2004-03-18

Transmission Strategies for Wireless Multi-user, Multiple-Input, Multiple-Output Communication Channels

Quentin H. Spencer
Brigham Young University - Provo

TRANSMISSION STRATEGIES FOR WIRELESS MULTI-USER, MULTIPLE-INPUT, MULTIPLE-OUTPUT COMMUNICATION CHANNELS

by

Quentin H. Spencer

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

Brigham Young University

April 2004

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Quentin H. Spencer

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	A. Lee Swindlehurst, Chair
Date	Randal W. Beard
Date	Brian D. Jeffs
Date	Michael A. Jensen
Date	Michael D. Rice

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Quentin H. Spencer in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

A. Lee Swindlehurst
Chair, Graduate Committee

Accepted for the Department

Michael A. Jensen
Graduate Coordinator

Accepted for the College

Douglas M. Chabries
Dean, College of Engineering and Technology

ABSTRACT

TRANSMISSION STRATEGIES FOR WIRELESS MULTI-USER, MULTIPLE-INPUT, MULTIPLE-OUTPUT COMMUNICATION CHANNELS

Quentin H. Spencer

Electrical and Computer Engineering

Doctor of Philosophy

Multiple-Input, Multiple-Output (MIMO) processing techniques for wireless communication are of interest for next-generation systems because of their potential to dramatically improve capacity in some propagation environments. When used in applications such as wireless LAN and cellular telephony, the MIMO processing methods must be adapted for the situation where a base station is communicating with many users simultaneously. This dissertation focuses on the downlink of such a channel, where the base station and all of the users have antenna arrays. If the transmitter has advance knowledge of the users' channel transfer functions, it can use that information to minimize the inter-user interference due to the signals that are simultaneously transmitted to other users. If the transmitter assumes that all receivers treat the interference as noise, finding a solution that optimizes the use of resources is very difficult. This work proposes two classes of solutions to this problem. First, by forcing some or all of the interference to zero, it is possible to achieve a sub-optimal solution in closed-form. Second, a class of iterative solutions can

be derived by extending optimal algorithms for multi-user downlink beamforming to accommodate receivers with multiple antennas. The closed-form solutions generally require less computation, but the iterative solutions offer improved performance are more robust to channel estimation errors, and thus may be more useful in practical applications. The performance of these algorithms were tested under realistic channel conditions by testing them on channels derived from both measurement data and a statistical model of an indoor propagation environment. These tests demonstrated both the ability of the channel to support multiple users, and the expected amount of channel estimation error due to movement of the users, with promising results. The success of any multi-user MIMO processing algorithm is ultimately dependent on the degree of correlation between the users' channels. If a base station is required to support a large number of users, one way to ensure minimal correlation between users' channels is to select groups of users whose channels are most compatible. The globally optimal solution to this problem is not possible without an exhaustive search, so a channel allocation algorithm is proposed that attempts to intelligently select groups of users at a more reasonable computational cost.

ACKNOWLEDGMENTS

In addition to his helpful expertise and valuable feedback on my work, my advisor, Lee Swindlehurst, also deserves credit for first encouraging me to consider pursuing a Ph.D. I am also very thankful to my wife, Heather, for her continual support and encouragement throughout this lengthy process.

Contents

List of Figures

Chapter 1

Introduction

In the 55 years since Claude Shannon's groundbreaking work, "A Mathematical Theory of Communications" [1], created the science of information theory, much effort has been devoted to discovering ways of achieving what Shannon showed to be theoretically possible. The invention of block codes, convolutional codes, trellis codes, and most recently turbo codes [2] have each made incremental performance improvements to the point that it is now possible to come quite close to the "Shannon bound". In modern wireless communication systems, a particular challenge has been to design coding systems that can adequately compensate for the effects of multipath propagation, which leads to fading and inter-symbol interference.

In addition to coding techniqes, array processing has long been an effective tool in improving the performance of wireless communications systems. The earliest application of array processing methods to wireless systems was the use of arrays at the receiving end of the link, a scenario referred to as a single-input, multiple-output (SIMO) channel. The use of simple adaptive array algorithms in this type of a channel is an effective way of improving the SNR of the channel without any change in the transmitted power, and it has the ability to eliminate co-channel or multipath interference. The same methods generally apply as well to the multiple-input, single-output (MISO) channel, where the transmitter has an array, and the receiver does not. Until recently, this was less common because it generally requires that the transmitter have advance knowledge of the channel, often referred to as side information or channel state information (CSI), which is not always practical.

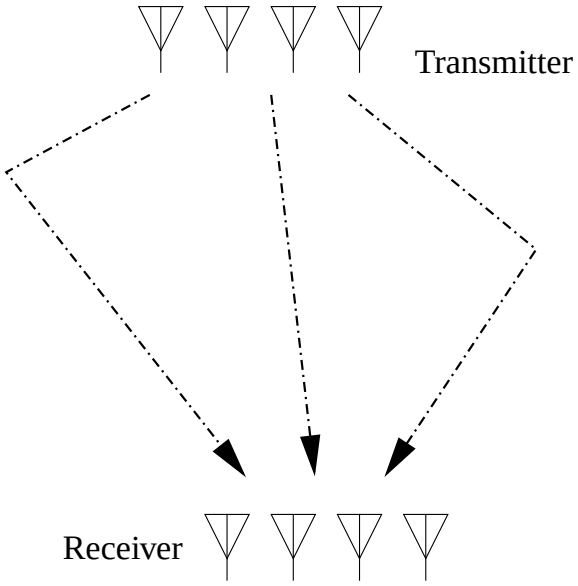

Figure 1.1: An illustration of a MIMO channel.

1.1 MIMO Channels

Beginning around 1998, several researchers demonstrated that the use of multiple-input, multiple-output (MIMO) channels could result in dramatic gains in channel capacity compared to single-input, single-output (SISO), SIMO or MISO channels [3–6]. The key to this is the use of parallel data transmission. While the SIMO and MISO channels could improve the effective gain of a channel and cancel interference, they still only transmitted a single data stream. The MIMO channel, on the other hand, could be used to transmit data in parallel, resulting in a capacity increase that is a linear rather than logarithmic function of array size. However, this is only possible in channels with significant multipath components. In a channel with a single transmission path, only one data stream can be transmitted, and the traditional rules of array processing apply, but in channels with multipath, each path can be used to transmit separate data streams [6]. Thus, while traditional array processing algorithms had been used to mitigate the effects of multipath interference, MIMO channels can potentially use multipath as an advantage, rather than a disadvantage. This is illustrated in Figure 1.1, where a MIMO channel with 4 transmitters and receivers has 3 multipath components available.

The various methods of transmitting data in MIMO channels can be classified in terms of what CSI available at the transmitter [7]. For the "informed" transmitter, or the transmitter that has complete CSI available, the capacity is shown in [6] to be achieved by a singular value decomposition (SVD) of the channel matrix, where the transmitter uses the right singular vectors with proper power weighting as its beamforming vectors, and the receiver uses the left singular vectors to estimate the transmitted signal. This decomposes the channel into a set of orthogonal channels whose optimal power coefficients can be determined by the well-known solution to the parallel Gaussian channels problem [8]. On the other hand, when the transmitter does not have CSI, the best that can be done is transmission of equal power through all transmitting antennas, leaving the work of estimating and inverting the channel effects to the receiver [4]. While it is useful to understand the differences between channels with and without CSI at the transmitter, performance comparisons of the two channels [9, 10] have shown that the performance of the "blind" transmitter quickly approaches the performance of an informed transmitter at moderate SNR, and that there is only a significant performance difference at very low SNR. This is problematic for two reasons: first, many applications of wireless communications systems will require operating in a higher SNR region, and second, operating at a low SNR makes is much more difficult to obtain a good estimate of the channel. Thus, for practical applications involving single-user MIMO channels, it is difficult to justify the cost of obtaining channel information at the transmitter, and it is worthwhile to focus on transmission methods that do not require CSI.

