for a block of transmission [11], [12], and that the channel coefficients are known (at least partially) at the transmitters.

The rest of the paper is organized as follows. Section II presents a mathematical formulation of the problem for the flat-fading case [intersymbol interference (ISI)-free], and the main result is proved in Section III. For the frequency-selective fading (ISI) case, the use of orthogonal frequency division multiplexing (OFDM) is assumed for the sake of simplicity and clarity of the result. With OFDM, each user's channel can be modeled as M independent flat fading channels, and a similar result as in an ISI-free channel is derived in Section IV. In both flat and frequency-selective fading channels, channel coefficients are assumed to be independently fading with an i.i.d. Gaussian distribution. In Section V, a suboptimal transmission scheme based on partial channel information, which is available in a time division duplex (TDD) system, is introduced. The suboptimal scheme is shown to achieve a data rate close to the sum capacity and to be only mildly sensitive to channel estimation error.

II. PROBLEM FORMULATION FOR ISI-FREE CHANNELS

Consider a K -user synchronous multiple-access channel where an average power constraint p_j is imposed on user j in each block of transmission ($j = 1, \dots, K$). Full channel information is assumed to be available at all transmitters and the receiver for the analytical derivation. The case of partial channel information and the effect of channel estimation error are considered in Section V. The discrete-time base-band equivalent model is as follows:

$$\mathbf{y} = \sum_{j=1}^K H_j \mathbf{x}_j + \mathbf{n} \quad (1)$$

where $\mathbf{y} \in \mathcal{C}^{N_r \times 1}$ is the received signal, $\mathbf{x}_j \in \mathcal{C}^{N_t \times 1}$ is the transmit signal of user j , $\mathbf{n} \in \mathcal{C}^{N_r \times 1}$ is the Gaussian noise at the receiver, and $H_j \in \mathcal{C}^{N_r \times N_t}$ is the channel of user j . Let $Z = \mathcal{E}[\mathbf{n}\mathbf{n}^*]$ be the noise covariance matrix and $S_j = \mathcal{E}[\mathbf{x}_j \mathbf{x}_j^*]$, $j = 1 \dots K$ be the transmit covariance matrices. Then, the sum rate maximization problem with variables $\{S_j\}_{j=1}^K$ can be formulated as follow under the average power constraints $\{p_j\}$ [15]

$$\begin{aligned} & \text{maximize } \log_2 \det \left(\sum_{j=1}^K H_j S_j H_j^* + Z \right) - \log_2 \det(Z) \\ & \text{subject to } \text{Tr}(S_j) \leq p_j \text{ for } j = 1, \dots, K \\ & \quad S_j \geq 0 \text{ for } j = 1, \dots, K. \end{aligned} \quad (2)$$

Note that $\text{Tr}(A)$ denotes the trace of matrix A , and $A \geq 0$ means that matrix A is nonnegative definite. This paper considers an i.i.d. Gaussian channel model, in which elements of H_j are i.i.d. complex Gaussian random variables with mean zero and variance 0.5 in each of real and imaginary dimensions.

The previous formulation does not pose any restriction on the rank of S_j . If the solution of the above optimization problem is such that $\text{rank}(S_j) = 1$, then S_j can be represented by a vector,

i.e., $S_j = \mathbf{w}_j \mathbf{w}_j^*$, and the transmission can be done by beamforming in the \mathbf{w}_j direction. In this case, only a single coded data stream per user is needed to achieve the sum capacity, and a conventional single-antenna based code can be used. However, if the optimal solution is such that $\text{rank}(S_j) > 1$, multiple data streams per user are needed for space-multiplexing, and a multidimensional space-time code with a higher encoding and decoding complexity would be necessary.

III. OPTIMALITY OF BEAM-FORMING IN ISI-FREE CHANNELS

This section shows the optimality of beamforming for ISI-free channels.

Theorem 1: Consider a multiantenna Gaussian multiple-access channel with no ISI. If the number of users K is much larger than the number of receive antennas N_r , then the sum capacity is achieved only if almost all users are using beamforming. More precisely, the optimal transmit strategy is such that at most $N_r(N_r + 1)/2$ users can adopt space-multiplexing.

Proof: The sum rate maximization problem in (2) is a convex optimization problem over the set of nonnegative definite covariance matrices [14], [15]. Therefore, Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality [13]. The KKT conditions are the following (see Appendix A):

$$H_j^* \left(\sum_{i=1}^K H_i S_i H_i^* + Z \right)^{-1} H_j + U_j = \lambda_j I_{N_t} \quad (3)$$

$$\text{Tr}(S_j) = p_j \quad (4)$$

$$\text{Tr}(S_j U_j) = 0 \quad (5)$$

$$S_j, U_j, \lambda_j \geq 0 \quad (6)$$

where λ_j is the dual variable related to the power constraint of user j and U_j is the (N_t, N_t) nonnegative definite matrix related to the non-negative constraint on S_j . These equations need to be satisfied for all $j = 1, 2, \dots, K$. Let $r_j = \text{rank}(S_j)$. The theorem will be proved by deriving an upper bound on $\sum_{j=1}^K r_j$.

Since $\{S_j\}$ and Z are covariance matrices, they are Hermitian. In addition, $\{U_j\}$ is Hermitian due to (3). Therefore, (3) imposes at most $N_t(N_t + 1)/2$ scalar equations for each j , corresponding to the diagonal and upper triangular parts of matrices. The equations corresponding to the lower triangular parts are automatically satisfied as long as the equations corresponding to the upper triangular parts are satisfied. Since the elements of H_j are drawn from an i.i.d. Gaussian distribution, all $N_t(N_t + 1)/2$ equations are independent. With K users with independent channels, a total of $K(N_t(N_t + 1)/2)$ independent equations need to be satisfied to achieve (3) for all K users.

Now, the number of variables in the KKT conditions is counted. Define $C = (\sum_{i=1}^K H_i S_i H_i^* + Z)^{-1}$. Then, from (3)

$$H_j^* C H_j + U_j = \lambda_j I_{N_t}. \quad (7)$$

Note that C is common for all j 's. Since C is an (N_r, N_r) Hermitian matrix, it has at most $N_r(N_r + 1)/2$ degrees of freedom. Also, from K dual variables $\{\lambda_j\}_{j=1}^K$ related to the K power constraints, K degrees of freedom are obtained. Furthermore, the degrees of freedom in $\{U_j\}$ can be shown to

be $\sum_{j=1}^K (N_t(N_t + 1)/2 - r_j(r_j + 1)/2)$ in the following way: Since the rank of S_j is r_j , the maximum rank of U_j is $N_t - r_j$ from (5) and Lemma 1 and 2 in Appendix B. Then, the result directly follows from Lemma 3 in Appendix B.

