Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels

Quentin H. Spencer, *Student Member, IEEE*, A. Lee Swindlehurst, *Senior Member, IEEE*, and Martin Haardt, *Senior Member, IEEE*

***Abstract*—The use of space-division multiple access (SDMA) in the downlink of a multiuser multiple-input, multiple-output (MIMO) wireless communications network can provide a substantial gain in system throughput. The challenge in such multiuser systems is designing transmit vectors while considering the co-channel interference of other users. Typical optimization problems of interest include the capacity problem—maximizing the sum information rate subject to a power constraint—or the power control problem—minimizing transmitted power such that a certain quality-of-service metric for each user is met. Neither of these problems possess closed-form solutions for the general multiuser MIMO channel, but the imposition of certain constraints can lead to closed-form solutions. This paper presents two such constrained solutions. The first, referred to as "block-diagonalization," is a generalization of channel inversion when there are multiple antennas at each receiver. It is easily adapted to optimize for either maximum transmission rate or minimum power and approaches the optimal solution at high SNR. The second, known as "successive optimization," is an alternative method for solving the power minimization problem one user at a time, and it yields superior results in some (*e.g.,* low SNR) situations. Both of these algorithms are limited to cases where the transmitter has more antennas than all receive antennas combined. In order to accommodate more general scenarios, we also propose a framework for coordinated transmitter-receiver processing that generalizes the two algorithms to cases involving more receive than transmit antennas. While the proposed algorithms are suboptimal, they lead to simpler transmitter and receiver structures and allow for a reasonable tradeoff between performance and complexity.**

***Index Terms*—Antenna arrays, array signal processing, MIMO systems, signal design, space division multiaccess (SDMA), wireless LAN.**

I. Introduction

THERE has been considerable recent interest in wireless multiple-input, multiple-output (MIMO) communications systems, due to their potential for dramatic gains in channel capacity. To date, research has focused on the single-user point-to-point scenario where the transmitter and receiver each have arrays, and the presence of other co-channel users is not considered. More recently, attention has shifted to multiuser MIMO channels, where several co-channel users with arrays attempt to communicate with each other or with some central base station [1]–[7]. Research in this area has focused on two related optimization problems that are of particular interest: throughput maximization (capacity) and power control. To achieve (sum) capacity in a multiuser network, one maximizes the sum of the information rates for all users subject to a sum power constraint. On the other hand, the power control problem deals with minimizing the total transmitted power while achieving a prespecified minimum Quality-of-Service (QoS) level for each user in the network. In either case, a satisfactory solution must balance the desire for high throughput or good QoS at one node in the network with the resulting cost in interference produced at other nodes.

The capacity of the vector multiple access channel (where arrays are employed at the transmit and possibly all receive nodes in the network) has been studied in [8]–[10], and its connection with the broadcast channel has been explored in [11]. The particular challenge of the vector broadcast channel is that while the transmitter has the ability to coordinate transmission from all of its antennas, the receivers are grouped among different users that are typically unable to coordinate with each other [12]–[14]. The capacity of the broadcast channel has been studied recently in [15] and [16], for the special case where each user has only one antenna, and in [17], for users with arrays of arbitrary size. A feature common to some of the new work cited above is the use of a technique developed by Costa known as "dirty paper coding" [18]. The fundamental idea of this approach is that when a transmitter has advance knowledge of the interference in a channel, it can design a code to compensate for it, and the capacity of the channel is the same as if there were no interference. For the multiuser MIMO downlink, the interference due to signals transmitted to other users is known at the transmitter, and in principle, a precoder could be used to essentially undo its effects. The primary drawback of such schemes is that their use of nontraditional coding leads to increased complexity at both the transmitter and receiver.

For the special case where the base station has an array but all users employ single antennas, alternative solutions have been proposed in [19]–[22]. The more general problem considered in this paper, where each user may have multiple antennas, has been approached in two different ways. The first [23] employs an iterative method of canceling out interuser interference, allowing multiple data subchannels per user as in classical MIMO transmission methods. The second approach [24] generalizes the single-antenna algorithms to include beamforming at the receiver while still using only a single data subchannel per user. The iterative nature of these algorithms typically results in a high computational cost.

Manuscript received December 12, 2002; revised May 3, 2003. The associate editor coordinating the review of this paper and approving it for publication was Dr. Helmut Bölcskei.

Q. H. Spencer and A. L. Swindlehurst are with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602 USA (e-mail: qspencer@ieee.org; swindle@ee.byu.edu)

M. Haardt is with the Communications Research Laboratory, Ilmenau University of Technology, Ilmenau, Germany (e-mail: martin.haardt@tu-ilmenau.de)

Digital Object Identifier 10.1109/TSP.2003.821107

In this paper, we present three different noniterative algorithms for choosing downlink transmit vectors for the case where the users in the network have multiple antennas. The first, known as *block diagonalization* (BD), can be thought of as a generalization of channel inversion for situations with multiple antennas per user. The BD algorithm can be applied to either the throughput maximization or power control problems but is restricted to channels where the number of transmit antennas (n_T) is no smaller than the total number of receive antennas in the network (n_R). The second method is a *successive optimization* algorithm that addresses the power control problem one user at a time. It can outperform BD at low SNR, but it has the same limitation on channel dimensions. Finally, we propose a method for *coordinated transmit-receive* processing, which relaxes the $n_T \geq n_R$ requirement by combining either of the previous algorithms together with the method of [24]. This hybrid approach accommodates up to n_T users, regardless of their array sizes. The primary advantage of this and the other techniques proposed in the paper are that they provide efficient, closed-form solutions that yield a reasonable tradeoff between performance and computational complexity.

In the next section, we begin with the MIMO transmission model that will be assumed in the paper. Section III then outlines the BD algorithm for two cases: first, where the transmitter has complete channel information and, second, where it has incomplete or partial information. Section IV describes the successive optimization algorithm for achieving power control with arbitrary rate points. Section V discusses coordinated transmit-receive processing, which is a framework for extending the first two algorithms to handle larger channel geometries, and finally, Section VI presents simulation results comparing the algorithms under various conditions.

II. MIMO Transmission Model and Channel Capacity

A flat-fading MIMO channel with n_T transmitters and n_R receivers is typically modeled by an $n_R \times n_T$ matrix $\mathbf{H}$ so that the received signal $\mathbf{x}$ is

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

where $\mathbf{s}$ is the signal vector, and $\mathbf{n}$ represents additive noise. In the flat-fading case, each element of $\mathbf{H}$ is viewed as the transmission coefficient linking one of the transmit antennas with one of the receive antennas. However, this model can also be easily extended to include frequency selective fading by writing the overall channel matrix as a block matrix whose component blocks implement a convolution with the time-domain impulse response of a particular antenna pair [12]. Thus, any optimization algorithm for a flat-fading channel can easily be extended to include frequency selective channels. For simplicity, flat fading will be assumed here.

We focus on MIMO transmission systems that include linear pre- and post-processing performed at the transmitter and receiver [25]:

$$\hat{\mathbf{d}} = \mathbf{D}(\mathbf{H}\mathbf{M}\mathbf{d} + \mathbf{n})$$

where $\mathbf{d}$ is a data vector of arbitrary dimension m, and the actual transmitted signal $\mathbf{s} = \mathbf{M}\mathbf{d}$ is generated using an $n_T \times m$ modulation matrix $\mathbf{M}$ that includes all channel precoding done at the transmitter. The received signal $\mathbf{x}$ is converted into an estimate of the original transmitted data $\mathbf{d}$ by an $m \times n_R$ demodulation matrix $\mathbf{D}$.

Consider a multiuser downlink channel with K users and a single base station. The base has n_T antennas, and the jth receiver has n_{Rj} antennas. The total number of antennas at all receivers is defined to be $n_R = \sum n_{Rj}$. We will use the notation $\{n_{R1}, \ldots, n_{RK}\} \times n_T$ to represent such a channel (as opposed to writing $n_R \times n_T$ as in a point-to-point MIMO channel). For example, a $\{2,2\} \times 4$ channel has a four-antenna base and two two-antenna users. The channel matrix from the base to the j^{th} user is denoted by $\mathbf{H}_j$ and the associated modulation matrix by $\mathbf{M}_j$. The signal at the jth receiver is thus

$$\mathbf{x}_j = \sum_{i=1}^{K} \mathbf{H}_j \mathbf{M}_i \mathbf{d}_i + \mathbf{n}_j \tag{1}$$

$$= \mathbf{H}_j \mathbf{M}_j \mathbf{d}_j + \mathbf{H}_j \tilde{\mathbf{M}}_j \tilde{\mathbf{d}}_j + \mathbf{n}_j \tag{2}$$

where $\tilde{\mathbf{M}}_j$ and $\tilde{\mathbf{d}}_j$ are, respectively, defined as the modulation matrix and transmit vector for all users other than user j combined:

$$\tilde{\mathbf{M}}_j = [\mathbf{M}_1 \quad \ldots \quad \mathbf{M}_{j-1} \quad \mathbf{M}_{j+1} \quad \ldots \quad \mathbf{M}_K] \tag{3}$$

$$\tilde{\mathbf{d}}_j^T = [\mathbf{d}_1^T \quad \ldots \quad \mathbf{d}_{j-1}^T \quad \mathbf{d}_{j+1}^T \quad \ldots \quad \mathbf{d}_K^T]. \tag{4}$$

The high capacity potential of single-user channels can be realized by transmitting multiple data subchannels in parallel. The optimal way of doing this depends on what information is available to the transmitter about the channel. If $\mathbf{H}$ is known perfectly to the transmitter, capacity is achieved by choosing $\mathbf{M}$ as the right singular vectors of $\mathbf{H}$ and weighting the transmit power into each vector using water-filling on the corresponding singular values [12]. If $\mathbf{H}$ is unknown, the ergodic capacity can be achieved for i.i.d. Gaussian channels by choosing $\mathbf{M} = \alpha\mathbf{I}$ [13]. The difference in performance between these two approaches has been shown to be small at high SNR [26]. At lower SNR, the water-filling solution yields some improvement in performance, but this must be balanced against the cost of obtaining knowledge of the channel at the transmitter. On the other hand, in a MIMO channel where a single base station is simultaneously transmitting to multiple independent receivers and generating co-channel interference, the situation becomes considerably different. In such cases, channel information at the transmitter provides a considerable advantage, *particularly* at high SNR, since it can be used for interference mitigation.