An alternative approach to the MIMO channel is to use the additional antennas to improve diversity while transmitting only one data stream. It has been shown that a $n_T \times n_R$ channel has diversity order $n_T n_R$. Methods of doing this include the codes based on "orthogonal designs" [11, 12], which are able to achieve near optimal diversity at both the transmitter and receiver. A review and comparison of these two differing transmission approaches can be found in [13].

1.2 The Multi-User Channel

We have considered MIMO systems in the context of a single-user point-to-point channel. However, many of today's wireless communications systems, such as cellular and wireless LAN systems, are based on a central hub or base station that simultaneously communicates with a group of users. Traditionally, in multi-user systems, multiple access has been achieved by either time-division, frequency-division, or code-division multiplexing, known respectively as TDMA, FDMA, and CDMA. The the use of arrays at the base station of such systems has resulted in the idea of Space-Division Multiple Access (SDMA), in which the spatial diversity of the signals received at the base station are used to separate signals that may be transmitted using the same time, frequency, or code sequence.

The uplink and downlink of a multi-user channel each have slightly different challenges, and are thus often treated separately in the literature. The downlink, where the base is transmitting to a group of users is often referred to as the "broadcast channel", and the uplink, where the base is receiving signals from a group of users, is often referred to as the "multiple access channel". A considerable number of recent publications have studied the application of MIMO processing methods to both the broadcast and multiple access channels [14–20].

The simplest multi-user MIMO channel is where the base has an array of antennas, but each user only has one antenna. A more challenging problem to consier is when all users are allowed to have an arbitrary number of antennas. The capacity of the uplink, also known as the vector multiple access channel (where arrays are employed at the transmit and possibly all receive nodes in the network) has been studied in [21–23], and its connection with the broadcast channel has been explored in [24]. 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 [4–6]. The capacity of the broadcast channel has been studied recently in [25, 26] for the special case where each user has only one antenna, and in [27] 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

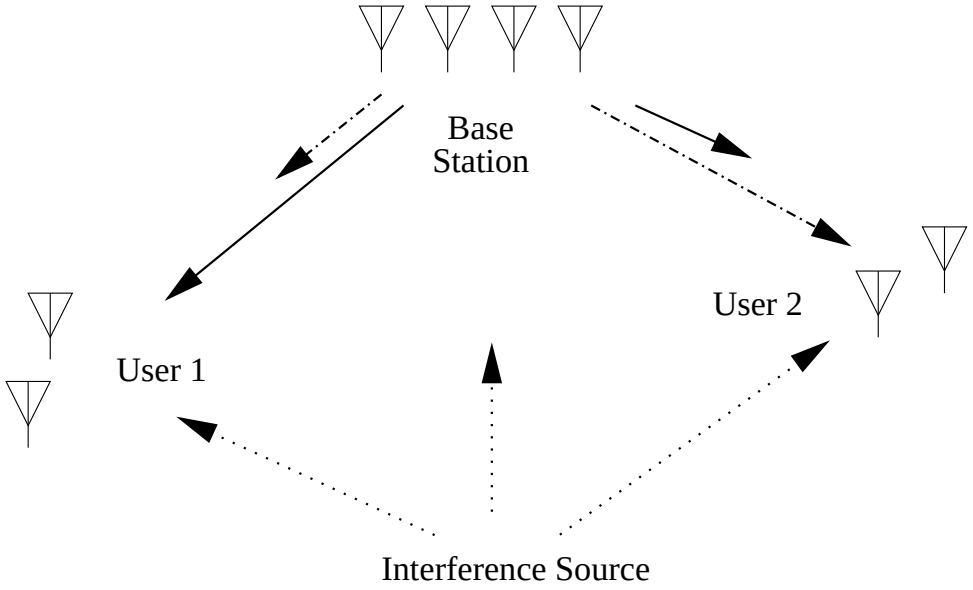

Figure 1.2: An illustration of a multi-user MIMO channel.

paper coding" [28]. Costa showed that when a communication channel is subject to interference that is known to the transmitter, the capacity is unchanged by the interference. This is achieved when the transmitter uses coding techniques that avoid the interference [29]. In multi-user transmission channels, the interference can be regarded as the signal intended for other users, which is known to the transmitter, so in principle a precoder could be used to avoid its effects. The primary drawback of such schemes is that their use of non-traditional coding leads to increased complexity at both the transmitter and receiver. Alternatives to the dirty-paper approach have been proposed [30–33], but so far they have generally only dealt with the special case of when all users employ single antennas. An illustration of a multi-user MIMO channel is shown in Figure 1.2. In this example, each user receives some interference from the signal intended for the other user, as well as some interference from an external source. While the external interference is not considered here, it is possible to use CSI at the transmitter to avoid the inter-user interference.

There are some specific optimization problems that are of particular interest when designing transmission algorithms for MIMO channels. The most fundamental is the capacity problem, or maximization of throughput subject to a constraint on the total transmitted power. This is important to understanding the capabilities of a channel, but

often a system designer will be more concerned with solving the closely related "power control" problem: minimize total transmitted power subject to satisfying a constraint on the data transmission rate and error rate. For a single-user channel, these two problems are virtually identical. For the multi-user channel, the problem is more complex. To achieve (sum) capacity in a multi-user 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 pre-specified 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. A third optimization problem related to power control has also been discussed in [30]. In this case, the transmit power is fixed and the transmitter attempts to maximize the amount by which the system exceeds the SINR requirement for all users, such that the SINR margin is the same for all users.

The problem of optimizing throughput for multi-user MIMO systems where each user may have multiple antennas has been approached in two different ways so far. The first [34] employs an iterative method of canceling out inter-user interference, allowing multiple data sub-channels per user as in classical MIMO transmission methods. The second approach [35] generalizes the single-antenna algorithms to include beamforming at the receiver, while still using only a single data sub-channel per user. The iterative nature of these algorithms typically results in a high computational cost.

1.3 Summary and New Contributions

This dissertation focuses on the downlink multi-user MIMO channel. In contrast to the dirty-paper algorithms, which take an information theoretic approach and emphasize new coding methods designed to achieve the channel capacity, the algorithms here generally take a signal processing approach, attempting to solve the various optimization problems by choosing good transmit vectors, independent of the particular signaling and coding methods in use. In the long term, it is likely that dirty-paper methods will be developed that are superior in performance and have acceptable complexity. However, because

they require all new code designs, they are not usable with current communications standards and protocols. As a result, the signal processing algorithms proposed here show greater promise over the short to medium term because they can be seamlessly integrated with existing communications systems to improve capacity. Chapter 2 introduces the general MIMO problem and the multi-user MIMO downlink problem and discusses the previous work in this area in greater detail. While it is mostly a review, it also includes the new contribution of a derivation of single-user MIMO capacity for some situations where the transmitter has partial channel information (originally published in [36]).

Chapter 3 focuses on methods of solving both the throughput maximization and power control problems in closed form. The most basic way of achieving this is nulling out all inter-user interference, which has been referred to here as "Block-Diagonalization". This concept was recently introduced in the literature [37–39], though it is developed in more detail here, including showing how it can be adapted to solve both the throughput maximization and power control problem, and how it can be adapted for cases where the transmitter has only partial channel knowledge. A second approach that allows some inter-user interference is also proposed, and proves to be more effective in some channels. In addition, a non-iterative approach for coordinated transmitter-receiver processing is proposed. These contributions are published in [40].

