Low Complexity User Selection Algorithms for Multiuser MIMO Systems with Block Diagonalization

Zukang Shen, *Student Member, IEEE,* Runhua Chen, *Student Member, IEEE,*
Jeffrey G. Andrews, *Member, IEEE,* Robert W. Heath, Jr. *Member, IEEE,* and
Brian L. Evans, *Senior Member, IEEE*

Abstract

Block Diagonaliation (BD) is a precoding technique that eliminates inter-user interference in downlink multiuser multiple-input-multiple-output (MIMO) systems. The number of simultaneously supportable users with BD is limited by the number of transmit and receive antennas. In a downlink MIMO system with a large number of users, the basestation may select a subset of users to serve in order to maximize the total throughput. The brute-force search for the optimal user set, however, is computationally prohibitive. We propose two low complexity suboptimal user selection algorithms for multiuser MIMO systems with BD. Both algorithms aim to select a subset of users such that the total throughput is nearly maximized. The first user selection algorithm greedily maximizes the total throughput, whereas the criterion of the second algorithm is based on the channel energy. We show that both algorithms have linear complexity in the number of users and achieve around 95% of the total throughput of the complete search method in simulations.

Index Terms

Downlink, Multiuser, MIMO, Sum Capacity, Frobenius Norm

Submitted to *IEEE Transactions on Signal Processing* as a Correspondence. EDICS: MSP-MULT: MIMO Multiuser and Multi-Access Schemes.

The authors are with the Wireless Networking and Communications Group in the Department of Electrical and Computer Engineering at The University of Texas at Austin, Austin, TX 78712 USA. E-mail: {shen, rhchen, jandrews, rheath, bevans}@ece.utexas.edu. Z. Shen was supported by Texas Instruments. R. Chen is supported by SBC Laboratories. All authors were supported by an equipment donation from Intel.

I. INTRODUCTION

Multiple-input-multiple-output (MIMO) systems have drawn a lot of attention in the past decade. A pioneering paper on point-to-point MIMO channel capacity by Telatar [1] showed that the MIMO channel capacity scales linearly with the minimum number of transmit and receive antennas in Rayleigh fading channels. For Gaussian broadcast multiuser MIMO channels, it was conjectured in [2] and recently proven in [3] that Dirty Paper Coding (DPC) [4] can achieve the capacity region. The sum capacity in a multiuser broadcast channel is defined to be the maximum aggregation of all users' data rates. Although DPC can achieve the sum capacity [2], deploying DPC in real-time systems is impractical because of the complicated encoding and decoding schemes [5]. An alternative and more practical precoding technique for downlink broadcast MIMO channels is Block Diagonalization (BD) [6]–[10]. With BD, each user's signal is multiplied by a precoding matrix before transmission. Every user's precoding matrix is restricted to be in the null space of all other users' channels. Hence if the channel matrices of all users are perfectly known at the transmitter, each user perceives an interference-free channel. On the other hand, BD is inferior in terms of sum capacity relative to DPC, since the users' transmit signal covariance matrices are generally not optimal.

Due to the rank condition imposed by the fact that each user's precoding matrix must lie in the null space of all other users' channels, the number of users that can be simultaneously supported with BD is limited by the number of transmit antennas, the number of receive antennas, and the richness of the channels [6]. In this paper, we consider the problem of choosing a subset of users that maximize the total throughput (defined as the aggregate error-free capacity) for a multiuser system with a large number of users. We assume that every user utilizes all his receive antennas. A brute-force complete search over all possible user sets guarantees that the total throughput is maximized. The complexity, however, is prohibitive if the number of users in the system is large. For example, if $\hat{K}$ is the maximum number of users that can be simultaneously supported by BD and K is the total number of users, then the complete search for the optimal user set has combinatoric complexity because every i ($1 \leq i \leq \hat{K}$) out of K users must searched.

A user selection algorithm for downlink multiuser MISO systems has been proposed in [11], where the users are equipped with one receive antenna and zero-forcing beamforming is performed at the transmitter, which is equivalent to BD. The algorithm in [11] constructs a set of semi-orthogonal users whose total throughput is close the sum capacity achieved by DPC. Analogous to the user selection problem is the antenna selection problem where the transmitter and receiver select a subset of antennas to transmit and receive signals. A low complexity antenna selection algorithm is proposed in [12] that achieves almost the same outage capacity as the optimal selection method. Antenna selection has also been considered in downlink multiuser MIMO systems with BD [13], where it has been shown that a significant reduction in symbol error rate can be achieved even with one extra transmit antenna.