The channel modulation and demodulation matrices can be viewed as attempting to diagonalize the product $\mathbf{DHM}$. Although the optimal solution is not necessarily diagonal, it will generally be near-diagonal in most situations. The BLAST approach [14], which does not use any channel precoding, essentially leaves the task of diagonalization to the receiver. On the other hand, the water-filling solution breaks the channel down into its dominant subspaces so that optimal power loading into the subchannels can be performed. In this case, the diagonalization is accomplished by a combination of both $\mathbf{M}$ and $\mathbf{D}$.

For a multiuser system with an array at the transmitter and K single-antenna receivers, no coordination is possible among the receivers, so channel diagonalization (if desired) must been done entirely by the transmitter. Perfect diagonalization is only possible for $n_T \geq K$ and can be achieved using *channel inversion*, e.g., by choosing $\mathbf{M} = \mathbf{H}^{\dagger}$, where $\mathbf{H}^{\dagger}$ is the pseudo-inverse of $\mathbf{H}$ [15], [27]. On the other hand, when each of the K users has multiple antennas, complete diagonalization of the channel at the transmitter is suboptimal since each user is able to coordinate the processing of its own receiver outputs. If we define the network channel and modulation matrices $\mathbf{H}_S$ and $\mathbf{M}_S$ as

$$\mathbf{H}_S = [\mathbf{H}_1^T \quad \mathbf{H}_2^T \quad \dots \quad \mathbf{H}_K^T]^T$$
$$\mathbf{M}_S = [\mathbf{M}_1 \quad \mathbf{M}_2 \quad \dots \quad \mathbf{M}_K]$$

the optimal solution under the constraint that all interuser interference be zero is one, where $\mathbf{H}_S\mathbf{M}_S$ is block diagonal. Like channel inversion, a block-diagonal solution imposes two conditions—one on the dimensions and one on the independence of the component $\mathbf{H}_j$ matrices—although it will be shown in the next section that the conditions are somewhat less strict for a block-diagonal solution. However, there is still a limitation on how many users can be accomodated simultaneously. These conditions are not as restrictive as they appear when viewed in the context of a system that uses SDMA in conjunction with other multiple access methods (TDMA, FDMA, etc.). Consider a base station with a small number of antennas and a large group of users, where an SDMA-only solution is impractical. A more realistic implementation would divide the users into subgroups (organized so that the dimension requirements are satisfied within each group) whose members are multiplexed spatially, while the subgroups themselves are assigned different time or frequency slots. The linear independence condition can be met by intelligently grouping the users to avoid placing two users with highly correlated channels in the same subgroup.

An algorithm for achieving a block-diagonal solution is presented in the following section.

III. Block Diagonalization Algorithm

This section outlines a procedure for finding the optimal transmit vectors $\mathbf{M}_S$ such that all multiuser interference is zero. Since the resulting product $\mathbf{H}_S\mathbf{M}_S$ will be block diagonal, the algorithm is referred to here as BD. Note that when $n_{Rj} = 1$ for all users, this simplifies to a complete diagonalization, which can be achieved using a pseudo-inverse of the channel. While complete diagonalization could also be applied when $n_{Rj} > 1$ and would have the advantage of simplifying the receiver (each antenna would receive only one signal), it comes at the cost of reduced throughput or requiring higher power at the transmitter, particularly when there is significant spatial correlation between the antennas at the receiver. The two approaches are compared in the simulation results of Section VI.

A. BD for Throughput Maximization

To eliminate all multi-user interference, we impose the constraint that $\mathbf{H}_j\mathbf{M}_j = \mathbf{0}$ for $i \neq j$. With a sum power constraint, the achievable throughput for the resulting block-diagonal system is

$$C_{\text{BD}} = \max_{\mathbf{M}_S,\, \mathbf{H}_i\mathbf{M}_j=\mathbf{0},\, i\neq j} \log_2 \left| \mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{H}_S\mathbf{M}_S\mathbf{M}_S^*\mathbf{H}_S^* \right| \quad (5)$$

$$= \max_{\mathbf{H}_i\mathbf{M}_j=\mathbf{0},\, i\neq j} \sum_{j=1}^{K} \log_2 \left| \mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{H}_j\mathbf{M}_j\mathbf{M}_j^*\mathbf{H}_j^* \right| \leq C_S \quad (6)$$

where C_S represents the sum capacity of the system, and * indicates the Hermitian transpose. If we define $\tilde{\mathbf{H}}_j$ as

$$\tilde{\mathbf{H}}_j = [\mathbf{H}_1^T \quad \dots \quad \mathbf{H}_{j-1}^T \quad \mathbf{H}_{j+1}^T \quad \dots \quad \mathbf{H}_K^T]^T \quad (7)$$

the zero-interference constraint forces $\mathbf{M}_j$ to lie in the null space of $\tilde{\mathbf{H}}_j$. This definition allows us to define the dimension condition necessary to guarantee that all users can be accomodated under the zero-interference constraint. Data can be transmitted to user j if the null space of $\tilde{\mathbf{H}}_j$ has a dimension greater than 0. This is satisfied when $\text{rank}(\tilde{\mathbf{H}}_j) < n_T$. So for any $\mathbf{H}_S$, block diagonalization is possible if $n_T > \max\{\text{rank}(\tilde{\mathbf{H}}_1), \dots, \text{rank}(\tilde{\mathbf{H}}_K)\}$. Thus, it is theoretically possible to support some situations where both $n_R > n_T$ and $\text{rank}(\mathbf{H}_S) > n_T$ (for example, the {3,3} $\times 4$ channel). Assuming the dimension condition is satisfied for all users, let $\tilde{L}_j = \text{rank}(\tilde{\mathbf{H}}_j) \leq n_R - n_{Rj}$, and define the singular value decomposition (SVD)

$$\tilde{\mathbf{H}}_j = \tilde{\mathbf{U}}_j\tilde{\mathbf{\Sigma}}_j \left[\tilde{\mathbf{V}}_j^{(1)} \quad \tilde{\mathbf{V}}_j^{(0)} \right]^* \quad (8)$$

where $\tilde{\mathbf{V}}_j^{(1)}$ holds the first $\tilde{L}_j$ right singular vectors, and $\tilde{\mathbf{V}}_j^{(0)}$ holds the last $(n_T - \tilde{L}_j)$ right singular vectors. Thus, $\tilde{\mathbf{V}}_j^{(0)}$ forms an orthogonal basis for the null space of $\tilde{\mathbf{H}}_j$, and its columns are, thus, candidates for the modulation matrix $\mathbf{M}_j$ of user j.

Let $\bar{L}_j$ represent the rank of the product $\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}$. In order for transmission to user j to take place under the zero-interference constraint, $\bar{L}_j \geq 1$ is necessary. In general, $\bar{L}_j$ is bounded by $L_j + \tilde{L}_j - n_T \leq \bar{L}_j \leq \min\{L_j, \tilde{L}_j\}$ [28]. A sufficient condition for $\bar{L}_j \geq 1$ is that at least one row of $\mathbf{H}_j$ is linearly independent of the rows of $\tilde{\mathbf{H}}_j$. To satisfy this condition, one should take care to avoid spatially multiplexing users with highly correlated channel matrices. Note that both the dimension and independence conditions allow certain cases that cannot be handled by channel inversion. The channel inversion approach would require that all rows of $\mathbf{H}_j$ be linearly independent of $\tilde{\mathbf{H}}_j$. While this is not necessary for block diagonalization, it would still be beneficial, resulting in a higher value of $\bar{L}_j \geq 1$ and, thus, greater degrees of freedom for the final solution. Assuming that the independence condition is satisfied for all users, we now define the matrix

$$\mathbf{H}'_S = \begin{bmatrix} \mathbf{H}_1\tilde{\mathbf{V}}_1^{(0)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H}_K\tilde{\mathbf{V}}_K^{(0)} \end{bmatrix}. \quad (9)$$

The system capacity under the zero-interference constraint can now be written as

$$C_{\text{BD}} = \max_{\mathbf{M}'_S} \log_2 \left| \mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{H}'_S\mathbf{M}'_S\mathbf{M}'^*_S\mathbf{H}'^*_S \right|. \quad (10)$$

The problem is now to find a matrix $\mathbf{M}'_S$ that maximizes the determinant. This is now equivalent to the single-user MIMO capacity problem, and the solution is to let $\mathbf{M}'_S$ be the right singular vectors of $\mathbf{H}'_S$, weighted by water-filling on the corresponding singular values [12]. Thus, a solution for $\mathbf{M}'_S$ based on an SVD and water-filling is the solution that maximizes sum capacity for the system under the zero-interference constraint.