For the KKT conditions in (3)–(6) to have a valid solution, the total number of degrees of freedom needs to be equal to or larger than the number of independent equations. Therefore, the following inequality is required:

$$\begin{aligned} K \frac{N_t(N_t + 1)}{2} &\leq \frac{N_r(N_r + 1)}{2} + K \\ &\quad + \sum_{j=1}^K \left(\frac{N_t(N_t + 1)}{2} - \frac{r_j(r_j + 1)}{2} \right) \\ \Rightarrow \sum_{j=1}^K \frac{r_j(r_j + 1)}{2} &\leq \frac{N_r(N_r + 1)}{2} + K. \end{aligned} \quad (8)$$

Furthermore, since $\sum_{j=1}^K r_j \leq \sum_{j=1}^K r_j(r_j + 1)/2$, the following can be concluded:

$$\sum_{j=1}^K r_j \leq \frac{N_r(N_r + 1)}{2} + K. \quad (9)$$

On the other hand, it is not difficult to see that $K \leq \sum_{j=1}^K r_j$. First, note that $\text{Tr}(S_j) = p_j$, because any unused power can be used to strictly increase the sum rate. Therefore, $r_j \geq 1$. By summing up over all users, $K \leq \sum_{j=1}^K r_j$ is obtained. Now, from (9) and $K \leq \sum_{j=1}^K r_j$, the following bound is obtained:

$$K \leq \sum_{j=1}^K r_j \leq \frac{N_r(N_r + 1)}{2} + K \quad (10)$$

$$1 \leq r_j \text{ for } j = 1, 2, \dots, K. \quad (11)$$

If $K \gg N_r$, (10) and (11) imply $r_j = 1$ for almost all users. Clearly, the maximum number of users allowed to have $r_j > 1$ is $N_r(N_r + 1)/2$. ■

In a single-user multiantenna system, capacity can be achieved from an SVD of the matrix channel followed by a water-filling on the singular values. This generally results in a transmit covariance matrix whose rank is larger than one. In contrast to the single-user case, Theorem 1 states that almost all users need to transmit rank one signals (beamforming) in order to achieve the sum capacity in a multiuser uplink transmission environment. This holds even when N_t is large. Thus, while the rank of each user's transmit covariance matrix S_j can be up to N_t , it is necessary for almost all users to transmit rank-one signals to achieve the sum capacity. In such case, spatial diversity is fully utilized at the multiuser level and not at the individual user level.

It can be seen from (10) and (11) that each user should use at least one space dimension, and the extra $N_r(N_r + 1)/2$ dimensions can be optionally used by any set of users. For instance, one user could be using the whole extra dimensions, two users could be using half of extra dimensions each, or the extra dimensions might not be used by any user at all. The distribution of the use of the extra space dimensions depends on the channel

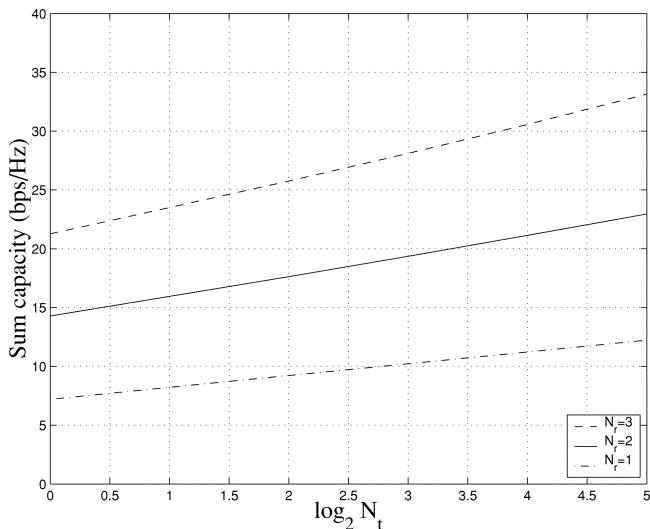


Fig. 3. Sum capacity as a function of $\log_2 N_t$ ($K = 15$, $N_r = 1, 2, 3$, and average SNR of 10 dB per user).

realization, and the optimal choice could be anything as long as (8), (10), and (11) are satisfied.

The assumption of independent channels is crucial for the proof of Theorem 1. As an example, if all users have exactly the same channel, $H_j = H$ for all j , then the total number of equations that needs to be satisfied is fewer than $K(N_t(N_t + 1)/2)$, and the conclusion will not follow.¹ However, Theorem 1 does hold for any realistic channel model that results in $K(N_t(N_t + 1)/2)$ independent equations. In particular, Theorem 1 holds for the i.i.d. Gaussian channel model.

Theorem 1 holds asymptotically as $K \gg N_r$. However, even for a relatively low value of K/N_r , numerical simulation indicates that beamforming is near optimal. As an example, a Monte Carlo simulation shows that when $N_t = 3$, $N_r = 3$, $K = 10$, $Z = I_{N_r}$, and SNR = $p_j = 10$ dB for all users, the probability that $r_j > 1$ is approximately 0.5%.

If K is close to or smaller than N_r , many users need to have transmit covariance matrices whose ranks are larger than one. However, (8) and (10) still provide bounds on the sum rank, from which the value of r_j can be predicted. A similar argument for ISI channels with $K < N_r$ is used in Section V.

In Fig. 3, the average sum capacity of a flat-fading channel is plotted as a function of $\log_2 N_t$ for $K = 15$, $N_r = 1, 2, 3$, $Z = I_{N_r}$, SNR = 10 dB per user, and 100 random channel realizations. As $K \gg N_r$, beamforming is optimal for almost all users. In this case, the sum capacity increases linearly with $\log_2 N_t$. This can be explained as follows. Consider the case with $N_r = 1$. With beamforming at each transmitter, doubling the number of transmit antennas results in a 3-dB gain in the received signal SNR on the average, because the transmitter can transmit more power in the channel direction. A 3-dB power gain is equivalent to one additional bit at high SNR, and thus, doubling N_t gives a capacity gain of 1 bps/Hz. For $N_r > 1$,

¹As long as $K \gg N_r$, there exists a resource allocation scheme achieving sum capacity with almost all users beamforming. In other words, the optimal power allocation scheme is not unique for $H_j = H$, and beamforming is one way to achieve the sum capacity.

the effect of doubling N_t is not so obvious because the power increase corresponding to the N_r by N_t matrix channel is complicated². From simulations, the capacity gain per doubling N_t is observed to be linear with N_r when both N_t and N_r are in a reasonable range. In Fig. 3, the average gains are 1 for $N_r = 1$, 1.78 for $N_r = 2$, and 2.42 b/doubling for $N_r = 3$. In general, it is observed that an exponential increase of N_t is required for a linear capacity increase when $K \gg N_r$.

IV. RANK BOUND FOR OFDM SYSTEMS

In this section, an extension of Theorem 1 to ISI channels is considered assuming OFDM modulation.