Chapter 4 focuses on iterative solutions to the same problems. The closed-form solutions have the advantage of reduced computational cost, but iterative solutions allow greater flexibility in the channel dimensions that can be supported, and can offer improved performance. The new contribution of this chapter is a hybrid algorithm that combines the zero-forcing methods of Chapter 3 with existing interference-balancing solutions which allows arbitrary array sizes at both transmitters and receivers, multiple data streams transmitted to individual users, and corrupted channel estimates at the transmitter. In addition, there are some computational reductions that can be achieved in special cases. These new contributions appear separately in [41, 42].

Chapter 5 applies some of the new transmission methods to realistic channel conditions. Up to this point, most of the literature has used idealized assumptions about channel statistics to test the algorithms. Here we consider channels from two sources:

channel measurements and statistical models. Channel measurements are useful because they give the most accurate possible representation of channel behavior, but they can be time-consuming and expensive to collect. Statistical models can be a useful alternative because they make it easy to generate a large number of test cases which can closely match behaviour of real channels when the model is designed well. This work also appears separately in [43]

Chapter 6 deals with the channel allocation problem in the context of the spatial multiplexing methods from Chapters 3 and 4. The general channel allocation problem for cellular systems has been studied extensively, and recent work has been done to study the effects of using adaptive arrays at the base stations. Usually, adaptive arrays are proposed as a means of increasing channel re-use in neighboring cells. In this chapter, we consider the challenge of channel re-use within a single cell, and how to appropriately decide which users should share channels so as to avoid putting users together that have highly correlated channels. If users with antenna arrays are considered, the problem of resource allocation from a system point of view is very complex. This appears to be a completely open problem at present, and is likely to be an important area of future research. This chapter presents a discussion of some of the issues involved, and proposes an algorithm for finding a good allocation of resources. These results will appear in [44, 45]

Chapter 2

The Downlink Multiplexing Problem

This chapter provides some background on the multi-user MIMO problem, including mathematical models, capacity, and existing transmission algorithms already in the literature. We begin with a detailed discussion of the single-user MIMO problem, and then show how the models, capacity derivations, and transmission schemes have been extended in the multi-user case.

2.1 The Single-User MIMO Problem

2.1.1 Channel Model

The MIMO channel is typically modeled using a matrix multiplication. Consider a MIMO channel with n_T transmitters and n_R receivers. If the transmitted signal is the n_T dimensional vector $\mathbf{s}$, and the received signal is the n_R dimensional vector:

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}\,. \tag{2.1}$$

where $\mathbf{n}$ is an additive noise term and $[\mathbf{H}]_{i,j}$ represents the transfer function from the j^{th} transmitter to the i^{th} receiver. The use of a single coefficient assumes a flat-fading or narrowband channel. The narrowband assumption is assumed to hold when the delay spread is small compared to the symbol rate of the transmitted signal. When this is not true, the broadband or frequency selective fading channel suffers from inter-symbol interference and a fading characteristic that varies significantly across the frequency band. For a traditional single-input single-output (SISO) channel, this is typically modeled using a convolution

operation. The MIMO model can also be easily generalized for frequency selective fading using a block structure [6] for $\mathbf{H}$ that combines the matrix channel with a convolution:

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{n_R} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,n_T} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{n_R,1} & \dots & \mathbf{H}_{n_R,n_T} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{n_T} \end{bmatrix} + \mathbf{n}\,. \tag{2.2}$$

In this case each of the blocks is a convolution matrix of dimension $(N+\nu) \times N$, the transmitted signal has dimension Nn_T, and the received signal and noise vectors have dimension $(N+\nu)n_R$, where N is the block length for transmission, and ν is the maximum delay spread in samples. While the algorithms presented here will generally assume narrowband fading, using this structure they could easily be generalized for wideband fading, although the computational cost would certainly be increased.

The channel matrix can also be expressed as the sum over L multipath components [6]:

$$\mathbf{H} = \sum_{l=1}^{L} \beta_l \begin{bmatrix} a_{R,1}(\theta_{R,l})\mathbf{I}_{N+\nu} \\ \vdots \\ a_{R,n_R}(\theta_{R,l})\mathbf{I}_{N+\nu} \end{bmatrix} \mathbf{G}_l \begin{bmatrix} a_{T,1}(\theta_{T,l})\mathbf{I}_N & \dots & a_{T,n_T}(\theta_{T,l})\mathbf{I}_N \end{bmatrix}, \tag{2.3}$$

where $\mathbf{G}_l$ is the convolution matrix

$$\mathbf{G}_l = \begin{bmatrix} g_l(0) & & & & \mathbf{0} \\ \vdots & \ddots & & & \\ g_l(\nu) & & \ddots & & \\ & & \ddots & & g_l(0) \\ & & & \ddots & \vdots \\ \mathbf{0} & & & & g_l(\nu) \end{bmatrix}, \tag{2.4}$$

which represents the pulse shaping function when sampled at the time delay associated with the l^{th} multipath component. The coefficients $a_{R,1}(\theta) \dots a_{R,n_R}(\theta)$ represent the steering vector at the receiver at an arbitrary angle θ, and $a_{T,1}(\phi) \dots a_{T,n_T}(\phi)$ are the corresponding steering vector at the transmitter at an angle ϕ. The angles $\theta_{R,l}$ and $\theta_{T,l}$ respectively represent the angles of arrival and departure for the l^{th} path. The gain of the l^{th} path is β_l. In [6] it is proven that the rank of $\mathbf{H}$ has an upper bound of

$$\min\{NL, (N+\nu)n_R, Nn_T\}\,.$$

In the narrowband case, we can assume a block size of 1, and that ν is 0, so the rank becomes $\min\{L, n_R, n_T\}$. The rank of $\mathbf{H}$ is important because of its close connection to capacity, which will be shown in the next section.

One additional property of radio propagation channels that must also be considered in the MIMO context is how they vary over time. This is particularly important for applications that assume mobility of one or both ends of the channel. Two applications for which MIMO transmission has been considered are wireless LAN environments and cellular environments. Wireless LANs are a natural fit because of the rich multipath environment in the indoor channel [46, 47], and because a laptop computer is large enough to easily fit several antennas. In this type of channel, mobility speeds are likely to be quite slow, so the channel can be viewed as very slowly time-varying. Cellular channels are also a possible application of MIMO transmission, but they are slightly more challenging. Firstly, the smaller size of the mobile devices will make it difficult to fit multiple antennas, and secondly, the mobility can happen at much faster speeds. In this dissertation, we assume that the wireless channels are quasi-static, meaning that any time variation is slow with respect to the size of blocks of data that are transmitted. The quasi-static assumption is mainly adopted because it simplifies analysis and the algorithms presented here rely heavily upon it, but for wireless LAN applications, the assumption is realistic. For cellular applications, the algorithms presented here may still be usable, but will need to be adapted somewhat. For example, the hybrid algorithm in Chapter 4 allows the inclusion of noise statistics in the channel estimate, which could be used to accomodate time variation. Chapter 5 includes a study of indoor channel measurement data to determine the validity of the quasi-static assumption.