The block structure of $\mathbf{H}'_S$ allows the SVD to be determined individually for each user, rather than computing a single large SVD. Define the SVD

$$\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j \begin{bmatrix} \mathbf{\Sigma}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_j^{(1)} & \mathbf{V}_j^{(0)} \end{bmatrix}^* \tag{11}$$

where $\mathbf{\Sigma}_j$ is $\bar{L}_j \times \bar{L}_j$, and $\mathbf{V}_j^{(1)}$ represents the first $\bar{L}_j$ singular vectors. The product of $\tilde{\mathbf{V}}_j^{(0)}$ and $\mathbf{V}_j^{(1)}$ now produces an orthogonal basis of dimension $\bar{L}_j$ and represents the transmission vectors that maximize the information rate for user j subject to producing zero interference. Thus, we define the modulation matrix as

$$\mathbf{M}_S = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)} & \mathbf{V}_1^{(1)} & \tilde{\mathbf{V}}_2^{(0)} & \mathbf{V}_2^{(1)} & \cdots & \tilde{\mathbf{V}}_K^{(0)} & \mathbf{V}_K^{(1)} \end{bmatrix} \mathbf{\Lambda}^{1/2} \tag{12}$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose elements λ_i scale the power transmitted into each of the columns of $\mathbf{M}_S$.

With $\mathbf{M}_S$ chosen as in (12), the capacity of the BD method in (6) becomes

$$C_{\mathrm{BD}} = \max_{\mathbf{\Lambda}} \log_2 \left| \mathbf{I} + \frac{\mathbf{\Sigma}^2\mathbf{\Lambda}}{\sigma_n^2} \right| \tag{13}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & & \\ & \ddots & \\ & & \mathbf{\Sigma}_K \end{bmatrix}. \tag{14}$$

The optimal power loading coefficients in $\mathbf{\Lambda}$ are then found using water-filling on the diagonal elements of $\mathbf{\Sigma}$, assuming a total power constraint P. A summary of the BD algorithm is given below.

Sum Capacity Block Diagonalization

1) For $j = 1, \ldots, K$:
Compute $\tilde{\mathbf{V}}_j^{(0)}$, the right null space of $\tilde{\mathbf{H}}_j$.
Compute the SVD

$$\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j \begin{bmatrix} \mathbf{\Sigma}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_j^{(1)} & \mathbf{V}_j^{(0)} \end{bmatrix}^*.$$

2) Use water filling on the diagonal elements of $\mathbf{\Sigma}$ to determine the optimal power loading matrix $\mathbf{\Lambda}$ under power constraint P.

3) Set
$$\mathbf{M}_S = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(1)} & \tilde{\mathbf{V}}_2^{(0)}\mathbf{V}_2^{(1)} & \cdots & \tilde{\mathbf{V}}_K^{(0)}\mathbf{V}_K^{(1)} \end{bmatrix} \mathbf{\Lambda}^{1/2}.$$

B. BD for Power Control

The problem with sum capacity maximization in a multiuser channel is that such an approach may result in one or two "strong" users (large $\mathbf{H}_j$) taking a dominant share of the available power, potentially leaving weak users with little or no throughput. Consequently, in practice, the dual problem is often of more interest, i.e., minimize power output at the transmitter subject to achieving a desired arbitrary rate (a measure of QoS) for each user. For the single-user MIMO channel, these two optimization problems are essentially equivalent. Things are different for the multiuser case, however, and achieving a set of arbitrary rate points is much more complex. This problem is addressed for the case where each user has a single antenna in [19] and [20]. We investigate below the more general case where all users may have multiple receive antennas.

If there are K users with desired rates $R_1, R_2, \ldots, R_K$, then in general, we must simultaneously solve K equations of the following form:

$$2^{R_j} = \left| \mathbf{I} + \left(\sigma_n^2\mathbf{I} + \sum_{i=1, i\neq j}^{K} \mathbf{H}_j\mathbf{M}_i\mathbf{M}_i^*\mathbf{H}_j^* \right)^{-1} \mathbf{H}_j\mathbf{M}_j\mathbf{M}_j^*\mathbf{H}_j^* \right| \tag{15}$$

such that $\mathrm{tr}(\mathbf{M}_S\mathbf{M}_S^*)$ is minimized. This a nonlinear system of equations with as many as $n_T n_R$ unknowns. Because single-user MIMO capacity is a monotonic function of the given power constraint, the converse problem of minimizing transmitted power for a given rate can be solved by water-filling. Extending this idea to the multiuser case, if the dependence of the equations can be removed by the addition of constraints, as done in the previous section with the throughput maximization problem, the power minimization problem can also be solved in closed form. However, as before, it may result in a solution that is not globally optimal.

There are at least two ways to impose constraints so that an explicit solution to the system of equations in (15) is possible. We discuss one based on BD here and propose another in the following section. In step 2 of the BD algorithm described above, water-filling with a total power constraint of P is performed with the singular values $\mathbf{\Sigma}_j$ from all users collected together. As an alternative, we replace this step by one that performs a water-filling solution separately for each user, where the power constraint for the user (denoted P_j) is scaled so that the rate requirement is satisfied. The BD procedure removes all interdependence in the equations and allows an explicit solution to each of the individual determinant maximizations. The algorithm is outlined in detail below.

Block Diagonalization for Power Control

1) For each user $j = 1, \ldots, K$:
Compute $\tilde{\mathbf{V}}_j^{(0)}$, the right null space of $\tilde{\mathbf{H}}_j$.
Compute the SVD

$$\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j \begin{bmatrix} \mathbf{\Sigma}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_j^{(1)} & \mathbf{V}_j^{(0)} \end{bmatrix}^*.$$

Use water-filling on the diagonal elements of $\mathbf{\Sigma}_j$ to calculate the power loading matrix $\mathbf{\Lambda}_j$ that achieves the power constraint P_j corresponding to rate R_j.

2) Form $\mathbf{\Lambda}$ using the diagonal blocks $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_K$.

3) Set
$$\mathbf{M}_S = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(1)} & \tilde{\mathbf{V}}_2^{(0)}\mathbf{V}_2^{(1)} & \cdots & \tilde{\mathbf{V}}_K^{(0)}\mathbf{V}_K^{(1)} \end{bmatrix} \mathbf{\Lambda}^{1/2}.$$

C. Partial Channel Knowledge

Thus far, we have assumed that the transmitter has knowledge of each channel matrix $\mathbf{H}_j$. In certain instances, this can be achieved using training data in a time-division duplex system or by means of channel feedback from the receiver. However, there are situations where it is possible to only obtain partial rather than full channel state information. In this section, we show how the BD algorithm can be implemented for cases such that $\mathbf{H}_j = \mathbf{A}_j\mathbf{B}_j$, where $\mathbf{B}_j$ is known but $\mathbf{A}_j$ is not [29]. One case where this model is applicable occurs when temporal averages are performed on the subspaces of $\mathbf{H}_j$ [30], and due to fast time variation, the signal subspace is more stable than the corresponding singular values. Another occurs in conjunction with "physical" channel models based on individual multipath components. For example, if $\mathbf{H}_j$ is composed of contributions from L_j multipath rays, we may write

$$\begin{aligned}\mathbf{H}_j &= \mathbf{\Phi}_{R,j}\mathbf{\Gamma}_j\mathbf{\Phi}_{T,j}\\ &\equiv [\boldsymbol{\alpha}_{R,j}(\theta_{j,1}) \quad \dots \quad \boldsymbol{\alpha}_{R,j}(\theta_{j,L_j})] \begin{bmatrix}\gamma_{j,1} & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & \gamma_{j,L_j}\end{bmatrix}\\ &\quad\times \begin{bmatrix}\boldsymbol{\alpha}_{T,j}^H(\phi_{j,1})\\ \vdots \\ \boldsymbol{\alpha}_{T,j}^H(\phi_{j,L_j})\end{bmatrix}\end{aligned} \tag{16}$$

where $\boldsymbol{\alpha}_{R,j}(\theta_{j,i})$ is the $n_{Rj}\times 1$ steering vector at receiver j for the ith multipath signal arriving from angle $\theta_{j,i}$, $\boldsymbol{\alpha}_{T,j}(\phi_{j,i})$ is the $n_T\times 1$ steering vector at the transmitter for the corresponding transmit angle of departure $\phi_{j,i}$, and $\gamma_{j,i}$ is the complex gain for the corresponding path. Under this model, the transmitter may be able to estimate uplink angles of arrival ($\phi_{j,i}$), but in the absence of feedback, it may have no information about either $\mathbf{\Phi}_{R,j}$ and $\mathbf{\Gamma}_j$. Thus, we associate $\mathbf{A}_j$ with $\mathbf{\Phi}_{R,j}\mathbf{\Gamma}_j$ (unknown) and $\mathbf{B}_j$ with $\mathbf{\Phi}_{T,j}$ (known) in the factorization $\mathbf{H}_j = \mathbf{A}_j\mathbf{B}_j$.