A. Extension to OFDM

An OFDM system decomposes the ISI channel into M independent ISI-free subchannels. In the rest of this section, all users are assumed to be synchronized, and channel coefficients of different subchannels are assumed to be independent. If a power spectrum mask constraint applies to each user, the problem can be formulated exactly same as in the ISI-free case for each subchannel, and Theorem 1 directly applies to each subchannel. In this section, a more general power constraint, where each user's total transmit power over all subchannels is limited, is considered. Using superscripts to denote subchannel indexes, the l th subchannel can be modeled as follows:

$$\mathbf{y}^l = \sum_{j=1}^K H_j^l \mathbf{x}_j^l + \mathbf{n}^l. \quad (12)$$

Let Z^l and S_j^l represent the noise covariance matrix in subchannel l and the transmit-covariance matrix of user j in subchannel l , respectively. Then, the sum rate maximization problem for the OFDM system can be formulated as the following:

$$\begin{aligned} & \text{maximize} \quad \sum_{l=1}^M \log_2 \det \left(\sum_{j=1}^K H_j^l S_j^l H_j^{l*} + Z^l \right) - \sum_{l=1}^M \log_2 \det(Z^l) \\ & \text{subject to} \quad \sum_{l=1}^M \text{Tr}(S_j^l) \leq p_j \text{ for } j = 1, \dots, K \\ & \quad \quad \quad S_j^l \geq 0 \text{ for } j = 1, \dots, K \text{ and } l = 1, \dots, M. \end{aligned} \quad (13)$$

The KKT conditions for this problem are (see Appendix A)

$$H_j^{l*} \left(\sum_{i=1}^K H_i^l S_i^l H_i^{l*} + Z^l \right)^{-1} H_j^l + U_j^l = \lambda_j I_{N_t} \quad (14)$$

$$\sum_{l=1}^M \text{Tr}(S_j^l) = p_j \quad (15)$$

$$\text{Tr}(S_j^l U_j^l) = 0 \quad (16)$$

$$S_j^l, U_j^l, \lambda_j \geq 0. \quad (17)$$

²Lower and upper bounds on expectation can be derived from [16] by investigating the distributions of maximum and minimum eigenvalues of Wishart matrices.

TABLE I

MULTIPLE-ANTENNA MULTIPLE-ACCESS CHANNELS. GAUSSIAN CHANNELS WITH $K = 10$, $N_t = 3$, $N_r = 3$, $Z = I_{N_r}$, SNR = $p_j/M = 10$ dB AND $M = 20$ ARE ASSUMED. THE ITERATIVE WATER-FILLING ALGORITHM IN [15] WAS USED TO FIND THE OPTIMAL SPECTRUM, AND ONE MILLION SIMULATIONS WERE PERFORMED TO FIND THE EMPIRICAL DISTRIBUTIONS. (a) A TYPICAL RANK STRUCTURE OF SUM CAPACITY ACHIEVING SPECTRUM. (b) EMPIRICAL PROBABILITY DISTRIBUTION OF r_j^l . (c) EMPIRICAL PROBABILITY DISTRIBUTION OF $\sum_{j=1}^K r_j^l$.

Subchannel(l)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
User 1 (r_1^l)	1	0	0	1	1	0	0	0	0	0	0	1	0	1	0	1	0	1	1	0
User 2 (r_2^l)	1	1	1	0	0	1	1	1	0	0	1	1	0	1	1	1	0	0	1	0
User 3 (r_3^l)	1	0	1	1	1	0	0	1	1	1	0	0	1	0	0	1	1	0	0	0
User 4 (r_4^l)	0	1	1	1	1	0	1	0	1	0	1	0	0	0	0	0	1	1	0	1
User 5 (r_5^l)	1	1	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	1	0
User 6 (r_6^l)	0	1	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	0	0	1
User 7 (r_7^l)	1	0	0	1	1	0	1	0	0	1	1	1	0	0	1	1	0	1	0	0
User 8 (r_8^l)	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	1	1	0	1
User 9 (r_9^l)	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	1	1	0	1	0
User 10 (r_{10}^l)	1	0	1	1	0	1	1	0	0	0	0	1	1	0	1	0	0	1	1	1
$\sum_{j=1}^K r_j^l$	6	5	5	6	6	4	6	5	4	5	5	5	4	4	6	6	5	5	5	4

(a)

$\Pr(r_j^l = 0) = 0.5169$	$\Pr(r_j^l = 1) = 0.4793$	$\Pr(r_j^l = 2) = 0.0039$	$\Pr(r_j^l = 3) = 0$
---------------------------	---------------------------	---------------------------	----------------------

(b)

$\Pr(\sum_{j=1}^K r_j^l \leq 2)$	$\Pr(\sum_{j=1}^K r_j^l = 3)$	$\Pr(\sum_{j=1}^K r_j^l = 4)$	$\Pr(\sum_{j=1}^K r_j^l = 5)$	$\Pr(\sum_{j=1}^K r_j^l = 6)$	$\Pr(\sum_{j=1}^K r_j^l = 7)$	$\Pr(\sum_{j=1}^K r_j^l = 8)$	$\Pr(\sum_{j=1}^K r_j^l = 9)$	$\Pr(\sum_{j=1}^K r_j^l \geq 10)$
0	0.0654	0.3008	0.3875	0.1967	0.0447	0.0048	0.0002	0

(c)

These conditions need to be satisfied for all users ($j = 1, \dots, K$) and all subchannels ($l = 1, \dots, M$) in order to achieve the sum capacity. By defining $r_j^l = \text{rank}(S_j^l)$ and using these KKT conditions in a similar way as in the proof of Theorem 1, the following Theorem 2 can be proved.

Theorem 2: In an OFDM-based multiantenna Gaussian multiple-access channel, where each user has M independent subchannels, the sum capacity is achieved only if the following condition is satisfied in each subchannel l :

$$\sum_{j=1}^K r_j^l \leq \sum_{j=1}^K \frac{r_j^l(r_j^l + 1)}{2} \leq \frac{N_r(N_r + 1)}{2} + a^l \quad (18)$$

where $\{a^l\}$ are auxiliary variables that satisfy $0 \leq a^l$ and $\sum_{l=1}^M a^l \leq K$.

Proof: Considering each subchannel l separately, the goal is to derive an inequality similar to (10). From KKT conditions (14)–(17), all the primal and dual variables except for $\{\lambda_j\}$ can be decoupled for each subchannel l . Since $\{\lambda_j\}$ remain as global variables over all subchannels, it is convenient to introduce auxiliary variables $\{a^l\}$ to represent the number of degrees of freedom used in each subchannel l . There are a total of K degrees of freedom stemming from $\{\lambda_j\}_{j=1}^K$, so $\sum_{l=1}^M a^l \leq K$. Now, for each subchannel l , (18) can be proved by following the same derivations as in (8) and (9) with K replaced by a^l . ■

Unlike the frequency-nonselctive channels considered in the previous section, K in (9) is distributed over M independent subchannels in (18) via auxiliary variables $\{a^l\}_{l=1}^M$. Since each

user's power constraint is over its all subchannels, a total of K dual variables exists in the optimization problem, and the K variables need to be shared among M subchannels. Also, $r_j^l \geq 1$ does not need to hold anymore, because user j does not need to transmit in every subchannel. [Equation (15) can be satisfied as long as user j transmits nonzero power in at least one subchannel.] Therefore, the optimality of beamforming cannot be directly concluded for the OFDM system. However, the upper bound on the sum rank suggests a similar beamforming result as in the ISI-free case.

From (18), at most $N_r(N_r + 1)/2 + a^l$ space dimensions can be used in each subchannel l . Since the objective is to maximize the sum rate, the space dimensions with large channel gains are more likely to be selected among the $N_r(N_r + 1)/2 + a^l$ space dimensions. This is especially true when K is large as the dimension selection process becomes more competitive. Now, assume that the spatial channel gains follow the same probability distribution for every user. If the space-channel gains of all users are considered together, it is unlikely for multiple dimensions of a single user's space dimensions to have large enough gains and appropriate directions to be selected among the $N_r(N_r + 1)/2 + a^l$ dimensions, because channels of all users have the same distribution. Therefore, when K is large, the optimal solution is likely to be many users beamforming rather than a few users space-multiplexing. This can be confirmed from simulations. Table I(a) shows an example of a typical optimal rank structure. In each subchannel, from N_r to $N_r(N_r + 1)/2$ users have the best space dimensions to transmit,

and they transmit rank-one signals. In Table I(b), the empirical distribution of r_j^l is summarized. Note that the probability of a user space-multiplexing for a given subchannel ($= \Pr(r_j^l \geq 2)$) is around 0.39%. In Table I(c), the empirical distribution of $\sum_{j=1}^K r_j^l$ is summarized. In each subchannel, about 4 ~ 6 data streams are used with about 90% probability.