2.1.2 Single-User Capacity

The capacity of a channel, or the theoretical bound on transmission rate under a constraint on transmitted power, is an important analysis tool for communication channels. While it represents only an upper bound, it is an important indicator of the potential of a particular channel. This section reviews the derivation of capacity for the single-user MIMO channel. In discussing capacity, it is important to make some distinctions. First, capacity

is an inherent property of a given channel. Some of the algorithms presented in subsequent chapters have a maximum achievable throughput that is close to, but not the same as the system capacity, and should not be confused with the actual capacity of the channel. Also, the capacity of the channel changes as a function of the information available to the transmitter. This does not mean that there is more than one capacity, but rather, the transmitter that has knowledge of the channel can be regarded from the information theoretic point of view as a different channel than the "uninformed" or "blind" transmitter [48–50].

We assume linear pre- and post-processing at the transmitter and receiver, modeled by the matrices $\mathbf{M}$ and $\mathbf{W}$, respectively. If the transmitter transmits m symbols simultaneously, we model this with a m dimensional data vector $\mathbf{d}$, implying that $\mathbf{M}$ is $n_T \times m$ and $\mathbf{W}$ is $n_R \times m$. The transmitted signal $\mathbf{s}$ is $\mathbf{Md}$, and the received signal $\mathbf{x}$ is

$$\mathbf{x} = \mathbf{HMd} + \mathbf{n}, \tag{2.5}$$

and the receiver's estimate of the transmitted signal $\hat{\mathbf{d}}$ is

$$\hat{\mathbf{d}} = \mathbf{W}^* \left(\mathbf{HMd} + \mathbf{n}\right). \tag{2.6}$$

The capacity of such a channel is [8]:

$$C = \mathcal{I}(\mathbf{x};\mathbf{s}) = \mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{s}) \tag{2.7}$$

$$= \log_2 \det[\pi e \mathbf{R}_x] - \log_2 \det[\pi e \mathbf{R}_n] \tag{2.8}$$

$$= \max_{\mathbf{R}_x} \log_2 \frac{|\mathbf{R}_x|}{|\mathbf{R}_n|}, \tag{2.9}$$

where $\mathcal{I}$ denotes mutual information, $\mathcal{H}$ denotes entropy, $\mathbf{R}_x$ is the covariance of the received signal $\mathbf{x}$, and $\mathbf{R}_n$ is the covariance of the noise term $\mathbf{n}$ in equation (2.5). If the limit on the total transmitted power is P, the capacity for a single-user channel takes the form:

$$C = \max_{\mathbf{M},\, \text{s.t. } \mathrm{tr}(\mathbf{MM}^*) \le P} \log_2 \frac{|\mathbf{R}_n + \mathbf{HMM}^*\mathbf{H}^*|}{|\mathbf{R}_n|} \tag{2.10}$$

$$= \max_{\mathbf{M},\, \text{s.t. } \mathrm{tr}(\mathbf{MM}^*) \le P} \log_2 |\mathbf{I} + \mathbf{R}_n^{-1}\mathbf{HMM}^*\mathbf{H}^*|. \tag{2.11}$$

The optimal $\mathbf{M}$ depends on what information about $\mathbf{H}$ is available.

Informed Transmitter

If $\mathbf{H}$ is known perfectly to the transmitter, then the capacity (called C_{IT} for "informed transmitter") is

$$C_{\text{IT}} = \max_{\mathbf{M}',\mathbf{\Lambda}_Z} \log_2 \left|\mathbf{I} + \mathbf{M}'^{*}\mathbf{H}^{*}\mathbf{R}_n^{-1}\mathbf{H}\mathbf{M}'\mathbf{\Lambda}_Z\right| \tag{2.12}$$

$$\leq \sum_{i=1}^{m} \log_2(1 + [\mathbf{W}]_{ii}\lambda_{Z,i}) \ , \tag{2.13}$$

where $\mathbf{W}$ is defined as $\mathbf{M}'^{*}\mathbf{H}^{*}\mathbf{R}_n^{-1}\mathbf{H}\mathbf{M}'$. In equation (2.12), $\mathbf{M}$ is factored into two components: a set of steering vectors each with unit gain and a diagonal power weighting matrix, so that $\mathbf{M} = \mathbf{M}'\mathbf{\Lambda}_Z^{1/2}$. The inequality in (2.13) is due to the fact that for any positive semidefinite matrix $\mathbf{A}$,

$$|\mathbf{A}| \leq \prod_i [\mathbf{A}]_{ii} \, . \tag{2.14}$$

Equality can be achieved in (2.14) when $\mathbf{A}$ is diagonal, so the $\mathbf{M}'$ which maximizes equation (2.12) independently of $\mathbf{\Lambda}_Z$ is the solution that makes $\mathbf{W}$ diagonal. $\mathbf{M}'$ is therefore derived from the eigenvalue decomposition of $\mathbf{H}^{*}\mathbf{R}_n^{-1}\mathbf{H}$:

$$\mathbf{H}^{*}\mathbf{R}_n^{-1}\mathbf{H} = \mathbf{M}'\mathbf{\Lambda}_{HR_n}^2\mathbf{M}'^{*} \, , \tag{2.15}$$

where $\mathbf{\Lambda}_{HR_n}^2$ is the diagonal matrix containing the eigenvalues. At this point, we have decomposed the channel into a set of orthogonal channels with gains of $\lambda_{HR_n,n}^2$, the diagonal elements of $\mathbf{\Lambda}_{HR_n}^2$. The noise power in each channel will be one. The maximization can now be solved by determining the optimal power distribution among the channels, which is contained in the diagonal elements of $\mathbf{\Lambda}_Z$. The solution to this problem is the well-known "water-filling" solution [8]:

$$\lambda_{Z,n} = \left(\mu + \frac{1}{\lambda_{HR_n,n}^2}\right)^{+} \, , \tag{2.16}$$

where $(z)^{+} \equiv \max\{0, z\}$, and μ is chosen so that the total transmitted power constraint is satisfied. The corresponding capacity of this channel then becomes

$$C_{\text{IT}} = \sum_{n=1}^{L} \log_2\left(1 + \lambda_{Z,n}\lambda_{HR_n,n}^2\right) \, . \tag{2.17}$$

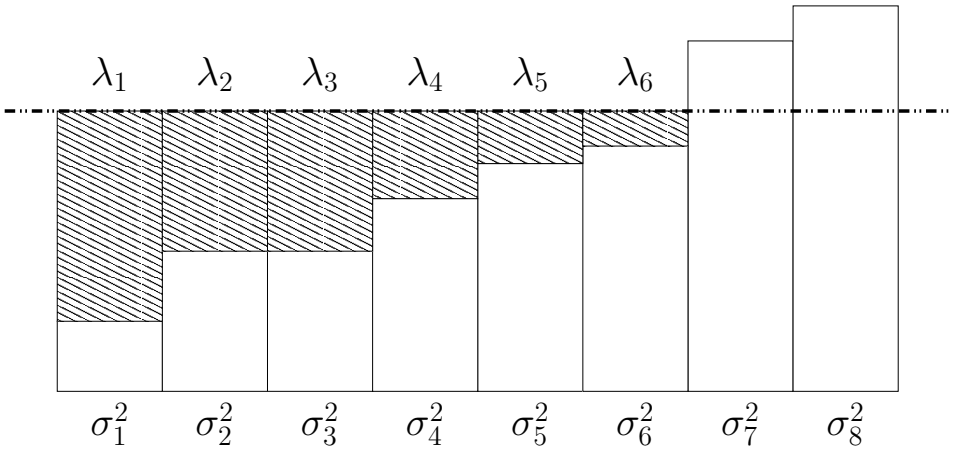

Figure 2.1: Illustration of water-filling to achieve capacity.