Assume $\mathbf{H}_j = \mathbf{A}_j\mathbf{B}_j$, where $\mathbf{A}_j$ is $n_R \times L_j$, $\mathbf{B}_j$ is $L_j \times n_T$, and $L_j \leq n_{Rj}$. Here, the condition $\mathbf{H}_i\mathbf{M}_j = 0, i \neq j$, which is necessary to make the system block diagonal, is equivalent to $\mathbf{B}_i\mathbf{M}_j = 0, i \neq j$. Thus, we define the matrix $\tilde{\mathbf{B}}_j$ as in (3):

$$\tilde{\mathbf{B}}_j = [\mathbf{B}_1^T \quad \dots \quad \mathbf{B}_{j-1}^T \quad \mathbf{B}_{j+1}^T \quad \dots \quad \mathbf{B}_K^T]^T. \tag{17}$$

Let the SVD of $\tilde{\mathbf{B}}_j$ be $\tilde{\mathbf{U}}_{B_j}\tilde{\mathbf{\Sigma}}_{B_j}[\tilde{\mathbf{V}}_{B_j}^{(1)} \quad \tilde{\mathbf{V}}_{B_j}^{(0)}]^*$, where $\tilde{\mathbf{V}}_{B_j}^{(0)}$ corresponds to the right null space of $\tilde{\mathbf{B}}_j$. The optimal modulation matrix for user j, subject to the constraint that the interuser interference is zero, is now of the form $\tilde{\mathbf{V}}_{B_j}^{(0)}\mathbf{M}_j'$ for some choice of transmit vectors $\mathbf{M}_j'$. The system capacity of the BD approach in this case is thus (18)–(21), shown at the bottom of the page. Equation (20) is a "high SNR" approximation achieved by dropping the identity matrix in the previous equation. The last equation has two terms, one of which is dependent on the noise and terms unknown to the transmitter and the second of which contains only known variables and the transmit vectors $\mathbf{M}_j'$. Thus, at high SNR, the optimal transmit matrix will only depend on the part of the channel that is known ($\mathbf{B}_j$) and not on the part that is unknown ($\mathbf{A}_j$). Equation (21) can be maximized by choosing $\mathbf{M}_j'$ to diagonalize the matrix inside the determinant, which is accomplished by letting it equal the right singular vectors of $\mathbf{B}_j\tilde{\mathbf{V}}_{B_j}^{(0)}$. In the standard MIMO capacity maximization problem, there is still a sum inside the determinant at this point due to the noise term, which leads to the water-filling solution. However, because the noise term has been removed using the high SNR approximation, the determinant is now maximized by equally dividing the power among each spatial dimension.

IV. Successive Optimization Algorithm

In this section, we describe another way of constraining the power control problem in order to achieve a closed-form solution. In the approach described here, we solve the equations one user at a time, optimizing each transmit matrix such that it does not interfere with any of the previous users. User j must optimize its transmit power to compensate for the interference received from users $1,\dots,j-1$ and subject to the constraint that it does not interfere with any of those users. We refer to this approach as successive optimization (SO) and describe it in detail below. The capacity-achieving schemes in [11], [15], and [17] have a similar structure, but they assume at each successive step that the interfering signals are known completely and use knowledge of these signals in coding the next signal. Here, the only information used are the statistics of the interfering signals from previous steps, and hence, the solution will be valid as long as the channel and the users' statistics are stationary.

$$C_{\text{BD}} = \max_{\mathbf{M}_j', j=1,\dots,K} \sum_{j=1}^{K} \log_2 \left|\mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{A}_j\mathbf{B}_j\tilde{\mathbf{V}}_{B_j}^{(0)}\mathbf{M}_j'\mathbf{M}_j'^*\tilde{\mathbf{V}}_{B_j}^{(0)*}\mathbf{B}_j^*\mathbf{A}_j^*\right| \tag{18}$$

$$= \max_{\mathbf{M}_j', j=1,\dots,K} \sum_{j=1}^{K} \log_2 \left|\mathbf{I} + \frac{1}{\sigma_n^2}\mathbf{A}_j^*\mathbf{A}_j\mathbf{B}_j\tilde{\mathbf{V}}_{B_j}^{(0)}\mathbf{M}_j'\mathbf{M}_j'^*\tilde{\mathbf{V}}_{B_j}^{(0)*}\mathbf{B}_j^*\right| \tag{19}$$

$$\approx \max_{\mathbf{M}_j', j=1,\dots,K} \sum_{j=1}^{K} \left[\log_2 \left|\frac{1}{\sigma_n^2}\mathbf{A}_j^*\mathbf{A}_j\right| + \log_2 |\mathbf{B}_j\tilde{\mathbf{V}}_{B_j}^{(0)}\mathbf{M}_j'\mathbf{M}_j'^*\tilde{\mathbf{V}}_{B_j}^{(0)*}\mathbf{B}_j^*|\right] \tag{20}$$

$$= \sum_{j=1}^{K} \log_2 \left|\frac{1}{\sigma_n^2}\mathbf{A}_j^*\mathbf{A}_j\right| + \max_{\mathbf{M}_j', j=1,\dots,K} \sum_{j=1}^{K} \log_2 |\mathbf{M}_j'^*\tilde{\mathbf{V}}_{B_j}^{(0)*}\mathbf{B}_j^*\mathbf{B}_j\tilde{\mathbf{V}}_{B_j}^{(0)}\mathbf{M}_j'|. \tag{21}$$

Assuming that user j's signal is not interfered with by any subsequent user's transmissions $(j+1,\ldots,K)$, the noise and interference matrix for user j is

$$\mathbf{R}_{ni,j} = \sigma_n^2\mathbf{I} + \sum_{i=1}^{j-1}\mathbf{H}_j\mathbf{M}_i\mathbf{M}_i^*\mathbf{H}_j^*. \tag{22}$$

Define the SVD of the previous $j-1$ users' combined channel matrix as

$$\hat{\mathbf{H}}_j = [\mathbf{H}_1^T \quad \mathbf{H}_2^T \quad \ldots \quad \mathbf{H}_{j-1}^T]^T = \mathbf{U}_j\mathbf{\Lambda}_j\left[\tilde{\mathbf{V}}_j^{(1)} \quad \tilde{\mathbf{V}}_j^{(0)}\right]^*. \tag{23}$$

If the rank of $\hat{\mathbf{H}}_j$ is $\hat{L}_j$, then $\tilde{\mathbf{V}}_j^{(0)}$ contains the last $n_T-\hat{L}_j$ right singular vectors. As in the BD solution, we force the modulation matrix $\mathbf{M}_j$ to lie in the null space of $\hat{\mathbf{H}}_j$ by setting $\mathbf{M}_j=\tilde{\mathbf{V}}_j^{(0)}\mathbf{M}_j'$ for some choice of $\mathbf{M}_j'$. We now need to solve (24), shown at the bottom of the page, such that $\mathrm{tr}(\mathbf{M}_j'\mathbf{M}_j'^*)$ is minimized. Under the constraints we have imposed, the solution can be found independently for each user. Finding $\mathbf{M}_j'$ to maximize the determinant leads to a water-filling solution using the following SVD:

$$\tilde{\mathbf{V}}_j^{(0)*}\mathbf{H}_j^*\left(\sigma_n^2\mathbf{I}+\sum_{i=1}^{j-1}\mathbf{H}_j\mathbf{M}_i\mathbf{M}_i^*\mathbf{H}_j^*\right)^{-1}\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)} = \mathbf{W}_j\mathbf{\Lambda}_{H,j}\mathbf{W}_j^*. \tag{25}$$

The values of $\mathbf{\Lambda}_{H,j}$, the noise power, and the total power constraint are used to compute the power loading coefficients $\mathbf{\Lambda}_{Z,j}$ by means of the water-filling solution, and the modulation matrix for user j then becomes

$$\mathbf{M}_j = \tilde{\mathbf{V}}_j^{(0)}\mathbf{W}_j\mathbf{\Lambda}_{Z,j}^{1/2} \tag{26}$$

where the water-filling coefficients in $\mathbf{\Lambda}_{Z,j}$ are chosen such that the rate requirement R_j is satisfied. The total transmitted power for all users is then the sum of the elements of all $\mathbf{\Lambda}_{Z,j}$.