In the extreme case where the number of subchannels, M , goes to infinity, the effect of auxiliary variables $\{a^l\}_{l=1}^M$ becomes negligible. In [17], assuming infinite M , where M represents the number of fading states of a flat-fading multiple-access channel, it was shown that at most $N_r(N_r + 1)/2$ users can simultaneously transmit in each fading state.

B. Single-Antenna Systems

The general rank bound for a multiple-antenna system can now be specialized to the single-antenna case ($N_t = N_r = 1$). By replacing N_r with 1 and by noting that r_j^l is either 0 or 1 for single-antenna cases, (18) becomes the following:

$$\sum_{j=1}^K r_j^l = \sum_{j=1}^K \frac{r_j^l(r_j^l + 1)}{2} \leq 1 + a^l \quad (19)$$

$$\sum_{l=1}^M a^l \leq K, \quad 0 \leq a^l. \quad (20)$$

Therefore, it can be seen that at most K subchannels can have $\sum_{j=1}^K r_j^l \geq 2$, which means that at most K subchannels can be shared by two or more users in a single-antenna system. In fact, the bound can be reduced to $K - 1$ for the single-antenna case.

Theorem 3: For a single-antenna multiple-access channel with K users each with M independent subchannels, the optimal transmit strategy that achieves the sum capacity is such that at most $K - 1$ subchannels can be shared by two or more users. In the limit, the probability of a subchannel being shared by two or more users goes to zero as the number of subchannels M goes to infinity.

Proof: For a single antenna multiple access channel, where all of K users and the base station have only one antenna each, the KKT conditions (14)–(17) can be written as the following:

$$\frac{|h_j^l|^2}{\left(\sum_{i=1}^K |h_i^l|^2 s_i^l + z^l\right)} + u_j^l = \lambda_j \quad (21)$$

$$\sum_{l=1}^M s_j^l = p_j \quad (22)$$

$$s_j^l u_j^l = 0 \quad (23)$$

$$s_j^l, u_j^l, \lambda_j \geq 0. \quad (24)$$

Note that the lower case letters are used for representing scalars. Because of (23), $u_j^l = 0$ if $s_j^l \neq 0$. Therefore

$$\frac{|h_j^l|^2}{\left(\sum_{i=1}^K |h_i^l|^2 s_i^l + z^l\right)} = \lambda_j, \text{ if } s_j^l \neq 0. \quad (25)$$

If there are $L(l)$ users (with indices $j_1, j_2, \dots, j_{L(l)}$) with positive transmit power in subchannel l , the following equations need to hold in order to satisfy (25):

$$\frac{|h_{j_1}^l|^2}{\lambda_{j_1}} = \dots = \frac{|h_{j_{L(l)}}^l|^2}{\lambda_{j_{L(l)}}} = \sum_{i=1}^K |h_i^l|^2 s_i^l + z^l. \quad (26)$$

With probability one, (26) forms $L(l) - 1$ independent equality equations since $\{h_{j_1}^l, h_{j_2}^l, \dots, h_{j_{L(l)}}^l\}$ are random realizations from an i.i.d. Gaussian distribution.

Now, the total number of independent equations needed to satisfy the KKT conditions (21)–(24) for all M subchannels is $\sum_{l=1}^M \max(0, L(l) - 1)$. Note that “max” operation is needed to prevent a subchannel, that is not used by any user, from being negatively counted. Since only K variables $\{\lambda_j\}_1^K$ exist for satisfying (25) for all $s_j^l > 0$, the following inequality can be derived:

$$\sum_{l=1}^M \max(0, L(l) - 1) \leq K.$$

However, if $\sum_{l=1}^M \max(0, L(l) - 1) = K$, the solution for $\{\lambda_j\}_1^K$ is a zero vector which does not allow (25) from having a meaningful solution. Therefore

$$\sum_{l=1}^M \max(0, L(l) - 1) \leq K - 1. \quad (27)$$

From (27), it is clear that the number of subchannels with $L(l) > 1$ is at most $K - 1$.

The probability of a subchannel being shared by 2 or more users is upper bounded by $(K - 1)/M$, because at most $K - 1$ subchannels out of M subchannels can be shared by two or more users. Therefore, assuming a fixed K , the probability converges to 0 as $M \rightarrow \infty$. ■

In the extreme case of $M = \infty$, the sum capacity achieving strategy is an frequency division multiple access (FDMA)-like power allocation. This was derived by Cheng and Verdú in [18], and a similar result was derived by Knopp and Humblet for fading channels in [19]. Theorem 3 is a more general result in the sense that it states the case of $M < \infty$ as well as the case of $M = \infty$.

Recently, orthogonal frequency division multiple access (OFDMA) has received a significant amount of attention due to its inherent ability to utilize multiuser diversity [22], [23]. In an OFDMA system, each subchannel is allocated to a single user whose channel gain for the subchannel is sufficiently large, and the overall spectral efficiency is optimized by carefully assigning subchannels. Theorem 3 implies that such an orthogonal allocation of subchannels can achieve a sum rate that is very close to the sum capacity as $(K - 1)/M \rightarrow 0$.

An example of optimal power allocation that achieves the sum capacity over a single-antenna multiple-access channel is shown in Table II for $K = 5$ users each with a single transmit antenna, a base station with a single receive antenna, $M = 20$ independent subchannels with i.i.d. Gaussian distributed channels, and an average SNR = 10 dB per user ($\text{SNR} = p_j/MZ$). In

TABLE II
SINGLE-ANTENNA MULTIPLE-ACCESS CHANNELS. GAUSSIAN CHANNELS WITH $K = 5$, $N_t = 1$, $N_r = 1$, $Z = 1$ AND $\text{SNR} = p_j/M = 10$ dB ARE ASSUMED. (a) A TYPICAL POWER ALLOCATION ACHIEVING SUM CAPACITY FOR $M = 20$. EACH SUBCHANNEL, EXCEPT FOR THREE SUBCHANNELS, IS EXCLUSIVELY USED BY A SINGLE USER. (b) EMPIRICAL PROBABILITY DISTRIBUTION OF $\sum_{j=1}^K r_j^l$. AS THE NUMBER OF SUBCHANNELS, M , GOES TO ∞ , $\text{Pr}\sum_{j=1}^K r_j^l = 1$ CONVERGES TO 1.