Figure 2.1 illustrates the water-filling problem for channels with equal gains and unequal noise powers, labeled in the diagram as $\sigma_1^2 \ldots \sigma_8^2$. This is equivalent to the problem considered here if the channel gains are normalized so that noise powers are equal. The term water-filling is used because graphically, the solution is equivalent to pouring a total amount of "water", or signal power, into an irregularly shaped container. Note that in the case pictured here, two of the channels are not used.

In addition to solving the capacity problem, the water-filling solution can also be used to solve a second optimization problem. If the group of channels are to be used to achieve a minimum total transmission rate while minimizing the total transmitted power, the same solution applies, except that μ in equation (2.16) is chosen to satisfy the rate requirement, rather than the power requirement. Using the water-pouring analogy, this corresponds to pouring an indefinite amount of water until the requirement is satisfied, as opposed to pouring a fixed amount of water, or transmitted power.

Thus far, we have considered the general case where the noise covariance $\mathbf{R}_n$ has any arbitrary structure. When the only source of noise at the receiver is thermal noise, then the noise will be spatially white, and $\mathbf{R}_n = \mathbf{I}$. In this case $\mathbf{M}'$ corresponds to the right singular values of $\mathbf{H}$. If there are other structured interference sources at the receiver, which are known to the transmitter, resulting in a non-diagonal $\mathbf{R}_n$, the decomposition is equivalent to a spatial noise-whitening filter.

We now briefly examine the asymptotic behavior of capacity at high SNR, repeating results originally found in [6]. We assume that the $\mathbf{H}$ matrix is normalized so that the noise power is 1, so the total transmit power P represents the SNR. If P is large, every one of the $\lambda_{Z,n}$ values will be nonzero, and the μ term will dominate the right side of (2.17), so that $\lambda_{Z,n} \to \mu = P/L$. So the capacity becomes

$$C \to \sum_{n=1}^{L} \log_2 \left(1 + \frac{P}{L}\lambda_{HR_n,n}^2\right) \tag{2.18}$$

$$\approx \sum_{n=1}^{L} \log_2(\lambda_{HR_n,n}^2) + L\log_2(P/L). \tag{2.19}$$

From this it is easy to see that at high SNR, capacity grows linearly with L, which is the rank of $\mathbf{H}$. This connection between rank and capacity, and the connection between rank and multipath evident from (2.3) illustrates an important problem. The early work on MIMO channels, most notably [4], demonstrated the potential of MIMO channels using simulations that assumed the best possible channel conditions, specifically that all members of $\mathbf{H}$ are Gaussian and independently, identically distributed (IID). While this may actually be the case in some channel conditions such as indoor channels which typically have significant multipath, many outdoor propagation environments tend to have fewer scatterers, so the multipath structure is dominated by a few paths, or even a single path, particularly where a line-of-sight component exists. These rank-deficient channels are often referred to as "keyhole" channels. A study of some propagation channels that have this effect is found in [51]. Channels with low rank or correlated fading characteristics have reduced capacity compared to uncorrelated full rank channels. Several recent works in the literature have studied capacity and optimal signaling schemes for channels with correlated fading [52–55]. As is the case with SISO channels, the fading properties of a MIMO channel will ultimately determine which transmission schemes make the most sense. Some of these transmission schemes are discussed later in this chapter.

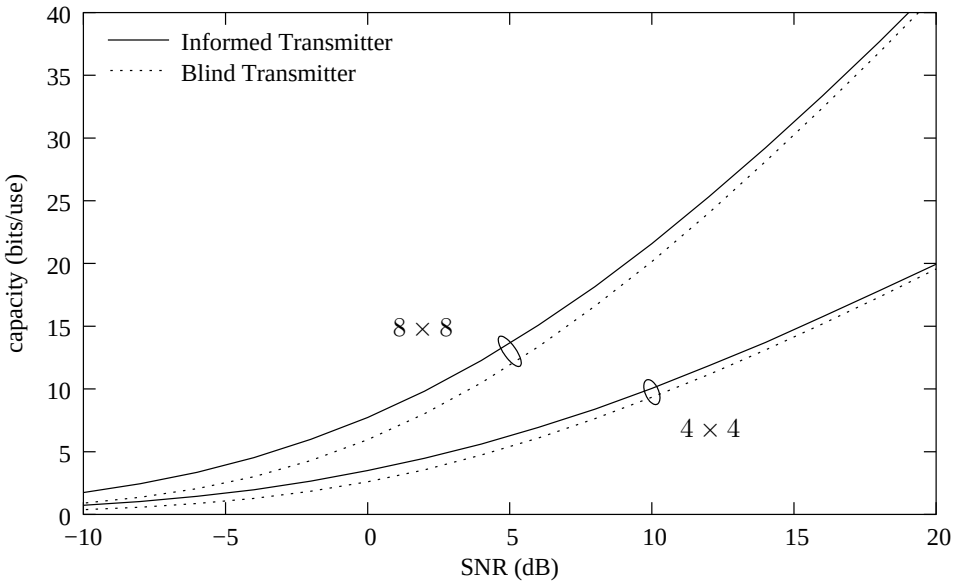

Figure 2.2: A comparison of capacity for blind and informed transmitters.

Blind Transmitter

We now consider the case where the transmitter does not know the channel. If the statistics of $\mathbf{H}$ are known, then it is possible to maximize the expected value of capacity, or ergodic capacity. If we assume that $\mathbf{H}$ has Gaussian IID elements, we want to maximize

$$E_{\mathbf{H}} \left[\log_2 \det \left(\mathbf{I}_{n_R} + \mathbf{H}\mathbf{R}_s\mathbf{H}^* \right) \right], \tag{2.20}$$

where $\mathbf{R}_s$ is the covariance of the transmitted signal. Let the SVD of $\mathbf{R}_s$ be $\mathbf{R}_s = \mathbf{U}\mathbf{D}\mathbf{U}^*$. Then the expectation is

$$E_{\mathbf{H}} \left[\log_2 \det \left(\mathbf{I}_{n_R} + (\mathbf{H}\mathbf{U})\mathbf{D}(\mathbf{H}\mathbf{U})^* \right) \right]. \tag{2.21}$$

Since $\mathbf{H}\mathbf{U}$ is distributed identically to $\mathbf{H}$, then $\mathbf{R}_s$ that maximizes the expectation will be diagonal. The optimal $\mathbf{R}_s$ that satisfies the power constraint is $\mathbf{R}_s = P/n_T\mathbf{I}$. So the capacity of a channel without CSI at the transmitter is:

$$C = \log_2 \det \left(\mathbf{I}_{n_R} + P/n_T\mathbf{H}\mathbf{H}^* \right). \tag{2.22}$$

Figure 2.2 illustrates the capacity of blind and informed transmitters as a function of SNR. The two sets of curves represent 4×4 and 8×8 channel matrices. The capacity

values represent a 10% outage probability of a set of randomly generated channels drawn from an IID Gaussian distribution. One clear trend is that as SNR increases, the difference between the blind and informed transmitters gets asymptotically smaller. This can be attributed to the fact that as the available power increases, all sub-channels in the water-filling solution are used, and the power distribution becomes closer to a uniform power distribution, which is equivalent to the power distribution in the blind transmitter case. A study of how SNR and other factors influence channel capacity is found in [9].