Using either the SO or BD methods results in a "rate region," which are the convex set of achievable rates for all users at a fixed total power level. To illustrate the properties of the two optimization algorithms, Figs. 1 and 2 show two-dimensional rate regions for a randomly chosen $\mathbf{H}$ matrix with four transmitters, and two users with two antennas each. Fig. 2 uses the same $\mathbf{H}$ as Fig. 1, except that the channel of user 2 is attenuated by 10 dB, thus creating the so-called "near–far" problem. For only two users, there are three possible regions: the region resulting from BD, and two regions for SO—one where user one is optimized before user two (U1) and one for the opposite case (U2). The BD rate regions are derived by equally dividing the power among the users and choosing the power loading coefficients by "local" water-filling, as in Section III-B, rather than globally. For comparison, an additional curve is shown for the case where the channel is unknown to the transmitter. This latter curve corresponds to transmission to a single user at a time; therefore, the rate region is the line connecting the blind channel capacities for the two users. Three sets of curves are shown, for system SNRs of 3, 10, and 20 dB, respectively. The point on each curve representing the maximum sum capacity is indicated with a "*." In Figs. 1 and 2, on the outermost (20-dB SNR) curves, the BD solution offers the highest sum capacity, but on the innermost set of curves (3 dB SNR), the region where BD offers a performance improvement over either of the SO curves is very small in Fig. 1 and nonexistent

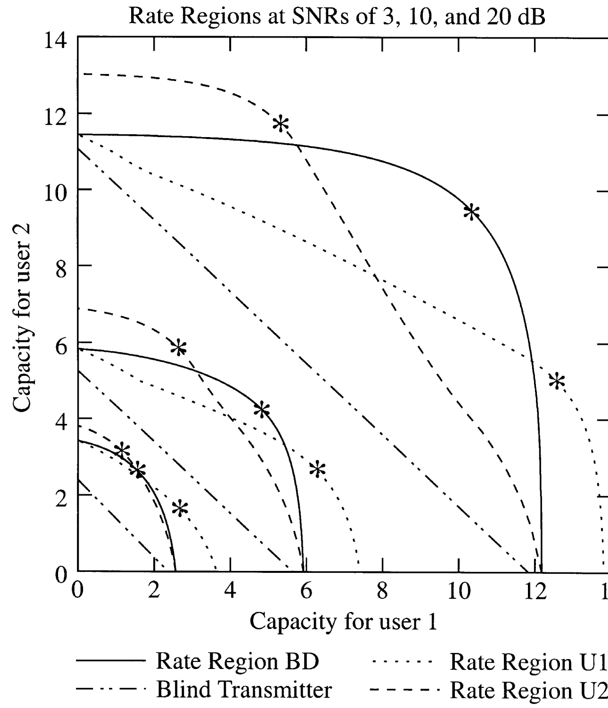

Fig. 1. Rate regions for a randomly generated $\mathbf{H}$ of dimension {2,2} × 4 at various power constraints.

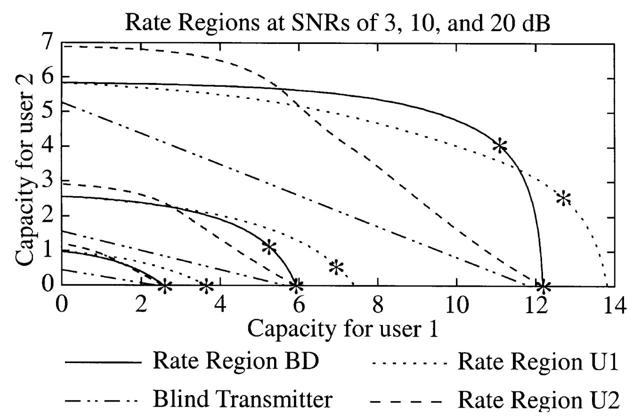

Fig. 2. Rate regions for a "Near–Far" $\mathbf{H}$ of dimension {2,2} × 4 with 10-dB difference between users.

$$2^{R_j} = \left|\mathbf{I}+\mathbf{M}_j'^*\tilde{\mathbf{V}}_j^{(0)*}\mathbf{H}_j^*\left(\sigma_n^2\mathbf{I}+\sum_{i=1}^{j-1}\mathbf{H}_j\mathbf{M}_i\mathbf{M}_i^*\mathbf{H}_j^*\right)^{-1}\mathbf{H}_j\tilde{\mathbf{V}}_j^{(0)}\mathbf{M}_j'\right| \tag{24}$$

in Fig. 2. This is not necessarily surprising, given the fact that the BD solution only approaches the true sum capacity at high SNR. It also implies that at low SNR, SO can yield better performance than any BD solution.

The asterisks on the BD curves represent the sum capacity optimization, which provides solutions that are generally good. However, suppose that a rate point $R_1 = 14$ and $R_2 = 2$ is desired. With a total SNR of 20 dB, this could be achieved with SO by putting user 1 first but could not be achieved with BD. Additionally, SNR differences between users could have a similar effect, as illustrated in Fig. 2. In this case, the BD solution results in a rate region that is strongly biased toward user 1, but using SO with user two first results in a more balanced rate region.

For a system of K users, there are $K!$ sequentially optimized solutions, and an important question is how to choose the best possible ordering. An algorithm for choosing a good ordering must have a lower computational cost than the "brute force" approach of computing all possible solutions and still have a high probability of choosing the best ordering. Empirical tests have revealed that when users have a different number of antennas, the best solution frequently chooses the users with smaller numbers of antennas to be optimized first. Furthermore, as illustrated in the previous rate region plots, power savings can sometimes be obtained by choosing users with attenuated channels first. One approach that performs reasonably well, but at a significant computational cost, is to measure the degree of orthogonality between the spaces spanned by $\mathbf{H}_j$ and $\tilde{\mathbf{H}}_j$. If $\mathbf{W}_j$ and $\tilde{\mathbf{W}}_j$ are orthonormal bases for $\mathbf{H}_j$ and $\tilde{\mathbf{H}}_j$, and σ_{W_j} is the smallest nonzero singular value of $\mathbf{W}_j\tilde{\mathbf{W}}_j^*$, then $\theta_j = \cos^{-1}(\sigma_{W_j})$ is the minimum angle between the subspaces spanned by the two matrices. A reasonable approach would be to schedule the users in order of increasing $\theta_j n_{Rj}$, but this only provides a computational savings over finding all possible solutions when there are a moderate to large number of users (four or more). More work is needed to investigate better ordering schemes.

An additional possibility is to combine SO with BD in a hybrid scheme. For example, when one user is likely to require high priority (low SNR, high rate requirement, small number of antennas, etc), it would be scheduled first in the successive optimization. If the remaining users have less stringent requirements that are more or less equivalent, one could simply find a block diagonal solution for them, subject to the additional constraint that they do not interfere with the first user. Some of the results in the next section lend support to this idea.

V. Coordinated Transmit-Receive Processing

The BD and SO algorithms discussed thus far rely on the condition that $n_T \geq n_R$. In general, the transmitter can send n_T interference free data streams, regardless of the number of users. In this section, we propose a framework for extending the applicability of the BD and SO algorithms to up to n_T users, regardless of the users' array sizes, by coordinating the processing between the transmitters and receivers. Our approach is based on the work of [24] for the power control problem. In [24], it was assumed that all users employ MMSE receivers. Since the transmitter already knows the channels and the signals to be transmitted, it can predict what the MMSE coefficients for each receiver will be. One data subchannel is transmitted to each user (thus allowing n_T users), an initial set of receiver vectors are assumed, and the optimal transmitter and receiver vectors are alternatively recomputed until the solution converges to one with minimum power. To avoid the computational cost of an iterative approach and to allow for more than one data stream per user (for which no iterative solution has yet been proposed), we propose a fast alternative method that uses a reasonable initial receiver estimate followed by application of either the BD or SO algorithms. In addition to reducing computation, this allows a blockwise optimization of the transmit vectors for cases where multiple data subchannels can be used.

Let m_j be the number of spatial dimensions used to transmit to user j, and let $\mathbf{W}_j$ be an $m_j \times n_{Rj}$ matrix consisting of the m_j beamformers user j will employ in receiving data from the base. We now define a new block matrix $\overline{\mathbf{H}}_S$:

$$\overline{\mathbf{H}}_S = \begin{bmatrix} \overline{\mathbf{H}}_1 \\ \vdots \\ \overline{\mathbf{H}}_K \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1^*\mathbf{H}_1 \\ \vdots \\ \mathbf{W}_K^*\mathbf{H}_K \end{bmatrix}. \tag{27}$$

The matrix $\overline{\mathbf{H}}_S$ has dimensions that are compatible with either the BD or SO algorithms when $\sum m_j \leq n_T$. Using $\overline{\mathbf{H}}_S$ in place of $\mathbf{H}_S$ in either algorithm allows some interuser interference to be transmitted, but this interference is eliminated at the output of the receiver beamformers since it is steered into the nulls of the $\mathbf{W}_j$ beampatterns. The problem then becomes one of choosing m_j and the beamformers $\mathbf{W}_j$ for each user.

The number of subchannels m_j allocated to each user must obviously be 1 when $K = n_T$, assuming that all users are to be accommodated. The question is somewhat more difficult when $K < n_T$. In such a case, the additional degrees of freedom available to the transmitter can either be used to still send only one data stream to each user, but with an increased gain, or to allocate additional subchannels to some or all users. If n_T is not sufficient to allocate a secondary subchannel to all users, the question of which user(s) should be given additional subchannels will likely depend on the optimization to be performed. If system throughput is the primary concern, the optimal solution may likely be to give extra channels to stronger users. If power control is the goal, it may be more beneficial to give the users with weaker channels the extra subchannels. Space does not permit a detailed discussion of the resource allocation problem here, but this is a topic of significant current interest.