Subchannel	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
User 1 (r_1^l)	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0
User 2 (r_2^l)	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0
User 3 (r_3^l)	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
User 4 (r_4^l)	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0
User 5 (r_5^l)	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0
$\sum_{j=1}^K r_j^l$	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	2	1	1

(a)

M	2	3	5	10	20	50	100
$\text{Pr}(\sum_{j=1}^K r_j^l = 0)$	0	0	0	0	0	0	0
$\text{Pr}(\sum_{j=1}^K r_j^l = 1)$	0.0214	0.1744	0.5600	0.7564	0.8733	0.9436	0.9680
$\text{Pr}(\sum_{j=1}^K r_j^l = 2)$	0.3318	0.6234	0.3934	0.2292	0.1230	0.0556	0.0317
$\text{Pr}(\sum_{j=1}^K r_j^l = 3)$	0.5382	0.1960	0.0450	0.0143	0.0037	0.0008	0.0003
$\text{Pr}(\sum_{j=1}^K r_j^l = 4)$	0.1075	0.0061	0.0016	0.0002	0.0001	0	0
$\text{Pr}(\sum_{j=1}^K r_j^l = 5)$	0.0011	0.0001	0	0	0	0	0

(b)

Table II(a), only three subchannels are shared which is consistent with Theorem 3. It can also be seen from Table II(b) that the probability of a subchannel being shared by two or more users, $\text{Pr}(\sum_{j=1}^K r_j^l \geq 2)$, goes to 0 as $M \rightarrow \infty$.

V. SUBOPTIMAL TRANSMIT ALGORITHM FOR TDD SYSTEMS

In previous sections, perfect channel information was assumed to be available for all users. However, this may not be true in real systems. In this section, a situation where only partial channel information is available is considered as a *toy* problem.

Consider a TDD system. In a TDD system, the uplink channel is strongly correlated with the downlink channel, and the channel estimation for transmission can be reliably done while receiving signals. Thus, assuming perfect channel correlation between uplink and downlink, H_j can be assumed to be available to user j . However, H_i , $i = 1, 2, \dots, j-1, j+1, \dots, K$ are not known to user j because user j does not have access to other users' received signals. For this system, a simple suboptimal transmit algorithm based on the beamforming result is considered in Section V-A. Then, the effect of channel estimation error is considered in Section V-B.

A. Algorithm and Simulation Results

In this TDD scenario, only each user's own channel information is available at each user's transmitter. The receiver is assumed to have access to all channel information. From previous sections, it is clear that the ranks of transmit covariances should

be upper bounded by a function of K , M , and N_r . Motivated by the upper bound result on $\sum_{j=1}^K r_j^l$, the following simple suboptimal algorithm is proposed for the allocation of power and dimension.

Algorithm 1: Suboptimal transmit algorithm: All users independently obey the following transmission scheme. Consider M SVD of M subchannels of user j . Define $N_{\min} = \min(N_t, N_r)$. Then, N_{\min} singular values and corresponding right and left singular vectors (beam vectors) are obtained from the SVD of each subchannel. Find the R largest singular values from the set of MN_{\min} singular values. Transmit R streams (beams) each with power p_j/R to the R right singular vector directions corresponding to the R singular values.

Even though each user does not have access to other users' channel information, cooperation among users is achieved by controlling R in Algorithm 1. Instead of each user greedily trying to use most of dimensions, Algorithm 1 suggests each user use only R dimensions out of MN_{\min} available dimensions. Based on the results in the last section, R could be chosen somewhere between MN_r/K and $MN_r(N_r + 1)/2K + 1$ for the highest data rates. Note that Algorithm 1 is considered only to understand an achievable data rate. In practice, data rate control, coding, and multiuser decoding need to be carefully designed, and performance loss can be incurred by an imperfect implementation.

Monte Carlo simulation results of sum rate as a function of R are shown in Fig. 4. N_t , N_r , and M are fixed at 5, 5, and 10 for both (a) and (b), and channels are randomly generated from an i.i.d. Gaussian distribution. The number of users K is 10 in (a) and 2 in (b). For comparison purpose, the sum capacity and the sum rate that can be achieved with an equal power allocation ($S_j^l = (p_j/MN_t)I_{N_t}$) are plotted together. As discussed before, the sum capacity can be achieved only if each transmitter

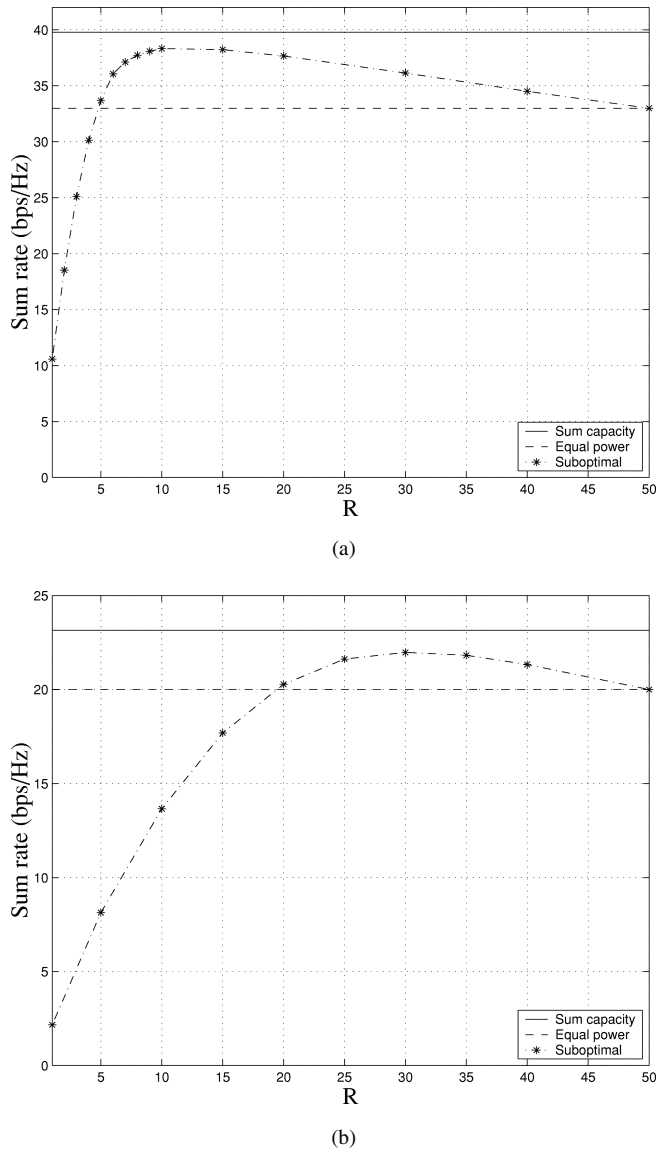


Fig. 4. Sum rates achievable with Algorithm 1 ($N_t = 5$, $N_r = 5$, $M = 10$, and SNR = 10 dB). (a) $K = 10$ (b) $K = 2$.

has access to all user channel information, and it is the maximum achievable data rate of a multiple-access channel. On the other hand, equal power allocation does not require any channel information at any transmitter, and it can be shown to be the optimal power allocation scheme when transmitters do not have access to any channel information [21]. Note that the amount of channel information needed for Algorithm 1 is in between these two extreme cases. As can be seen from Fig. 4, a sum rate close to the sum capacity can be achieved by Algorithm 1 as long as R is chosen appropriately. For $K = 10$ [Fig. 4(a)], selecting $R = 10$ results in a sum rate that is close to the sum capacity. For $K = 2$ [Fig. 4(b)], selecting $R = 30$ results in a sum rate close to the sum capacity. The analytical derivation of the optimal R , however, is a very difficult problem, because the optimal spectrum can be found only through a numerical procedure. Therefore, an extensive simulation is required to find the optimal R for a given system and channel statistics. In general, a larger K leads to a smaller optimal R .