In this paper, we propose two suboptimal user selection algorithms for BD with the aim of maximizing the total throughput while keeping the complexity low. Both algorithms iteratively select users until the maximum number of simultaneously supportable users are reached. The first user selection algorithm greedily maximizes the total throughput. In each user selection step, the algorithm selects a user who provides the maximum total throughput with those already selected users. While the first algorithm requires frequent singular value decomposition (SVD) of the channel matrices, the second proposed algorithm selects the users based on the channel energy, thus reducing the computational complexity. We show that the proposed algorithms achieve around 95% of the total throughput with the optimal user set, and the complexity of the proposed algorithms is linear in the total number of users K.

II. System Model and Background

In this section, we introduce the system model and briefly describe the block diagonalization method for multiuser MIMO systems presented in [6] [7]. In a downlink multiuser MIMO system with K users, we denote the number of transmit antenna at the base station as n_t and the number of receive antennas for the jth user as $n_{r,j}$. The transmitted symbol of user j is denoted as a N_j-dimensional vector $\mathbf{x}_j$, which is multiplied by a $n_t \times N_j$ precoding matrix $\mathbf{T}_j$ and sent to the basestation antenna array.

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

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

where the second item in the right-hand-side (RHS) of (1) is the interference seen by user j from other users' signals and $\mathbf{v}_j$ denotes the Additive Gaussian White Noise (AWGN) vector for user j with variance $E[\mathbf{v}_j\mathbf{v}_j^*] = \sigma^2\mathbf{I}$. Matrix $\mathbf{H}_j \in \mathbb{C}^{n_{r,j}\times n_t}$ denotes the channel transfer matrix from the basestation to the jth user, with each entry following an i.i.d. complex Gaussian distribution $\mathcal{CN}(0,1)$ [1], which is a valid channel model if the transmit and receive antennas are in rich-scattering environments and the antenna spacing is larger than the coherence distance. Other non-physical and physical MIMO channel models can be found in [14]. For analytical simplicity, we assume that $\text{rank}(\mathbf{H}_j) = \min(n_{r,j}, n_t)$ for all users. It is also assumed that the channels $\mathbf{H}_j$ experienced by different users are independent. The key idea of block diagonalization is to precode each user's data $\mathbf{x}_j$ with the precoding matrix $\mathbf{T}_j$, such that

$$\begin{aligned} &\mathbf{T}_j \in \mathbb{U}(n_t, N_j) \\ &\mathbf{H}_i\mathbf{T}_j = 0 \quad \text{for all } i \neq j \ \text{ and } 1 \leq i, j \leq K, \end{aligned} \tag{2}$$

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

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

$$\begin{aligned} \mathbf{y}_j &= \mathbf{H}_j\mathbf{T}_j\mathbf{x}_j + \sum_{k=1,k\neq j}^{K} \mathbf{H}_j\mathbf{T}_k\mathbf{x}_k + \mathbf{v}_j \\ &= \mathbf{H}_j\mathbf{T}_j\mathbf{x}_j + \mathbf{v}_j. \end{aligned} \tag{3}$$

Let $\widetilde{\mathbf{H}}_j = [\mathbf{H}_1^T \ \cdots \ \mathbf{H}_{j-1}^T \ \mathbf{H}_{j+1}^T \ \cdots \ \mathbf{H}_K^T]^T$. In order to satisfy the constraint in (2), $\mathbf{T}_j$ shall be in the null space of $\widetilde{\mathbf{H}}_j$. Let $\widetilde{N}_j$ denote the rank of $\widetilde{\mathbf{H}}_j$. Let the singular value decomposition of $\widetilde{\mathbf{H}}_j$ be $\widetilde{\mathbf{H}}_j = \widetilde{\mathbf{U}}_j\widetilde{\mathbf{\Lambda}}_j[\widetilde{\mathbf{V}}_j^1 \ \widetilde{\mathbf{V}}_j^0]^*$, where $\widetilde{\mathbf{V}}_j^1$ contains the first $\widetilde{N}_j$ right singular vectors and $\widetilde{\mathbf{V}}_j^0$ contains the last $(n_t - \widetilde{N}_j)$ right singular vectors of $\widetilde{\mathbf{H}}_j$. The columns in $\widetilde{\mathbf{V}}_j^0$ form a basis set in the null space of $\widetilde{\mathbf{H}}_j$, and hence the columns in $\mathbf{T}_j$ are linear combinations of those in $\widetilde{\mathbf{V}}_j^0$.