The results in Figure 2.2 call into question the value of obtaining channel information at the transmitter. In two-way communication systems, there are some ways to do this, such as using channel feedback in the reverse channel, or using estimates of the reverse channel to estimate the forward channel [36, 56, 57]. However, each of these comes at a cost, whether in computation or bandwidth, and poses different problems in estimation accuracy. At higher SNRs, the relatively small gap between channels with and without channel information at the transmitter is sufficiently small that it may be difficult to justify these costs. However, this assumes that the channel is full rank. When the channel is rank deficient, the gaps are larger, and having complete, or even only partial channel information available can be advantageous. To illustrate this, we include the derivation of capacity with partial channel information.

Capacity with Partial Channel Information

The problem of estimating a channel from measurements on the reverse channel is particularly challenging when Frequency Division Duplexing (FDD) is used. Direct estimates of the reverse channel do not correspond to the forward channel in this case. It is possible to estimate angle of arrival of the multipath components, provided the number of antennas is larger than the number of paths. However, as is seen in equation (2.3), synthesizing $\mathbf{H}$ from angles of the paths requires both angle of arrival and angle of departure, so angle of arrival information provides only partial information about the channel. An additional situation where only partial channel information may be available to the transmitter is illustrated in [57], where averaging of subspaces is used in fast time-varying channels, and the weighting of the subspace is considered to be less stable than the subspaces

themselves. This section presents a derivation of capacity for partial channel information originally found in [36].

Both of the partial information scenarios result in a situation where $\mathbf{H}$ can be factored into

$$\mathbf{H} = \mathbf{AB}\,, \tag{2.23}$$

where $\mathbf{B}$ is known to the transmitter, but $\mathbf{A}$ is not. This is particularly relevant in low-rank channels, where the rank $L \leq n_R$. We assume that $\mathbf{B}$ is $L \times n_T$. So the maximum dimension of the data vector $\mathbf{d}$ is L. The capacity of the channel (C_{PI}) is

$$C_{\text{PI}} = \max_{\mathbf{M},\text{s.t. tr}(\mathbf{MM}^*)\leq P} \log_2 \det\left(\mathbf{I} + \mathbf{ABMM}^*\mathbf{B}^*\mathbf{A}^*\right)\,, \tag{2.24}$$

At high signal to noise ratios, the second term in the determinant dominates, so we will simplify and reduce the matrix to only one term, similar to the high SNR approximation used in [58]. This cannot be done directly, since the second term is not full rank, so dropping the identity term results in a determinant equal to zero. However, it can be shown using the singular value decomposition of $\mathbf{A}$ that:

$$\begin{aligned} |\mathbf{I}_{n_R} + \mathbf{ABMM}^*\mathbf{B}^*\mathbf{A}^*| &= |\mathbf{I}_L + \mathbf{BMM}^*\mathbf{B}^*\mathbf{A}^*\mathbf{A}| && (2.25)\\ &= |\mathbf{A}^*\mathbf{A}||(\mathbf{A}^*\mathbf{A})^{-1} + \mathbf{BMM}^*\mathbf{B}^*| && (2.26)\\ &\approx |\mathbf{A}^*\mathbf{A}||\mathbf{BMM}^*\mathbf{B}^*|\,. && (2.27) \end{aligned}$$

The approximation in (2.27) is also a high SNR approximation, where the $(\mathbf{A}^*\mathbf{A})^{-1}$ term is dropped because the second term in (2.26) still dominates. So, we can say that

$$C_{\text{PI}} \approx \log_2 |\mathbf{A}^*\mathbf{A}| + \max_{\mathbf{M}} \log_2 |\mathbf{BMM}^*\mathbf{B}^*|\,. \tag{2.28}$$

Now the capacity has been split into two terms, the first of which is independent of our knowledge of the channel, and the second of which must be optimized. A similar transmit vector optimization problem is addressed in [59, 60] in the context of space-time coding, where an optimal code must be selected at the transmitter to maximize the diversity gain. The problem in this case is to maximize the second term in (2.28) subject to the power

constraint $\mathrm{tr}(\mathbf{M}^*\mathbf{M}) \leq P$. Using a Lagrange multiplier λ, we take the derivative with respect to the matrix $\mathbf{M}$ and set it to zero:

$$\frac{\partial}{\partial \mathbf{M}} \log |\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*| + \lambda \left[P - \mathrm{tr}(\mathbf{M}^*\mathbf{M})\right] = \tag{2.29}$$

$$2\mathbf{B}^*(\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*)^{-1}\mathbf{B}\mathbf{M} - 2\lambda\mathbf{M} = \mathbf{0}\,. \tag{2.30}$$

This results in:

$$\lambda\mathbf{M} = \mathbf{B}^*(\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*)^{-1}\mathbf{B}\mathbf{M}. \tag{2.31}$$

Pre-multiplying both sides by $\mathbf{M}^*$, taking a trace, and using the commutativity of matrices under the trace gives:

$$\lambda\,\mathrm{tr}(\mathbf{M}^*\mathbf{M}) = \mathrm{tr}\left(\mathbf{M}^*\mathbf{B}^*(\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*)^{-1}\mathbf{B}\mathbf{M}\right) \tag{2.32}$$

$$= \mathrm{tr}\,\mathbf{I}_L = L\,. \tag{2.33}$$

Solving for λ gives $\lambda = L/P$. Substituting this into the original expression gives an expression for $\mathbf{M}$, which can be simplified one step further:

$$\mathbf{M} = \frac{P}{L}\mathbf{B}^*(\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*)^{-1}\mathbf{B}\mathbf{M} \tag{2.34}$$

$$= \frac{P}{L}\mathbf{B}^*(\mathbf{M}^*\mathbf{B}^*)^{-1}. \tag{2.35}$$

Pre-multiplying $\mathbf{M}$ by $\mathbf{M}^*$ yields

$$\mathbf{M}^*\mathbf{M} = \frac{P}{L}\mathbf{I}_L\,, \tag{2.36}$$

which shows that in order to maximize the determinant and achieve the capacity, the columns of $\mathbf{M}$ must be orthogonal.

Using the expression for $\mathbf{M}$ and substituting it into $\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*$ gives the second term of the capacity in (2.28):

$$\mathbf{M}\mathbf{M}^* = \frac{P}{L}\mathbf{B}^*(\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*)^{-1}\mathbf{B}\mathbf{M}\mathbf{M}^* \tag{2.37}$$

$$\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^* = \frac{P}{L}\mathbf{B}\mathbf{B}^*(\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*)^{-1}\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^* \tag{2.38}$$

$$= \frac{P}{L}\mathbf{B}\mathbf{B}^* \tag{2.39}$$

$$\log_2 |\mathbf{B}\mathbf{M}\mathbf{M}^*\mathbf{B}^*| = \log_2 |\mathbf{B}\mathbf{B}^*| + L\log_2 \frac{P}{L}\,. \tag{2.40}$$

Substituting this back into the expression for capacity, it now simplifies as follows:

$$C_{\text{PI}} \approx \log_2 |\mathbf{A}\mathbf{A}^*| + \log_2 |\mathbf{B}\mathbf{B}^*| + L \log_2 \frac{P}{L} \tag{2.41}$$

$$= \log_2 |\mathbf{A}\mathbf{B}\mathbf{B}^*\mathbf{A}^*| + L \log_2 \frac{P}{L} \tag{2.42}$$

$$= \log_2 \left| \frac{P}{L} \mathbf{H}\mathbf{H}^* \right| . \tag{2.43}$$

This is a function of the total transmitted power P. The mean power per transmit antenna will be P/n_T. The noise is assumed to have unit variance, so the SNR per transmit antenna is also P/n_T. From equations (2.36) and (2.40) it can be inferred that the optimal transmit vectors $\mathbf{M}$ are a set of orthogonal vectors which span the column space of $\mathbf{B}^*$. The achievable throughput of this transmission scheme is actually equation (2.43) with the identity term that was previously removed added back in. So the capacity approximation is now:

$$C_{\text{PI}} \approx \log_2 \left| \mathbf{I}_{n_R} + \frac{P}{L} \mathbf{H}\mathbf{H}^* \right| . \tag{2.44}$$

Note that the difference between this expression and the capacity for a blind transmitter is the P/L term replaces the P/n_T term. So, if $L = n_T$, partial channel information does not provide any advantage. For low rank channels where $L < n_T$, there is an advantage in having partial information, because using the directivity of the transmitter to send information only into the space spanned by $\mathbf{B}$ offers better performance than no directivity.