R is a measure of transmission complexity. In general, smaller R means lower complexity, because R is the number of beams that need to be transmitted from each user. Therefore, R data streams are required for each user, and smaller R is preferred to reduce complexity. However, a choice of R much smaller than MN_{\min}/K makes many of the dimensions unused, and a high data rate can no longer be achieved. Note that an equal power allocation means $R = MN_t$, and it has the maximum possible encoding complexity.

B. Effect of Channel Estimation Error

Even for a TDD system, the estimation might not be accurate over a block of transmission due to the thermal noise or the channel variation before the completion of a block transmission. Define the true channel of user j 's l th subchannel as $H_{j,\text{true}}^l$. Each element of $H_{j,\text{true}}^l$ is drawn from an i.i.d. Gaussian distribution with mean zero and unit variance. Now, define the estimated channel information at user j 's transmitter as $H_{j,\text{est}}^l$. Then, the channel estimation error can be calculated as

$$H_{j,\text{error}}^l = H_{j,\text{true}}^l - H_{j,\text{est}}^l. \quad (28)$$

Since transmitter j assumes $H_{j,\text{est}}^l$ to be the true channel, it will find the R best beams based on $H_{j,\text{est}}^l$. As $H_{j,\text{error}}^l$, which is assumed to be independent to the true channel, becomes larger, the beams will be more dependent on $H_{j,\text{error}}^l$. Hence, the R beams will be selected more randomly by Algorithm 1.

Simulation results on the effect of channel estimation error is plotted in Fig. 5. For simulation purpose, each element of $H_{j,\text{error}}^l$ is assumed to have an i.i.d. Gaussian distribution with mean zero and variance σ^2 . For the case in Fig. 5, the sum rate is insensitive to the estimation error unless σ^2 is larger than -10 dB. Even at $\sigma^2 = -5$ dB, the degradation is mild. This can be explained as follows. Even though the variance of estimation error might not be negligible with respect to the average channel gain, there is a good chance of the error being much smaller than the R largest channel gains. Since the suboptimal algorithm utilizes only the R largest channel gains, it does not suffer much from the estimation error. In the extreme, if K is very large and R is very small, the suboptimal algorithm can be very robust to the estimation error, and even rough channel state information can be very useful for achieving high spectral efficiency.

In Fig. 5, channel estimation error is assumed at the transmitters, but not at the receiver. It is well known that the performance degradation can be significant even with small channel estimation error at the receiver: The optimal receiver needs to be able to decode jointly or successively multiple independent codes, and a slight error on channel information can cause a disaster for decoding. As can be seen from Fig. 5, however, the channel estimation error at the transmitters causes a mild loss, and even noisy channel information can be utilized to acquire a reasonable capacity and complexity gain. In summary, Figs. 4 and 5 suggest that the loss of other users' channel information may not cause much loss in capacity (Section V-A) and that the loss of accuracy on each user's own channel information may not cause much loss (Section V-B) either.

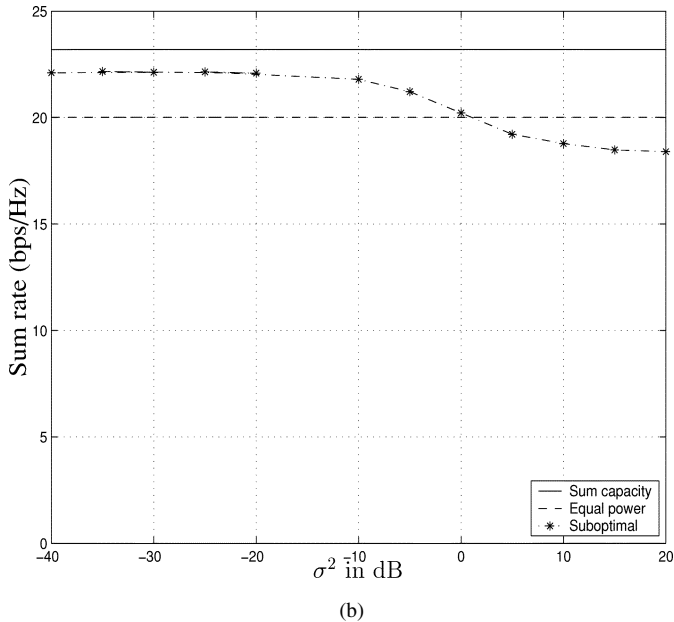
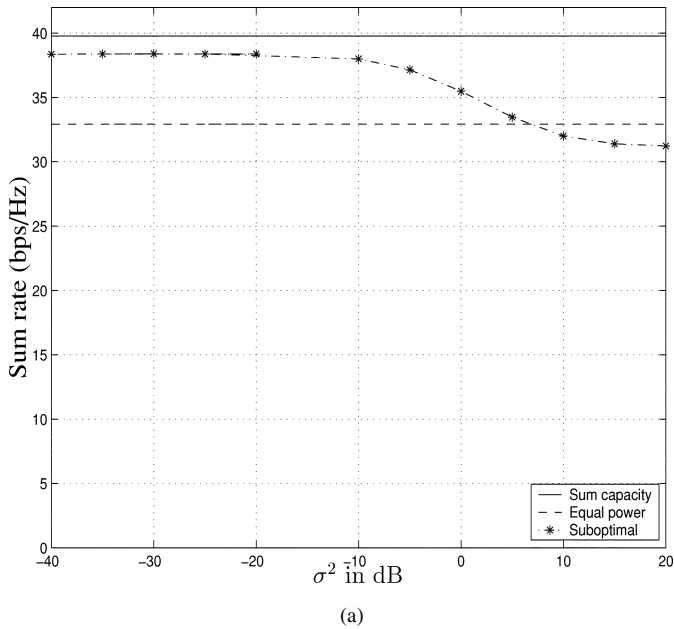


Fig. 5. Sum rates achievable with Algorithm 1 in the presence of channel estimation error ($N_t = 5$, $N_r = 5$, $M = 10$, and SNR = 10 dB). (a) $K = 10$, R is fixed at 10. (b) $K = 2$, R is fixed at 30.

VI. CONCLUSION

This paper investigates the asymptotic optimality of beamforming in synchronous multiantenna Gaussian multiple-access channels. The optimality of beamforming for ISI-free channels is derived assuming a large number of users compared to the number of receive antennas ($K \gg N_r$). This implies that only a logarithmic capacity gain can be achieved with a linear increase in the number of transmit antennas when K is large. For ISI channels, a similar result on the bound of sum rank is obtained assuming an OFDM modulation. As a by-product, Cheng and Verdú's result [18] and Knopp and Humblet's result [19] are extended for single antenna cases.

The optimality of beamforming is interesting mainly for two reasons. First, beamforming can be realized with minimum

complexity, and it allows a simple combination of multiple antennas with the conventional single-antenna transmit systems. Second, even though beamforming is in general far from being optimal in most single-user multiantenna systems, it is asymptotically optimal in multiple-access channels. In this case, the cooperation among users, by means of each user not using too many dimensions, is the key for achieving the sum capacity.

Based on the beamforming and the rank results, a simple sub-optimal transmit algorithm is considered for a situation where each user's transmitter has only its own channel information available. The suboptimal scheme achieves a sum rate that is very close to the sum capacity. Effect of channel estimation error is shown to be small through Monte Carlo simulations, which indicates that even moderate amount of channel information can be very useful for achieving a high data rate.