In the rest of the paper, we assume that every user has and uses the same number of receive antennas, i.e. $\{n_{r,j}\}_{j=1}^K = n_r$ for simplicity. With the assumption that each element in $\mathbf{H}_j$ is generated by an i.i.d. complex Gaussian distribution, it can be inferred from the rank condition in [6] that the maximum number of simultaneous users is $\left\lceil \frac{n_t}{n_r} \right\rceil$, where $\lceil \cdot \rceil$ is the ceiling operation.

III. Low Complexity User Selection Algorithms

In this section, we first define the sum capacity (i.e. the maximum total throughput) of BD for multiuser broadcast channels. Two suboptimal user selection algorithms are then proposed to reduce the complexity of finding the optimal user set.

Consider a set of channels $\{\mathbf{H}_j\}_{j=1}^K$ for a multiuser MIMO system. Let $\mathcal{K} = \{1, 2, \cdots, K\}$ denote the set of all users, and $\mathcal{A}_i$ be a subset of $\mathcal{K}$, where the cardinality of $\mathcal{A}_i$ is less than or equal to the maximum number of simultaneous users $\hat{K}$. Let $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ denote the effective channel after precoding for user $j \in \mathcal{A}_i$, then the total throughput achieved with BD applied to the user set $\mathcal{A}_i$ with total power P can be expressed as

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

where $\mathbf{Q}_j = E[\mathbf{x}_j\mathbf{x}_j^*]$ is user j's input covariance matrix of size $N_j \times N_j$ and $\mathbf{H}_{\mathcal{A}_i}$ denotes the set of channels for those users in $\mathcal{A}_i$. Notice that the solution to the RHS of (4) can be obtained by the water-filling algorithm over the eigenvalues of $\{\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*\}_{j\in\mathcal{A}_i}$ with total power constraint P, as discussed in [6].

Let $\mathcal{A}$ be the set containing all possible $\mathcal{A}_i$, i.e. $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots\}$, then the sum capacity (maximum total throughput) with BD can be defined as

$$C_{BD}(\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_K, P, \sigma^2) = \max_{\mathcal{A}_i\in\mathcal{A}} C_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2). \tag{5}$$

Denote $\hat{K} = \left\lceil \frac{n_t}{n_r} \right\rceil$ as the maximum number of simultaneous users, and the Cardinality of $\mathcal{A}$ is $|\mathcal{A}| = \sum_{i=1}^{\hat{K}} {}_K C_i$, where ${}_n C_m$ denotes the combination of n choosing m. Hence, it is clear that a brute-force exhaustive search over $\mathcal{A}$ is computationally prohibitive if $K \gg \hat{K}$.

A. Capacity-Based Suboptimal User Selection Algorithm

The exhaustive search method needs to consider roughly $\mathcal{O}(K^{\hat{K}})$ possible user sets. In this section, we present a suboptimal algorithm whose complexity is $\mathcal{O}(\hat{K}K)$.

Let s_i denote the user index selected in the ith iteration, i.e. $s_i \in \{1, 2, \cdots, K\}$ and $1 \leq i \leq \hat{K}$. Let Ω denote the set of unselected users and Υ denote the set of selected users. The capacity-based user selection algorithm is described in Table I. In words, the algorithm first selects the single user with the highest capacity. Then, from the remaining unselected users, it finds the user that provides the highest total throughput together with those selected users. The algorithm terminates when $\hat{K}$ users are selected or the total throughput drops if more users are selected (the total throughput may decrease with an additional user because the size of the null space for every user reduces in order to meet the zero inter-user interference requirement). Clearly, the proposed algorithm needs to search over no more than $\hat{K}K$ user sets, which greatly reduces the complexity compared to the exhaustive search method. Since the user selection criterion is based on the sum capacity, we name the above algorithm the capacity-based suboptimal user selection algorithm. Its throughput performance will be shown in Section V.