When the values of m_j have been determined, it is then necessary to determine the $\mathbf{W}_j$ matrices. The approach in [24] is to assume an initial set of $\mathbf{W}_j$ matrices and then iteratively compute $\mathbf{M}_S$ and $\mathbf{W}_j$, given the known receiver structure. To avoid the computational expense of an iterative solution, we propose the use of an intelligent initial value for the set of $\mathbf{W}_j$ matrices, followed by computation of the BD solution for the resulting $\overline{\mathbf{H}}_S$. As shown in the simulations, this approach can result in a near-optimal solution. An obvious candidate for $\mathbf{W}_j$, and the one we propose below, is to use the m_j dominant left singular vectors of $\mathbf{H}_j$. An outline of how coordinated transmit-receive processing can be used in conjunction with BD is given in the following algorithm description.

Coordinated Tx-Rx BD Algorithm

1) For $j = 1, \ldots, K$:
  Compute the SVD $\mathbf{H}_j = \mathbf{U}_j \mathbf{\Sigma}_j \mathbf{V}_j^*$.
2) Determine m_j, which is the number of subchannels for each user.
3) For $j = 1, \ldots, K$:
  Let $\mathbf{W}_j$ be the first m_j columns of $\mathbf{U}_j$.
  Calculate $\overline{\mathbf{H}}_j = \mathbf{W}_j^* \mathbf{H}_j$.
4) Apply the BD algorithm using $\overline{\mathbf{H}}_S$ in place of $\mathbf{H}_S$.

Note that since the beamformers $\mathbf{W}_j$ represent only a guess by the transmitter at the optimal receiver structure, they do not necessarily correspond to what the receiver will actually use. The optimal receiver will be the product of the first m_j columns of $\mathbf{U}_j$ from the BD algorithm and $\mathbf{W}_j$.

This coordinated processing can be used in conjunction with the SO algorithm as well by using SO in the place of the BD algorithm in step 4. We make the following observations. First, for channels with $m_j > 1$, the optimal receiver is no longer $\mathbf{W}_j$ but a combination of $\mathbf{W}_j$ and the left singular vectors from the second SVD in the BD algorithm. In addition, when $m_j = 1$ for all users, the BD simplifies to a weighted pseudo-inverse of $\overline{\mathbf{H}}_S$. The coordinated Tx-Rx algorithms simplify to the standard BD and SO algorithms, when dimensions permit, by initializing them with $\mathbf{W}_j = \mathbf{I}$.

In the simulation results that follow, we use coordinated processing with block diagonalization to compare the performance of a {4,4} × 4 channel for different numbers of subchannels per user.

VI. Simulation Results

In order to compare the maximum achievable throughput of the BD algorithm with other implementations, several special cases are considered. First, the number of antennas for each user (n_{Rj}) is held constant, so that for K users and $n_{Rj} = M$, the total number of receive antennas is $n_R = MK$. We consider in particular the {1,1,1,1} × 4 and {2,2} × 4 channels. All data were generated assuming the elements of $\mathbf{H}_S$ are independent complex Gaussian random variables with zero mean and unit variance.

As mentioned earlier, channel inversion is one method that has already been proposed for transmit vector selection [25]. For cases where $n_T \geq n_R$, this provides a solution that perfectly diagonalizes $\mathbf{H}_S$ subject to the constraint that equal power is transmitted to each receive antenna. For sake of comparison, the performance of this algorithm will be included in the plots that follow. To obtain the capacity of such a scheme, the transmit power must be scaled to meet the power constraint. Define $\mathbf{H}_S^\dagger$ as the pseudo-inverse of $\mathbf{H}_S$. Then, the modulation matrix that satisfies the power constraint P is

$$\mathbf{M}_S = \frac{\sqrt{P}}{\|\mathbf{H}_S^\dagger\|_F} \mathbf{H}_S^\dagger. \tag{28}$$

The maximum achievable rate (R_{PI}) for this scheme is

$$R_{PI} = \log_2 |\mathbf{I} + \mathbf{H}_S \mathbf{M}_S \mathbf{M}_S^* \mathbf{H}_S^*| \tag{29}$$

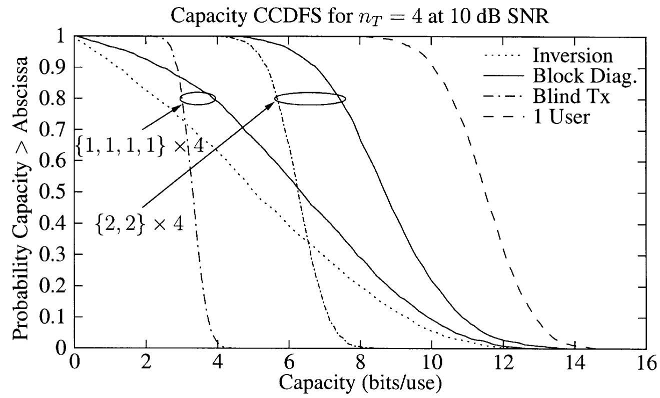

Fig. 3. Complementary cumulative distribution functions of sum capacity for Gaussian channels for four transmitters.

$$= \log_2 \left| \mathbf{I}n_R + \left(\frac{\rho}{\|\mathbf{H}_S^\dagger\|_F^2} \right) \mathbf{I}_L \right| \tag{30}$$

$$= L \log_2 \left[1 + \rho \left(\sum_{n=1}^{L} \sigma_{\mathbf{H}_S,n}^{-2} \right)^{-1} \right] \tag{31}$$

where L is the rank of $\mathbf{H}_S$, and $\sigma_{\mathbf{H}_S,n}$ is its nth singular value. Note that in this implementation of channel inversion, water-filling is not performed, and thus, all users are ensured an equal rate. The BD algorithm implemented with $n_{Rj} = 1$ reduces to channel inversion but with water-filling employed to maximize throughput. The plots that follow include results for both channel inversion and BD when $n_{Rj} = 1$, and any performance difference between the two can be attributed to the use of water-filling over equal-power transmission.

In the plots that follow, "Inversion" refers to the channel inversion algorithm of (28), "Block Diag" is the sum capacity BD algorithm of Section III.A, and "Blind Tx" is the capacity for the case where no channel information is available and the users are time multiplexed. As SNR $\rightarrow \infty$, we expect the achievable throughput of the BD algorithm to approach the sum capacity for $\mathbf{H}_S$.

Fig. 3 compares the probability distributions of sum capacity for the {1,1,1,1} × 4, {2,2} × 4, and single-user 4 × 4 channels. The SNR is 10 dB, and all channels are independent and identically distributed (IID) Gaussian. There is only one line representing the channel inversion algorithm because its performance is identical for any configuration with the same total n_R and n_T. This does not apply for simulations presented later, when the spatial correlation of the receive antennas is taken into account. It is interesting to note in Fig. 3 that at low outage probabilities, the case where each receiver has only one antenna produces better results when channel knowledge is not assumed and the users are simply time multiplexed. For the case of two antennas at each receiver, the average capacity gain derived from exploiting channel knowledge using the BD algorithm is around 30%. Note that BD outperforms channel inversion at all outage probabilities.

Fig. 4 shows the capacity as a function of the transmitter array size with the outage probability fixed at 0.1. The capacity gains of the BD algorithm are quite sizable here, up to a factor of 4 for the {1,1,1,1} $\times n_T$ channel, and a factor of 2 for the {2,2} $\times n_T$ channel. This is due to the ability of the BD algorithms to opti-

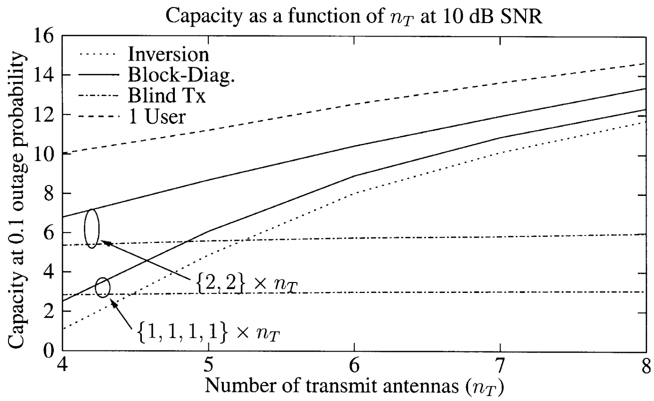

Fig. 4. Capacity as a function of transmitter array size at a SNR of 10 dB.