APPENDIX A KKT CONDITIONS FOR PARALLEL GAUSSIAN MULTIPLE-ACCESS CHANNELS

In this Appendix, the KKT conditions of the maximization problem (13) are shown to be (14)–(17). KKT conditions (3)–(6) for ISI-free channels can then be obtained by setting $M = 1$. A similar proof can be found from [15] for ISI-free case. First, (13) can be reformulated into the following equivalent form:

$$\text{minimize } - \sum_{l=1}^M \log \det(T^l) \quad (29)$$

$$\text{subject to } T^l \leq \sum_{j=1}^K H_j^l S_j^l H_j^{l*} + Z^l \text{ for } l=1, \dots, M \quad (30)$$

$$\sum_{l=1}^M \text{Tr}(S_j^l) \leq p_j \text{ for } j = 1, \dots, K \quad (31)$$

$$S_j^l \geq 0 \text{ for } j = 1, \dots, K \text{ and } l = 1, \dots, M. \quad (32)$$

The \log_2 was replaced by \log ($= \log_e$) without affecting the optimal solution. In addition, the constant $\sum_{l=1}^M \log \det(Z^l)$ for the objective is omitted for simplicity. Variables are $\{T^l\}$ and $\{S_j^l\}$. For each of the constraints in (30)–(32), associate dual variables $\{\Gamma^l\}$, $\{\lambda_j\}$, and $\{U_j^l\}$. Note that the first and the third constraints are matrix inequalities, so the dual variables $\{\Gamma^l\}$ and $\{U_j^l\}$ are matrices and the inner product is the trace of the matrix product. The power constraint is a constraint on real numbers, so its associated dual variable $\{\lambda_j\}$ is real. The Lagrangian of the optimization problem is (33) (shown at the bottom of the next page), where the fact $\text{Tr}(AB) = \text{Tr}(BA)$ is used. The objective of the dual problem is

$$g(\{\Gamma^l\}, \{\lambda_j\}, \{U_j^l\}) = \inf_{\{S_j^l\}, \{T^l\}} L(\{S_j^l\}, \{T^l\}, \{\Gamma^l\}, \{\lambda_j\}, \{U_j^l\}). \quad (34)$$

By differentiating with respect to S_j^l , we have

$$\lambda_j I_{N_t} = H_j^{l*} \Gamma^l H_j^l + U_j^l, \text{ for } j = 1, \dots, K \text{ and } l = 1, \dots, M. \quad (35)$$

Further, by differentiating with respect to T^l , we have

$$\frac{\partial}{\partial T^l} (-\log \det(T^l) + \text{Tr}(\Gamma^l T^l)) = 0 \quad (36)$$

which implies

$$(T^l)^{-1} = \Gamma^l. \quad (37)$$

Therefore

$$\begin{aligned} g(\{\Gamma^l\}, \{\lambda_j\}, \{U_j^l\}) \\ = \sum_{l=1}^M [\log \det(\Gamma^l) + N_r - \text{Tr}(\Gamma^l Z^l)] - \sum_{j=1}^K \lambda_j p_j \end{aligned}$$

where N_r is the number of receive antennas. The dual problem of (13) is then

$$\text{maximize } \sum_{l=1}^M [\log \det(\Gamma^l) + N_r - \text{Tr}(\Gamma^l Z^l)] - \sum_{j=1}^K \lambda_j p_j \quad (38)$$

$$\begin{aligned} \text{subject to } \lambda_j I_{N_t} \geq H_j^{l*} \Gamma^l H_j^l, \text{ for } j = 1, 2, \dots, K \\ \text{and } l = 1, \dots, M \quad (39) \\ \Gamma^l \geq 0, \text{ for } l = 1, \dots, M. \quad (40) \end{aligned}$$

Note that the only constraints on $\{U_j^l\}$ are nonnegative semi-definite constraints, so (35) is equivalent to the inequality in (38). Since the primal problem is convex, the dual problem achieves a maximum at the minimum value of the primal objective.

The KKT conditions are sufficient and necessary for achieving the optimality, because the primal constraints satisfy the Slater's condition [13]. The KKT conditions include the stationarity conditions on the Lagrangian (35) and (37), as well as the complementary slackness conditions

$$\text{Tr} \left[\Gamma^l \left(T^l - \sum_{j=1}^K H_j^l S_j^l H_j^{l*} - Z^l \right) \right] = 0, \text{ for } l = 1, \dots, M \quad (41)$$

$$\begin{aligned} \lambda_j (\text{Tr}(S_j^l) - p_j) = 0, \text{ for } j = 1, \dots, K \\ \text{and } l = 1, \dots, M \quad (42) \end{aligned}$$

$$\text{Tr}(U_j^l S_j^l) = 0, \text{ for } j = 1, \dots, K. \quad (43)$$

Consider the original optimization problem. Observe that at the optimum, we must have $T = \sum_{i=1}^K H_i^l S_i^l H_i^{l*} + Z^l$, and $\text{Tr}(S_j^l) = p_j$, $j = 1, \dots, K$ (otherwise, sum rate can be increased). So, only the last complementary slackness condition (43) is useful. Since the stationary and complementary slackness conditions, together with primal and dual constraints, are necessary and sufficient, the optimization problem can be

transformed into the problem of finding primal variables $\{S_j^l\}$, $\{T^l\}$, and dual variables $\{\Gamma^l\}$, $\{U_j^l\}$, $\{\lambda_j\}$ that satisfy

$$\begin{aligned} \lambda_j I_{N_t} = H_j^{l*} \left(\sum_{j=1}^K H_j^l S_j^l H_j^{l*} + Z^l \right)^{-1} H_j^l + U_j^l \\ \text{for } j = 1, \dots, K \text{ and } l = 1, \dots, M \quad (44) \end{aligned}$$

$$\text{Tr}(S_j^l) = p_j, \text{ for } j = 1, \dots, K \quad (45)$$

$$\text{Tr}(U_j^l S_j^l) = 0, \text{ for } j = 1, \dots, K \text{ and } l = 1, \dots, M \quad (46)$$

$$U_j^l, S_j^l, \lambda_j \geq 0, \text{ for } j = 1, \dots, K \text{ and } l = 1, \dots, M. \quad (47)$$

APPENDIX B LEMMAS

The following lemmas hold for (n, n) Hermitian non-negative matrices S and U .

Lemma 1: If $\text{Tr}(SU) = 0$, then $SU = 0$.

Proof: Since S and U are Hermitian non-negative definite, there exist A and B such that $S = AA^*$, $U = BB^*$. Then

$$\begin{aligned} \text{Tr}(SU) &= \text{Tr}(AA^*BB^*) = \text{Tr}(B^*AA^*B) \\ &= \text{Tr}((A^*B)^*A^*B) \end{aligned}$$

using $\text{Tr}(CD) = \text{Tr}(DC)$. However, $(A^*B)^*A^*B$ is non-negative definite, and $\text{Tr}((A^*B)^*A^*B) = 0$ means A^*B is a zero matrix. Therefore, SU is a zero matrix.

Lemma 2: If $SU = 0$ and $\text{rank}(S) = r$, then maximum rank of U is $n - r$.

Proof: Let the rank of U be p . Then output of $U\mathbf{v}$ spans p -dimensional space for $\{\mathbf{v} | \mathbf{v} \in \mathcal{C}^n\}$. Since input null space of S has dimension $n - r$, there exists a vector \mathbf{v}_0 such that $SU\mathbf{v}_0 \neq 0$ for $p > n - r$. This contradicts $SU = 0$. Therefore, $p \leq n - r$.