B. Frobenius Norm-Based Suboptimal User Selection Algorithm

Although the capacity-based suboptimal user selection algorithm greatly reduces the size of the search set, the algorithm still may not be cost-effective for real-time implementation because singular value decomposition, which is computationally intensive, is required for each user in each iteration to find the total throughput. In this section, we propose another suboptimal user selection algorithm which is based on channel Frobenius norm. The motivation is that the capacity is closely related to eigenvalues of the effective channel after precoding. Although the channel Frobenius norm cannot characterize the capacity completely, it is related to the capacity because the Frobenius norm indicates the overall energy of the channel, i.e. the sum of the eigenvalues of $\mathbf{HH}^*$ equals $||\mathbf{H}||_F^2$.

Let s_i denotes the user index selected in the ith iteration, i.e. $s_i \in \{1, 2, \cdots, K\}$ and $1 \leq i \leq \hat{K}$. Let Ω denote the set of unselected users and Υ denote the set of selected users. Let $\mathbf{V}_k$ be the basis for the row vector space of $\mathbf{H}_k$ after applying the Gram-Schmidt orthogonalization procedure to the rows of $\mathbf{H}_k$. The Frobenius norm-based user selection algorithm is described in Table II. The idea of the norm-based user selection algorithm is to select the set of users such that the sum of the effective channel energy of those selected users is as large as possible. Notice that steps 1 and 2 in the norm-based algorithm are independent with SNR, i.e. P. Once the $\hat{K}$ users are selected, step 3 makes the final user selection (possibly a subset of the $\hat{K}$ users chosen by steps 1 and 2) with the capacity-based algorithm, where the SNR is taken into consideration. Clearly, the norm-based algorithm requires fewer SVD operations than the capacity-based algorithm. Detailed computational complexity will be analyzed in Section IV.

IV. Computational Complexity Analysis

Since the primary motivation for the two proposed suboptimal algorithm is their reduced computational complexity, in the section we quantify their complexity and compare with the brute-force approach. The complexity is counted as the number of flops, denoted as ψ. A flop is defined to be a real floating point operation [17]. A real addition, multiplication, or division is counted as one flop. A complex addition and multiplication have two flops and six flops, respectively. Although flop counting cannot characterize the true computational complexity, it captures the order of the computation load, so suffices for the purpose of the complexity analysis in this paper.

A. Complexity of Typical Matrix Operations

For an $m \times n$ complex-valued matrix $\mathbf{H} \in \mathbb{C}^{m \times n}$, we first provide the flop count of several matrix operations that are frequently used in the suboptimal user selection algorithm. We assume $K \gg \hat{K}$, $\hat{K} n_r \approx n_t$, and $m \leq n$ in this section.

- Frobenius norm $||\mathbf{H}||_F^2$ takes $2mn$ real multiplications and $2mn$ real additions, hence the flop count is $4mn$.
- Gram-Schmidt orthogonalization GSO($\mathbf{H}$) takes $4m^2n - 2mn$ real multiplications; $4m^2n - 2mn$ real additions; and $2mn$ real divisions. The flop count for GSO is $8m^2n - 2mn$.

- Water-filling over n eigenmodes takes up to $\frac{1}{2}(n^2+3n)$ real multiplication; n^2+3n real additions; and $\frac{1}{2}(n^2+3n)$ real divisions. The flop count for water-filling is $2n^2+6n$.
- The flop count for SVD of real-valued $m \times n$ ($m \geq n$) matrices is $4m^2n+8mn^2+9n^3$ [17]. For complex-valued $m \times n$ ($m \leq n$) matrices, we approximate the flop count as $24mn^2+48m^2n+54m^3$ by treating every operation as complex multiplication.

B. Suboptimal User Selection Algorithm I: Capacity-Based Approach

1) $i=1$: SVD of $\mathbf{H}_k$ has $48n_r^2n_t+24n_rn_t^2+54n_r^3$ flops, water-filling needs $2n_r^2+6n_r$ flops, and the calculation of total throughput requires $2n_r$ flops. In total, step 1 has computational complexity $K\left(48n_r^2n_t+24n_rn_t^2+54n_r^3+2n_r^2+8n_r\right)$.
2) $i \geq 2$:
   For each $k \in \Omega$, to get $\mathbf{T}_k$ by SVD needs $48(i-1)^2n_r^2n_t+24(i-1)n_rn_t^2+54(i-1)^3n_r^3$ flops. To compute $\overline{\mathbf{H}}_k=\mathbf{H}_k\mathbf{T}_k$, the complexity of this multiplication is $8n_tn_r(n_t-(i-1)n_r)$. SVD of $\overline{\mathbf{H}}_k$ introduces $48n_r^2\left(n_t-(i-1)n_r\right)+24n_r\left(n_t-(i-1)n_r\right)^2+54n_r^3$ flops. Water-filling needs $2in_r(in_r+3)$ flops, whereas the total throughput calculation has complexity $2in_r$.