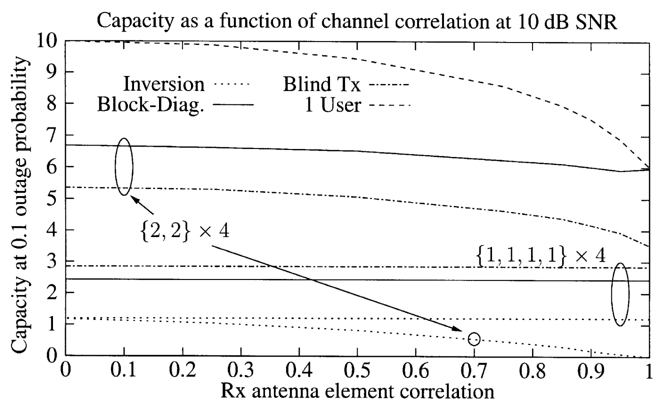

Fig. 5. Capacity as a function of channel correlation between Rx antennas at an SNR of 10 dB.

mally use the excess degrees of freedom available at the transmitter.

Fig. 5 shows the variation in performance as a function of channel spatial correlation. For this case, we illustrate the effects of correlated receive antennas but not transmit antennas. This is a realistic scenario in which the base station has significantly separated elements, but the mobile terminals have closely spaced antennas. The channels for different users are assumed to be uncorrelated. In order to reduce the effect of spatial correlation to a single parameter, each column of $\mathbf{H}_j$ is assumed to have covariance $\mathbf{R}$, with elements $R_{i,j} = \alpha^{|i-j|}$, where $0 < \alpha < 1$ is represented on the horizontal axis in the plot. The channel inversion algorithm now has two curves because the channel matrices for each user are independent, resulting in a completely independent channel $\mathbf{H}_j$ for the {1,1,1,1} $\times$ 4 case and a partially correlated $\mathbf{H}_S$ matrix for the {2,2} $\times$ 4 case (for no correlation, as was assumed in the previous figures, the $\mathbf{H}_S$ matrices are statistically identical). For the {2,2} $\times$ 4 case, as the channel becomes completely correlated, the capacity of the BD solution decreases slightly but less than the other algorithms.

Fig. 6 illustrates the performance of the BD algorithm for the case of partial channel information. Channels were generated for this example using angle-of-arrival information, as described by (16). We assume that only $\mathbf{\Phi}_{T,j}$ is known to the transmitter for each j and that it is used for the value of $\mathbf{B}_j$ in the BD algorithm. For the Monte Carlo trials used in this simulation, all angles of arrival are independent and uniformly distributed, and all multipath gains gains were generated as

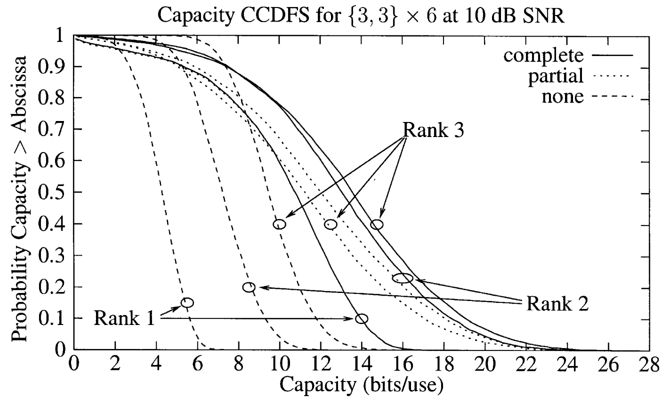

Fig. 6. Capacity CCDFs for different cases of partial channel information.

IID complex Gaussian random variables. The plot in Fig. 6 contains data for a {3,3} $\times$ 6 channel. Three algorithms are compared: first, BD with complete channel knowledge (labeled "complete"), second, BD with partial channel knowledge (labeled "partial"), and third, TDMA without any channel knowledge (labeled "none"). The results for complete and partial channel knowledge for rank-1 channels are close enough to be indistinguishable in the plot. It can be seen that as the rank of the channel decreases, the performance difference between full and partial channel knowledge decreases. In the rank 1 case, at a 10% outage probability, channel information (complete or partial) enables nearly double the capacity. At the same outage rate, both complete and partial channel knowledge provide a modest gain in capacity for rank-2 channels, but for full rank channels, partial information in this case provides no increase in capacity.

Fig. 7 shows the performance of SO for different ordering algorithms, together with the performance of BD. "Optimal" ordering is found by a global search, "Angle Algorithm" refers to ordering with increasing $\theta_j n_{Rj}$, as explained in Section IV, "Frobenius Norm" refers to ordering according to the Frobenius norm of $\mathbf{H}_j$ (so smaller $\mathbf{H}_j$ will tend to go first), and "Random" means random ordering. In all cases, there were six transmit antennas and three users. Fig. 7(a) shows the results for the {2,2,2} $\times$ 6 channel, and Fig. 7(b) shows results for a {1,2,3} $\times$ 6 channel. The fact that BD achieves better performance than even the best SO algorithm supports the idea of hybrid optimization mentioned at the end of the last section. It is obvious that the Frobenius norm, while simple to compute, is not a very good indicator for ordering (even worse than random ordering for equal array sizes), but the angle algorithm yields acceptable performance in both cases.

Fig. 8 compares some of the previous results with the performance of coordinated transmit-receive processing, using complementary cumulative density functions (CCDFs) similar to those in Fig. 3. Included for reference are the inversion and BD curves for the {2,2} $\times$ 4 channel. The {4,4} $\times$ 4 channel uses coordinated Tx-Rx processing with either one or two subchannels per user, labeled in the figure as "1 SC" or "2 SC," respectively. For the case of a single subchannel per user, we have shown the results of using an iterative approach as well (labeled "it." in the plot). The iterative algorithm was implemented using maximal ratio combining ($\mathbf{w}_j = \mathbf{H}_j \mathbf{m}_j$), and it alternates between

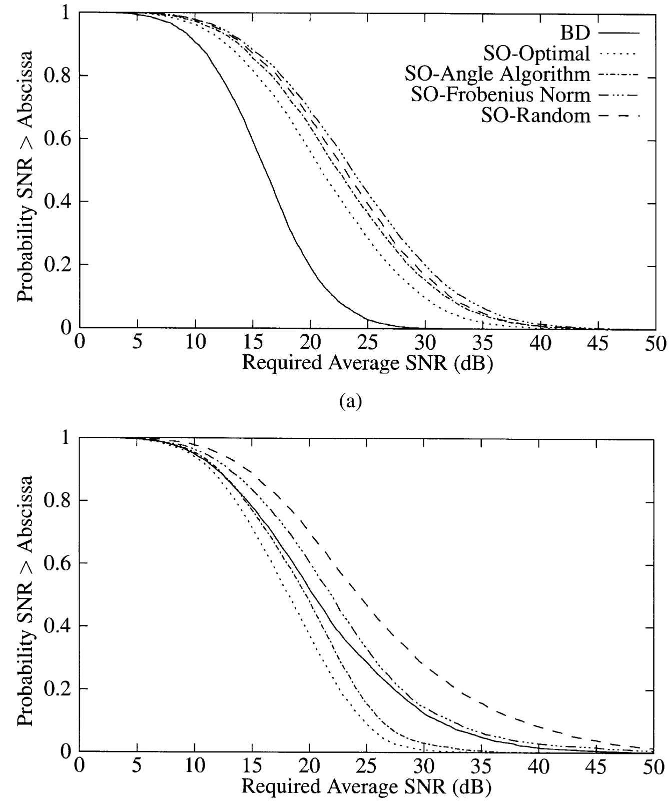

Fig. 7. Performance of successive optimization compared with BD for $n_T = 6$, random rate points in the interval [2], [8], and random channel gains in the interval [− 6, 6] dB.

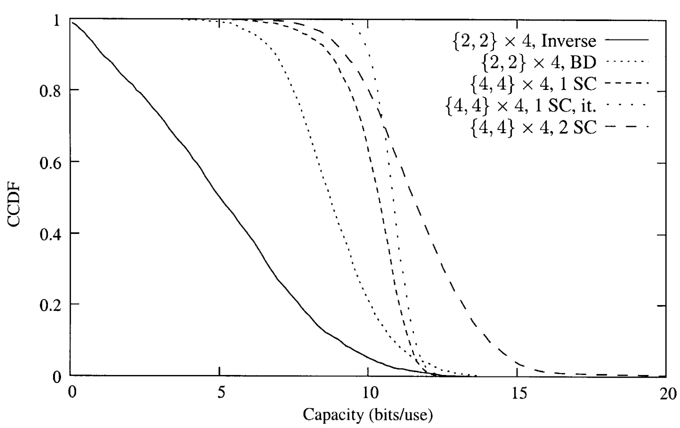

Fig. 8. Comparison of probability densities of capacity for different channel geometries and channel decomposition algorithms at a system SNR of 10 dB.

updating the receiver and transmitter vector until convergence (using $\|\mathbf{M}_{S,n-1} - \mathbf{M}_{S,n}\|_F$ as a convergence metric). This approach did not converge in our simulations when multiple subchannels per user were assigned. There are also potential numerical problems with such an approach even for single-channel cases if there is high correlation between users or if the channels are rank deficient. The iterative approach here shows some small gains in performance, but it is inferior in some cases to the noniterative two-subchannel approach, illustrating the benefit of the "block optimization" that characterizes the BD and SO algorithms.