Figure 2.3 compares the capacity of low-rank channels with complete, partial, and no channel information. The two sets of curves again represent 4×4 and 8×8 channels, but all of the channels in this case are rank 2. In this case, channels with partial information approach the performance of channels with complete information at high SNR, while channels with no information have a substantial performance loss.

2.1.3 MIMO Transmission Strategies

It is shown in [6] that the theoretical capacity of an $n_R \times n_T$ channel of rank L is limited by $\min\{n_R, n_T, L\}$. Except in cases where the number of multipath components is greater than the dimensions of $\mathbf{H}$, the rank will correspond to L. Rich multipath channels such as urban or indoor environments will clearly benefit from the fact that they

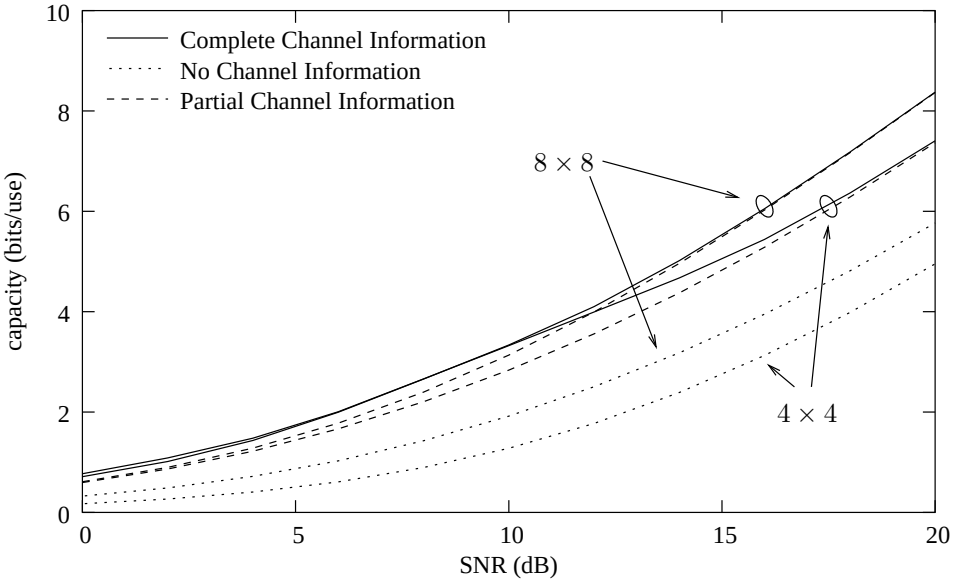

Figure 2.3: A comparison of capacity for complete, partial, and no channel information at the transmitter.

will have nearly full-rank $\mathbf{H}$ (depending on how big the arrays are), but there are many channels where L is 1, or close to 1. In this case, parallel data transmission is not possible. As a result, two different general approaches to MIMO transmission have evolved, one emphasizing parallel data transmission and the other emphasizing diversity. A comparison of these two approaches can be found in [10].

Spatial Multiplexing

The first proposed transmission method that used the parallel data transmission approach was the BLAST algorithm in [5], which has subsequently been elaborated upon [61]. The BLAST algorithm essentially consists of transmitting a different data symbol out of each transmit antenna. Mathematically this is equivalent to $\mathbf{M} = \mathbf{I}_{n_T}$. Each block of symbols contains training data which is used at the receiver to estimate the channel. If simple linear combining is used at the receiver, a good solution is using the training data to compute an MMSE estimate of the channel and invert it to decode the remaining transmitted data. However, it has been shown in [5] that some non-linear solutions to this problem can have better performance than the optimum linear solutions. Specifically, the

authors propose a scheme in which the strongest of the signals is estimated first and subtracted from the signal, and then repeating this process until all signals have been decoded. As will be discussed later, this type of ordered processing has proven useful in multi-user environments as well.

Subsequent improvements on this scheme have included improved decoding techniques at the receiver [62] and smarter transmission methods that compensate for time variations in the channel. This class of algorithm is capable of achieving very high data rates, but is limited to situations where channel conditions are favorable. A good review of BLAST and related transmission methods is [61].

In a spatial multiplexing model of MIMO transmission, the channel modulation and demodulation matrices can be viewed as attempting to diagonalize the product $\mathbf{W}^*\mathbf{H}\mathbf{M}$. The BLAST approach [5], which does not use any channel pre-coding, essentially leaves the task of the diagonalization to the receiver. On the other hand, the water-filling solution of [6] breaks the channel down into its dominant subspaces, so that optimal power loading into the sub-channels can be performed. In this case, the diagonalization is accomplished by a combination of both $\mathbf{M}$ and $\mathbf{W}$. This observation proves to be very important in the multi-user case.

Transmit Diversity

An alternative approach to the MIMO channel is to use the additional antennas to improve diversity rather than focusing on parallel data transmission. It has been shown that a $n_T \times n_R$ channel has diversity order $n_T n_R$. This approach is commonly referred to as "space-time" coding because it focuses on designing codes that improve reliability by spreading the data across both space and time. An early example of this is the trellis-based code proposed in [63]. Another approach is the codes based on "orthogonal designs" [11,12], which are able to achieve near optimal diversity at both the transmitter and receiver using a very simple transmit diversity scheme.

The space-time codes and spatial multiplexing schemes represent two fundamentally different approaches to the MIMO channel, one which emphasizes high data rates and the other which emphasizes high reliability. The choice of which of these should be

used depends on the particular channel characteristics and design requirements. However, it is not completely necessary to choose only one or the other. Some recent work has proposed some schemes which allow a combination of the two approaches to achieve a compromise between the benefits of diversity and multiplexing [64, 65].

2.2 Multi-User Wireless Channels

Even for SISO channels, a multi-user environment can be taken to mean a variety of things. Perhaps the most common is the cellular model of a base communicating with multiple users. This same model can also apply to a wireless LAN environment. On the other hand, a wireless LAN could also use a peer-to-peer networking concept, where a group of users are part of a network and each user may want to use the channel to communicate with one of the other users. The cellular model is important because it is already in wide use, and compared to other models, presents a relatively simple starting point for analysis of new transmission algorithms. The peer-to-peer model is more challenging, but is becoming an important research area because of applications such as ad-hoc networking. The cellular model is assumed in this dissertation, and extending the problem to the peer-to-peer model would be a logical next step in furthering the research presented here.

One important aspect of modern communication networks to consider is the idea of Quality of Service (QoS). First and second generation cellular systems were designed specifically with voice services in mind, while third generation sytems that are currently being implemented are designed with the intention of carrying a wide variety of voice and data services. Each type of service has specific requirements for transmission rate, error probability and latency, which can all be regarded as QoS. For example, voice communications require limited bandwidth, but very low latency, and a reasonably low bit error rate. On the other hand, data services will require higher rates, but have more lenient requirements on latency, and perhaps on error rates as well, since retransmission is more acceptable than in a voice channel. It is also possible for a wide variety of data rates to be made available to consumers, depending on what they are willing to pay for. The same is true to some extent for wireless LANs, except that they are less likely to be carrying voice

traffic. We assume here that questions of latency are the concern of higher layers in the networking model and that the primary concerns in the physical layer of a wireless network, which is the focus of this dissertation, are achieving the required rate and error probability. So, QoS will be taken to mean a requirement for a specific rate and bit error rate, with rate being the primary focus here.