Hence, the flop count of the capacity-based user selection algorithm is

$$\begin{aligned}\psi_c &\overset{(a)}{<} \sum_{i=2}^{\left\lceil \frac{n_t}{n_r}\right\rceil} \left\{\left[48i(i-1)^2+48i\right]n_r^2n_t\right. \\ &\quad +\left[24i(i-1)+32i\right]n_rn_t^2+\left(54i(i-1)^3+54i\right)n_r^3 \\ &\quad \left.+2i^2n_r^2+8in_r\right\}\times(K-i+1) \\ &\quad +K\left(48n_r^2n_t+24n_rn_t^2+54n_r^3+2n_r^2+8n_r\right) \\ &\approx \mathcal{O}\left(K\left\lceil\frac{n_t}{n_r}\right\rceil^5 n_r^3\right)\approx \mathcal{O}\left(K\left\lceil\frac{n_t}{n_r}\right\rceil^2 n_t^3\right), \end{aligned} \tag{6}$$

where the inequality in (a) is due to the upper bound of $(n_t-(i-1)n_r)$ by n_t in the calculation of $\overline{\mathbf{H}}_k$ and the SVD of $\overline{\mathbf{H}}_k$.

C. Suboptimal User Selection Algorithm II: Frobenius Norm Approach

1) $i=1$: The Frobenius norm of K users needs $4Kn_rn_t$ flop counts.
2) $i \geq 2$.
   For each $k \in \Omega$, we need $18(i-1)n_r^2n_t$ flops for $\widetilde{\mathbf{H}}_k=\mathbf{H}_k-\mathbf{H}_k\mathbf{V}^*\mathbf{V}$, which include the flops for both matrix multiplications and additions; $8(i-1)^2n_r^2n_t-2(i-1)n_rn_t$ flops for $\mathbf{W}_{s_j,k}$; $18(i-1)n_r^2n_t+4(i-1)n_rn_t$ flops for $\|\mathbf{H}_s-\mathbf{H}_s\mathbf{W}_{s,k}^*\mathbf{W}_{s,k}\|_F^2$; and $4(i-1)n_rn_t$ flops for $\|\widetilde{\mathbf{H}}_k\|_F^2$.

3) The complexity of applying the capacity-based algorithm to the selected $\hat{K} = \left\lceil \frac{n_t}{n_r} \right\rceil$ users is $\mathcal{O}\left\{ \left\lceil \frac{n_t}{n_r} \right\rceil^3 n_t^3 \right\}$, which is not dependent on K, and hence negligible compared to the complexity of steps 1 and 2.

Therefore, the total flops of the norm-based user selection algorithm is

$$\begin{aligned} \psi_n &\approx \sum_{i=2}^{\left\lceil \frac{n_t}{n_r} \right\rceil} \left\{ \left[8(i-1)^3 + 18(i-1)^2 + 18(i-1) \right] n_r^2 n_t + \right. \\ &\quad \left. [2(i-1)^2 + 4(i-1)] n_r n_t \right\} \times (K - i + 1) + 4K n_r n_t \\ &\approx \mathcal{O}\left(K \left\lceil \frac{n_t}{n_r} \right\rceil^4 n_r^2 n_t \right) \approx \mathcal{O}\left(K \left\lceil \frac{n_t}{n_r} \right\rceil^2 n_t^3 \right). \end{aligned} \quad (7)$$