Lemma 3: If $\text{rank}(U) = n - r$, maximum number of independent variables in U is $n(n+1)/2 - r(r+1)/2$.

Proof: Consider a singular value decomposition of U , $U = FDL^*$. D is a diagonal matrix with zeros on the last r diagonals. F and L are unitary matrices, and $F = L$ since U is Hermitian. Let $\tilde{F} = FD^{1/2}$. Then, $U = \tilde{F}\tilde{F}^*$. Note that the last r columns of \tilde{F} are zeros. Since the first $n - r$ columns of F are orthogonal, the first $n - r$ columns of \tilde{F} are also orthogonal. Because of the orthogonality between the second and the first column, at least one element of second column needs to be dependent on first column. Because of the orthogonality between the third and the first two columns, at least two elements of third column need to be dependent on

$$\begin{aligned} L(\{S_j^l\}, \{T^l\}, \{\Gamma^l\}, \{\lambda_j\}, \{U_j^l\}) \\ = \sum_{l=1}^M \left[-\log \det(T^l) + \text{Tr} \left(\Gamma^l (T^l - \sum_{j=1}^K H_j^l S_j^l H_j^{l*} - Z^l) \right) - \sum_{j=1}^K \text{Tr}(U_j^l S_j^l) \right] + \sum_{j=1}^K \lambda_j \left(\sum_{l=1}^M \text{Tr}(S_j^l) - p_j \right) \\ = \sum_{l=1}^M \left[-\log \det(T^l) + \text{Tr}(\Gamma^l T^l) - \text{Tr}(\Gamma^l Z^l) + \sum_{j=1}^K \text{Tr}[(\lambda_j I_{N_t} - H_j^{l*} \Gamma^l H_j^l - U_j^l) S_j^l] \right] - \sum_{j=1}^K \lambda_j p_j \quad (33) \end{aligned}$$

first two columns. In general, for c th column, at least $c - 1$ elements need to be dependent on first $c - 1$ columns to have orthogonality between c th column and the first $c - 1$ columns. Therefore, maximum number of independent variables in U is $n + (n - 1) + \dots + (n - (n - r - 1)) = n(n + 1)/2 - r(r + 1)/2$.

ACKNOWLEDGMENT

The authors would like to thank the editor of this paper for the careful reading and suggestions.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
- [2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.
- [3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [4] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [6] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418–2433, Nov. 1999.
- [7] R. S. Blum, J. H. Winters, and N. R. Sollenberger, "On the capacity of cellular systems with MIMO," in *Proc. IEEE Vehicular Technology Conf.*, Atlantic City, NJ, 2001, pp. 1220–1224.
- [8] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.
- [9] R. W. Heath, Jr., S. Sandhu, and A. J. Paulraj, "Antenna selection for spatial multiplexing with linear receivers," *IEEE Commun. Lett.*, vol. 5, pp. 142–144, Apr. 2001.
- [10] S. A. Jafar, S. Vishwanath, and A. J. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Int. Conf. Communications*, Helsinki, Finland, Jun. 2001.
- [11] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.
- [12] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273–1289, May 2001.
- [13] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
- [14] S. Wu and S. Boyd, "Design and implementation of a parser/solver for SDP's with matrix structure," in *Proc. IEEE Int. Symp. Computer-Aided Control System Design*, 1996, pp. 240–245.
- [15] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," in *Proc. Int. Symp. Information Theory (ISIT)*, Washington, D.C., 2001, p. 322.
- [16] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Dept. Math., Mass. Inst. Technol., Cambridge, MA, 1989.
- [17] W. Yu, W. Rhee, and J. M. Cioffi, "Optimal power control in multiple access fading channels with multiple antennas," in *Proc. Int. Conf. Communications*, June 2001, pp. 575–579.
- [18] R. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. 39, pp. 773–785, May 1993.
- [19] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. Int. Conf. Communications*, Seattle, WA, June 1995, pp. 331–335.
- [20] W. Rhee and J. M. Cioffi, "On the asymptotic optimality of beamforming in multi-antenna Gaussian multiple access channels," in *Proc. IEEE Globecom Conf.*, San Antonio, TX, Nov. 2001, pp. 891–895.
- [21] —, "Ergodic capacity of multi-antenna Gaussian multiple access channels," in *Proc. 35th Asilomar Conf. Signals, Systems Computers*, Asilomar, CA, Nov. 2001, pp. 507–512.
- [22] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.
- [23] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. IEEE Vehicular Technology Conf.*, Tokyo, Japan, May 2000.



Wonjong Rhee (M'02) received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 1996 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively. Since 2001, he has been with ArrayComm, Inc., San Jose, CA. His research interests are wireless communications with emphases on multiple antennas and multiuser diversity, design of cellular systems, joint design of physical and MAC layers, and multiuser information theory.



access networks.

Wei Yu (M'02) received the B.A.Sc. degree in computer engineering and mathematics from the University of Waterloo, Waterloo, ON, Canada, in 1997 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively.

Since 2002, he has been an Assistant Professor with the Electrical and Computer Engineering Department, the University of Toronto, Toronto, ON. His research interests include multiuser information theory, wireless communications, and broadband



Dr. Cioffi is currently on the Board of Directors of Marvell, ItEX, Teknovus, Clariphy, and Teranetics. He is on the Advisory Boards of Actelis, Ikanos, Charter Ventures, and Portview Ventures and is a member of the U.S. National Research Council's CSTB. He won the IEEE Communications Magazine Best Paper Award in 1991, the ANSI TI Outstanding Achievement Award in 1995, the University of Illinois Outstanding Alumnus Award in 1999, the IEEE Millennium Medal and the IEE J.J. Tomson Medal in 2000, and the IEEE Kobayashi Medal in 2001. He was an NSF Presidential Investigator from 1987 to 1992 and the Holder of the Hitachi America Professorship in Electrical Engineering in 2002. He became a Member of the National Academy of Engineering in 2001.

John M. Cioffi (S'81–M'84–SM'90–F'96) received the B.S. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 1978 and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1984.

He was with Bell Laboratories, from 1978 to 1984 and IBM Research, from 1984 to 1986. Since 1986, he has been a Professor of electrical engineering at Stanford University. In 1991, he founded Amati Com. Corporation, (which was purchased by Texas Instruments in 1997) and was the company's Officer/Director from 1991 to 1997. He has published over 200 papers and holds over 40 patents. His research interests include the area of high-performance digital transmission.

Dr. Cioffi is currently on the Board of Directors of Marvell, ItEX, Teknovus, Clariphy, and Teranetics. He is on the Advisory Boards of Actelis, Ikanos, Charter Ventures, and Portview Ventures and is a member of the U.S. National Research Council's CSTB. He won the IEEE Communications Magazine Best Paper Award in 1991, the ANSI TI Outstanding Achievement Award in 1995, the University of Illinois Outstanding Alumnus Award in 1999, the IEEE Millennium Medal and the IEE J.J. Tomson Medal in 2000, and the IEEE Kobayashi Medal in 2001. He was an NSF Presidential Investigator from 1987 to 1992 and the Holder of the Hitachi America Professorship in Electrical Engineering in 2002. He became a Member of the National Academy of Engineering in 2001.