VII. Conclusions

Two new approaches for optimizing information transfer in a multiuser channel have been presented here. Both are suboptimal in that they do not perfectly achieve the sum capacity of the channel, but the block diagonalization algorithm asymptotically approaches capacity at high SNR. The successive optimization algorithm is better suited to the problem of minimizing power output for a fixed set of transmission rates than it is to the problem of maximizing throughput for fixed power. In low SNR channels, it often performs better than block diagonalization, and it appears to also be a good choice for channels where users have different power levels or rate requirements. Both algorithms provide a straightforward, computationally efficient method of choosing "optimized" downlink transmit vectors and allow for a good tradeoff between performance and computational complexity. For channels whose dimensions will not support the block diagonalization or successive optimization algorithms directly, joint transmitter-receiver processing can be used to reduce the dimensionality of the problem so that these methods can be used. All of the algorithms have a fixed computational cost that is a function of the dimensions of the users' channel matrices. For a system with K users, the BD and SO algorithms both require $2K$ SVDs, and the joint transmitter-receiver version of the BD algorithm can require as many as $3K$ SVDs. Many of the alternatives are iterative algorithms for which the computational cost will be higher and cannot be known in advance. The algorithms presented here all have the advantage of a fixed computational cost and provide a sufficient performance advantage to justify the cost.

All of the algorithms presented require partial or complete knowledge of the channel at the transmitter. Past studies for the single-user channel have demonstrated that the gain from having such knowledge at the transmitter is often small, particularly at high SNR. In the multiuser case, however, the performance gap is much larger, and it increases rather than decreases as the SNR becomes large or as the number of transmit antennas grows. This may make the potentially high cost of obtaining channel knowledge at the transmitter more justifiable.

References

[1] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal blast processing with multiple-access interference," in *Proc. IEEE VTC Fall Conf.*, vol. 1, Boston, MA, Sept. 24–28, 2000, pp. 87–91.

[2] R. S. Blum, "MIMO capacity with interference," in *Proc. Conf. Inform. Sci. Syst.*, Mar. 2002.

[3] B. G. Agee, "Exploitation of internode MIMO channel diversity in spatially distributed multipoint communication networks," in *Proc. Asilomar Conf.*, Nov. 2001.

[4] R. W. Heath, M. Airy, and A. J. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Proc. Asilomar Conf.*, Nov. 2001.

[5] M. F. Demirkol and M. A. Ingram, "Power-controlled capacity for interfering MIMO links," in *Proc. IEEE VTC Fall Conf.*, vol. 1, Atlantic City, NJ, Oct. 7–11, 2001, pp. 187–191.

[6] R. S. Blum, J. H. Winters, and N. Sollenberger, "On the capacity of cellular systems with MIMO," in *IEEE VTC Fall Conf.*, vol. 2, Atlantic City, NJ, Oct. 7–11, 2001, pp. 1220–1224.

[7] J. Li, K. B. Letaief, M. A. Zhengxin, and Z. Cao, "Spatial multiuser access with MIMO smart antennas for OFDM systems," in *IEEE VTC Fall Conf.*, vol. 3, Atlantic City, NJ, Oct. 7–11, 2001, pp. 1553–1557.

[8] P. Viswanath, D. N. C. Tse, and V. Anantharam, "Asymptotically optimal waterfilling in vector multiple access channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 241–267, Feb. 2001.

[9] D. C. Popescu, O. Popescu, and C. Rose, "Interference avoidance for multiaccess vector channels," in *Proc. Int. Symp. Inform. Theory*, July 2002, p. 499.

[10] J. Wang and K. Yao, "Multiuser spatio-temporal coding for wireless communications," in *Proc. IEEE Wireless Commun. Networking Conf.*, vol. 1, Mar. 2002, pp. 276–279.

[11] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[12] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[13] G. J. Foschini and M. J. Gans, "On limits of wireless personal communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[14] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich scattering wireless channel," in *Proc. ISSSE*, Sept. 1998.

[15] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, submitted for publication.

[16] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inform. Theory*, submitted for publication.

[17] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[18] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.

[19] M. Schubert and H. Boche, "An efficient algorithm for optimum joint beamforming and power control," in *Proc. IEEE VTC*, vol. 4, 2002, pp. 1911–1915.

[20] H. Boche and M. Schubert, "Optimum SIR balancing using extended 1-Norm beamforming optimization," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 2002, pp. 2945–2948.

[21] M. Bengtsson and B. Ottersten, "Optimal and suboptimal beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed. Boca Raton: CRC, 2001.

[22] M. Bengtsson, "A pragmatic approach to multi-user spatial multiplexing," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Aug. 2002.

[23] K.-K. Wong, R. D. Murch, R. S.-K. Cheng, and K. B. Letaief, "Optimizing the spectral efficiency of multiuser MIMO smart antenna systems," in *Proc. Wireless Commun. Networking Conf.*, vol. 1, 2000, pp. 426–430.

[24] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint transmitter receiver diversity for efficient space division multiaccess," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 16–27, Jan. 2002.

[25] W. Qiu, H. Tröger, and M. Meurer, "Joint transmission (JT) in multi-user MIMO transmission systems," EURO-COST, COST 273 TD(02)008, 2002.

[26] D. W. Bliss, K. W. Forsythe, A. O. Hero, and A. L. Swindlehurst, "MIMO environmental capacity sensitivity," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Oct. 2000.

[27] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, pp. 1740–1751, Feb./Mar./Apr. 1994.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.

[29] Q. H. Spencer and A. L. Swindlehurst, "Some results on channel capacity when using multiple antennas," in *Proc. VTC Fall Conf.*, 2000.

[30] M. T. Ivrlač and J. A. Nossek, "Correlated fading in MIMO-systems-blessing or curse?," in *Proc. 39th Annu. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2001.

Quentin H. Spencer (S'00) received the B.S. and M.S. degrees in electrical engineering from Brigham Young University (BYU), Provo, UT, in 1994 and 1996, respectively. He is currently pursuing the Ph.D. degree at BYU.

From 1995 to 1996, he was a Research Assistant with the Department of Electrical Engineering, BYU, working on indoor channel propagation modeling. From 1996 to 1997, he was an Engineer with L-3 Comunnunications, Salt Lake City, UT, working on broadband military communications systems. Since 1998, he has been a research assistant at BYU, where his interests are in signal processing methods for MIMO channels. From 1999 to 2001, he was a control systems engineer with UniDyn Corporation, Orem, UT, and in 2001, he was a signal processing engineer with Inari Corporation, Draper, UT, working on powerline communications systems.

A. Lee Swindlehurst (SM'99) received the B.S., *summa cum laude*, and M.S. degrees in electrical engineering from Brigham Young University (BYU), Provo, UT, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1991.

From 1983 to 1984, he was with Eyring Research Institute, Provo, as a scientific programmer. From 1984 to 1986, he was a Research Assistant with the Department of Electrical Engineering, BYU, working on various problems in signal processing and estimation theory. He was awarded an Office of Naval Research Graduate Fellowship from 1985 to 1988, and during most of that time, he was affiliated with the Information Systems Laboratory at Stanford University. From 1986 to 1990, he was also with ESL, Inc., Sunnyvale, CA, where he was involved in the design of algorithms and architectures for several radar and sonar signal processing systems. He joined the faculty of the Department of Electrical and Computer Engineering, BYU, in 1990, where he now holds the position of Full Professor and is currently serving as Department Chair. From 1996 to 1997, he held a joint appointment as a visiting scholar at both Uppsala University, Uppsala, Sweden, and at the Royal Institute of Technology, Stockholm, Sweden. His research interests include sensor array signal processing for radar and wireless communications, detection and estimation theory, and system identification.

Dr. Swindlehurst is currently serving as Secretary of the IEEE Signal Processing Society, as a member of the Sensor Array and Multichannel Signal Processing Technical Committee in the same society, and is a past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the Technical Program Chair for the 1998 IEEE Digital Signal Processing Workshop and for the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. He is also a recipient of the 2000 IEEE W. R. G. Baker Prize Paper Award and is co-author of a paper that received a Signal Processing Society Young Author Best Paper Award in 2001.

Martin Haardt (S'90–M'98–SM'99) studied electrical engineering at the Ruhr University Bochum, Bochum, Germany, and at Purdue University, West Lafayeete, IN. He received the Dipl.-Ing. degree from the Ruhr University Bochum in 1991 and the Dok.-Ing. degree from Munich University of Technology, Munich, Germany, in 1996.

He is a Full Professor and Head of the Communications Research Laboratory at Ilmenau University of Technology, Ilmenau, Germany. In 1997, he joined Siemens Mobile Networks, Munich, where he was responsible for strategic research for third-generation mobile radio systems. He was the Director for International Projects and University Cooperations in the mobile infrastructure business of Siemens, Munich, where his work focused on mobile communications beyond the third generation. During his time at Siemens, he also taught in the international Master of Science in Communications Engineering program at Munich University of Technology. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, and numerical linear algebra.

Dr. Haardt received the Mannesmann Mobilfunk Innovations Award for outstanding research in mobile communications, the ITG Best Paper Award from the Association of Electrical Engineering, Electronics, and Information Technology (VDE), and the Rohde & Schwarz Outstanding Dissertation Award. Currently, he serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.