D. Optimal User Selection Algorithm: Complete Search

In the optimal user selection algorithm, the base-station conducts an exhaustive search over the $\sum_{i=1}^{\left\lceil \frac{n_t}{n_r} \right\rceil} {}_K C_i$ possible user sets. The complexity of this complete search method is

$$\begin{aligned} \psi_{cs} &\overset{(a)}{\geq} {}_K C_{\left\lceil \frac{n_t}{n_r} \right\rceil} \left\lceil \frac{n_t}{n_r} \right\rceil \left[\left(48 \left(\left\lceil \frac{n_t}{n_r} \right\rceil - 1 \right)^2 + 8 \right) n_r^2 n_t + 24 \left(\left\lceil \frac{n_t}{n_r} \right\rceil - 1 \right) n_r n_t^2 \right. \\ &\quad \left. + \left(54 \left(\left\lceil \frac{n_t}{n_r} \right\rceil - 1 \right)^3 + 2 \left\lceil \frac{n_t}{n_r} \right\rceil^2 + 126 \right) n_r^3 + 8 \left\lceil \frac{n_t}{n_r} \right\rceil n_r \right] \\ &\approx \mathcal{O}\left({}_K C_{\left\lceil \frac{n_t}{n_r} \right\rceil} \left\lceil \frac{n_t}{n_r} \right\rceil n_t^3 \right) \end{aligned} \quad (8)$$

where the inequality in (a) holds because only the case of picking $\hat{K} = \left\lceil \frac{n_t}{n_r} \right\rceil$ out of K users is considered to simplify the complexity analysis.

In summary, the proposed two suboptimal user selection algorithms have only a fraction of the complexity of the complete search method approximately equal to

$$\eta \approx \frac{K \left\lceil \frac{n_t}{n_r} \right\rceil}{{}_K C_{\left\lceil \frac{n_t}{n_r} \right\rceil}}. \quad (9)$$

Both the capacity-based and the Frobenius norm-based algorithms have linear complexity with K, because no more than $\hat{K}K$ user sets need to be searched over. The norm-based algorithm has slightly lower complexity than the capacity-based one because SVD is less frequently used in the norm-based algorithm. In our Matlab 7.0 implementation of the two proposed suboptimal algorithms, we observed that both algorithms take tens to hundreds of milliseconds (on a Pentium M 1.6 GHz PC) to select a user set, and the CPU run time is linear in the number of users. Further, the norm-based algorithm runs roughly two times faster than the capacity-based algorithm, for systems with a large number of users.

V. Simulation Results

In this section, we compare the performance of the following algorithms:

- iterative water-filling for dirty paper coding [15] (DPC),
- optimal user selection by complete search (BD Optimal),
- capacity-based user selection algorithm (BD c-algorithm),
- Frobenius norm-based user selection algorithm (BD n-algorithm),
- round-robin algorithm for $\hat{K}$ simultaneous users (BD no selection).

Figs. 1–3 show the ergodic sum capacity (averaged over 1000 channel realizations) vs. the number of users for $(n_t = 4, n_r = 2)$, $(n_t = 12, n_r = 4)$, and $(n_t = 8, n_r = 1)$ MIMO systems, where $\hat{K} = 2, 3, 8$, respectively. Fig. 3 shows only up to 16 users in the system due to the complexity of the exhaustive search method. In all simulations, the capacity-based and the norm-based user selection algorithms achieve around 95% of the total throughput of the complete search method. The capacity-based algorithm performs slightly better than the norm-based algorithm because its user selection criterion is directly based on the sum capacity. For low SNRs, e.g. $SNR = 0$ dB, the proposed algorithms achieve almost the same sum capacity as the exhaustive search method. This is true because beamforming to the user with the highest capacity, which is the first step in the capacity-based user selection algorithm, is asymptotically optimal for sum capacity of BD in the low SNR regime. For high SNRs, although the proposed algorithms may not always find the optimal user set due to their reduced search range, they can still achieve a significant part of the ergodic sum capacity of the exhaustive search method because both algorithms greedily try to maximize the total throughput. The sum capacity achieved by dirty paper coding (DPC) is also plotted in Figs. 1–3. In general, DPC achieves higher sum capacity than BD because DPC is optimal for the sum capacity of MIMO broadcast channels [2][3]. BD, however, still achieves a significant part of the DPC sum capacity. Further, the low complexity property of the BD algorithm (e.g. without the requirement for successively encoding and decoding user signals) makes it more suitable for practical implementations.

VI. Conclusion

Two suboptimal user selection algorithms for multiuser MIMO systems with block diagonalization are proposed in this paper. The goal is to select a subset of users to maximize the total throughput while keeping the complexity low. The brute-force complete search method yields the optimal user set with the sum capacity achievement. However, the complexity of the complete search algorithm is roughly $\mathcal{O}\{K^{\hat{K}}\}$, where K is the total number of users and $\hat{K}$ is the maximum number of simultaneous users. Simulations show that the proposed capacity-based and norm-based user selection algorithms achieve about 95% of the sum capacity whereas their complexity is $\mathcal{O}\{K\hat{K}\}$. Although the proposed user selection algorithms are greedy in nature, they can be easily extended to incorporate fairness, e.g. the rate proportional fairness in [16].

REFERENCES

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

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

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

[4] M. Costa, “Writing on Dirty Paper,” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

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

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

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

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

[9] Z. Pan, K. K. Wong, and T. S. Ng, “Generalized Multiuser Orthogonal Space-Division Multiplexing,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1969–1973, Nov. 2004.

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

[11] T. Yoo and A. J. Goldsmith, “Optimality of Zero-Forcing Beamforming with Multiuser Diversity,” in *IEEE International Conference on Communications*, vol. 1, pp. 542–546, May 2005.

[12] M. Gharavi-Alkhansari and A. B. Gershman, “Fast Antenna Subset Selection in MIMO Systems,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 339-347, Feb. 2004.

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

[14] K. Yu and B. Ottersten, “Models for MIMO Propagation Channels, a Review,” *Wiley Journal on Wireless Communications and Mobile Computing*, vol 2, no. 7, pp. 653–666, Nov. 2002.

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

[16] Z. Shen, J. G. Andrews, and B. L. Evans, “Adaptive Resource Allocation in Multiuser OFDM Systems with Proportional Fairness,” to appear in *IEEE Transactions on Wireless Communications*, Nov. 2005.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd Edition, The John Hopkisns Univ. Press, 1996.

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

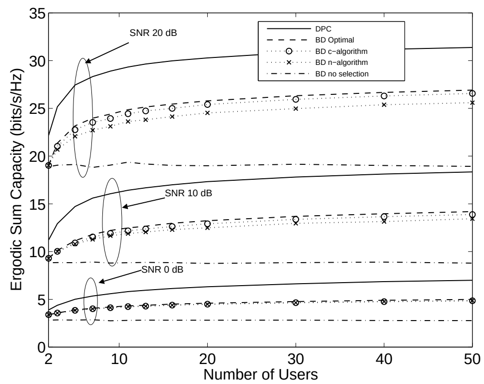

Fig. 1. Ergodic sum capacity vs. the number of users, where the number of transmit antennas is $n_t = 4$, the number of receive antennas is $n_r = 2$, and the maximum number of simultaneous users served by the block diagonalization algorithm is $\hat{K} = 2$.

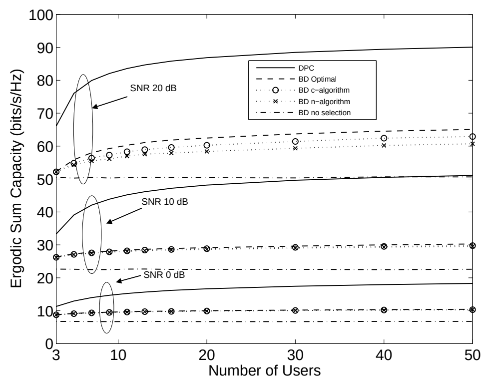

Fig. 2. Ergodic sum capacity vs. the number of users, where the number of transmit antennas is $n_t = 12$, the number of receive antennas is $n_r = 4$, and the maximum number of simultaneous users served by the block diagonalization algorithm is $\hat{K} = 3$.

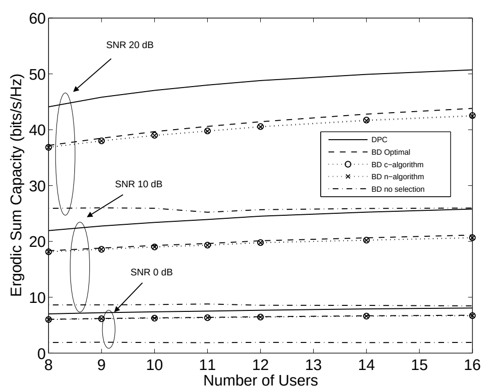

Fig. 3. Ergodic sum capacity vs. the number of users, where the number of transmit antennas is $n_t = 8$, the number of receive antennas is $n_r = 1$, and the maximum number of simultaneous users served by the block diagonalization algorithm is $\hat{K} = 8$.

TABLE I

CAPACITY-BASED SUBOPTIMAL USER SELECTION ALGORITHM

1) Initially, let $\Omega = \{1, 2, \cdots, K\}$ and $\Upsilon = \emptyset$. Let $s_1 = \arg\max_{k\in\Omega} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^*\right|$ where $\mathrm{Tr}(\mathbf{Q}_k) \leq P$ and $\mathbf{Q}_k$ is semi-positive definite. Let $\Omega = \Omega - \{s_1\}$ and $\Upsilon = \Upsilon + \{s_1\}$. Let $C_{temp} = \max_{k\in\Omega} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^*\right|$.
2) for $i = 2 : \hat{K}$
   a) For every $k \in \Omega$,
      i) Let $\overline{\Upsilon}_k = \Upsilon + \{k\}$.
      ii) Find the precoding matrix $\mathbf{T}_j$ for each $j \in \overline{\Upsilon}_k$, and obtain the effective channel $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ for each $j \in \overline{\Upsilon}_k$.
      iii) Perform a singular value decomposition (SVD) on $\overline{\mathbf{H}}_j$, and obtain the M singular values $\{\lambda_{j,m}\}_{i=1}^M$.
      iv) Water-fill over $\lambda_{j,m}^2$ for $j \in \overline{\Upsilon}_k$ and $1 \leq m \leq M$. Find the total throughput to the user set $\overline{\Upsilon}_k$, denoted as C_k.
   b) Let $s_i = \arg\max_{k\in\Omega} C_k$.
   c) If $\max_{k\in\Omega} C_k < C_{temp}$
      Algorithm terminated. The selected user set is Υ.
      else
      Let $\Omega = \Omega - \{s_i\}$ and $\Upsilon = \Upsilon + \{s_i\}$. And let $C_{temp} = \max_{k\in\Omega} C_k$.

TABLE II

FROBENIUS NORM-BASED SUBOPTIMAL USER SELECTION ALGORITHM

1) Initially, let $\Omega = \{1, 2, \cdots, K\}$ and $\Upsilon = \emptyset$. Let $s_1 = \arg\max_{k\in\Omega} ||\mathbf{H}_k||_F^2$. Let $\mathbf{V} = \mathbf{V}_{s_1}$. Let $\Omega = \Omega - \{s_1\}$ and $\Upsilon = \Upsilon + \{s_1\}$.
2) for $i = 2 : \hat{K}$
   a) For each $k \in \Omega$, let $\widetilde{\mathbf{H}}_k = \mathbf{H}_k - \mathbf{H}_k\mathbf{V}^*\mathbf{V}$. Then $\widetilde{\mathbf{H}}_k$ is in the null space of $\mathbf{V}$.
      for $j = 1 : i - 1$
      i) Let
      $$\hat{\mathbf{H}}_{s_j,k} = [\mathbf{H}_{s_1}^T \ \cdots \ \mathbf{H}_{s_{j-1}}^T \ \mathbf{H}_{s_{j+1}}^T \ \cdots \ \mathbf{H}_{s_{i-1}}^T \ \mathbf{H}_k^T]^T.$$
      ii) Let $\mathbf{W}_{s_j,k}$ be the row basis for $\hat{\mathbf{H}}_{s_j,k}$ after Gram-Schmidt orthogonalization.
   b) For each $s \in \Upsilon$, let $\widetilde{\mathbf{H}}_s = \mathbf{H}_s - \mathbf{H}_s\mathbf{W}_{s,k}^*\mathbf{W}_{s,k}$. Then $\widetilde{\mathbf{H}}_s$ is in the null space of $\hat{\mathbf{H}}_{s,k}$. Let
      $$s_i = \arg\max_{k\in\Omega}\left(\sum_{s\in\Upsilon} ||\widetilde{\mathbf{H}}_s||_F^2 + ||\widetilde{\mathbf{H}}_k||_F^2\right).$$
   c) Let $\Omega = \Omega - \{s_i\}$ and $\Upsilon = \Upsilon + \{s_i\}$. Apply the Gram-Schmidt orthogonalization procedure to $\widetilde{\mathbf{H}}_{s_i}$ and get $\widetilde{\mathbf{V}}_{s_i}$. Let $\mathbf{V} = [\mathbf{V}^T \ \ \widetilde{\mathbf{V}}_{s_i}^T]^T$.
3) Apply the capacity-based suboptimal user selection algorithm to the set Υ, and get the final selected user set and the total throughput.