Traditionally, the multi-user problem has been handled using three different multiplexing methods: time, frequency, and code division multiple access (TDMA, FDMA, and CDMA). TDMA means that the full frequency bandwidth of the channel is divided up into time slots and the channel is shared by allocating those time slots to the users that wish to share the channel. FDMA is similar, but the bandwidth is subdivided into frequency slots, which are allocated to the users. CDMA uses direct sequence spread spectrum, and allocates a different code sequence to each user sharing the channel. Each of these have different advantages which make them suited to particular types of channels. Some modern applications combine two or more of these.

A more recent type of multiplexing that has become prevalent with the use of arrays of antennas at the base stations of cellular systems has been dubbed space division multiple access (SDMA). This means that array processing methods are used at the base to allow two or more users to share the same channel. As will be shown later, the maximum number of users that can share the same channel in this way is a function of array size. Since in most cases it isn't realistic to build a big enough array to accommodate all users, SDMA also is generally used in conjunction with other multiplexing methods. Chapter 6 deals with the problem of deciding which users should share channels in a multi-user MIMO environment.

2.3 The Multi-User MIMO Channel

In the literature multi-user MIMO has been taken to refer both to the case where the users have only one antenna and when they have multiple antennas. Both of these channels can reasonably be considered MIMO in the sense that there are multiple inputs and multiple outputs regardless of the particular configuration. However, the primary focus of this dissertation is the multi-user channel where both the base and users all have antenna

arrays, and unless noted otherwise, this is what multi-user MIMO will be taken to mean. This section discusses how the single-user MIMO model and capacity can be extended to the multi-user MIMO case.

2.3.1 Channel Model

In a cellular channel, there are two communication problems to consider. The first is a group of users all transmitting to the same base station, commonly referred to as the uplink. The reverse, where the base attempts to transmit signals to multiple users, is commonly referred to as the downlink.

Like the single-user MIMO channel, the multi-user MIMO channel can also be represented using matrix multiplication. In single-user MIMO channels, the benefit of MIMO processing is gained from the coordination of the processing among all of the transmitters or receivers. We assume that in the multi-user channel, there is no coordination possible among the users. Such coordination would be very difficult to achieve, and remains largely an open problem, but has been recently considered [66]. A result of the lack of coordination between the users is that the problem differs somewhat between the uplink and downlink channels. We now consider each of these separately.

The Uplink

Assuming a narrowband propagation environment, let $\mathbf{H}_j$ represent the channel transfer matrix from the j^{th} user to the base station in the uplink, and let $\mathbf{M}_j$ and $\mathbf{d}_j$ represent the modulation matrix and data vectors transmitted by user j. Let K represent the number of users. Then the received signal at the base is

$$\mathbf{x} = \sum_{j=1}^{K} \mathbf{H}_j \mathbf{M}_j \mathbf{d}_j + \mathbf{n}\,. \tag{2.45}$$

This is also referred to in the information theory literature as the "vector multiple access channel". In this case the challenge is for the receiver to separate all of the simultaneously received signals, using multi-user detection, or other methods. Since the users are not able to coordinate with each other, there is little that can be done to optimize the transmitted signals. If some channel feedback is allowed from the transmitter back to the users, some

coordination may be possible, but it may require that each user know all users' channels, rather than its own. Otherwise the challenge in the uplink is mainly in the processing done by the base to separate the users.

The Downlink

The downlink channel, where the base is simultaneously transmitting to a group of users, is the primary focus of this dissertation. In the information theory literature, the case where a base transmits signals to multiple users, is called the "broadcast channel", and the MIMO version of it discussed here is referred to the "vector broadcast channel". We consider a multi-user downlink channel with K users and a single base station. The base has n_T antennas, and the j^{th} 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 4-antenna base and two 2-antenna users. Similar to the uplink, we define the channel matrix from the base to the j^{th} user as $\mathbf{H}_j$, and the associated modulation matrix as $\mathbf{M}_j$. The signal at the j^{th} receiver is thus:

$$\mathbf{x}_j = \sum_{i=1}^{K} \mathbf{H}_j \mathbf{M}_i \mathbf{d}_i + \mathbf{n}_j = \mathbf{H}_j \mathbf{M}_S \mathbf{d}_S + \mathbf{n}_j \tag{2.46}$$

$$= \underbrace{\mathbf{H}_j \mathbf{M}_j \mathbf{d}_j}_{\text{signal}} + \underbrace{\mathbf{H}_j \tilde{\mathbf{M}}_j \tilde{\mathbf{d}}_j}_{\text{interference}} + \underbrace{\mathbf{n}_j}_{\text{noise}}, \tag{2.47}$$

where the following definitions apply:

$$\mathbf{M}_S = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 & \ldots & \mathbf{M}_K \end{bmatrix} \tag{2.48}$$

$$\mathbf{d}_S = \begin{bmatrix} \mathbf{d}_1^T & \mathbf{d}_2^T & \ldots & \mathbf{d}_K^T \end{bmatrix}^T \tag{2.49}$$

$$\tilde{\mathbf{M}}_j = \begin{bmatrix} \mathbf{M}_1 & \ldots & \mathbf{M}_{j-1} & \mathbf{M}_{j+1} & \ldots & \mathbf{M}_K \end{bmatrix} \tag{2.50}$$

$$\tilde{\mathbf{d}}_j = \begin{bmatrix} \mathbf{d}_1^T & \ldots & \mathbf{d}_{j-1}^T & \mathbf{d}_{j+1}^T & \ldots & \mathbf{d}_K^T \end{bmatrix}^T. \tag{2.51}$$

The challenge in the downlink problem is to minimize the inter-user interference created by the transmitter. If the transmitter has knowledge of the users' channels, it is then aware of what interference is being created for user 2 by the signal it is transmitting to user

1, for example. So, it is possible for the transmitter to avoid interference by carefully designing the $\mathbf{M}_j$ matrices. Since the receivers have array processing capabilities, it is also possible to mitigate this problem on the receiving end, using what is known as multi-user detection (MUD). In this approach, the receiver is designed to detect and separate multiple signals, even though only one of them is of interest to the user. Obviously, this considerably increases the computational cost, which is often not acceptable since the users in cellular environments are usually small mobile devices. If the transmitter does not know the channel, and significant inter-user interference is not acceptable, the only available option is to resort to other traditional multiple access methods.

For future use, we also introduce the following matrices and notation related to transmission in the downlink channel. The system channel matrix for the downlink channel is:

$$\mathbf{H}_S = \begin{bmatrix} \mathbf{H}_1^T & \mathbf{H}_2^T & \dots & \mathbf{H}_K^T \end{bmatrix}^T. \tag{2.52}$$

We also define the following matrix which represents the channel matrix for all users other than user j:

$$\tilde{\mathbf{H}}_j = \begin{bmatrix} \mathbf{H}_1^T & \dots & \mathbf{H}_{j-1}^T & \mathbf{H}_{j+1}^T & \dots & \mathbf{H}_K^T \end{bmatrix}^T. \tag{2.53}$$

2.3.2 Capacity of the Multi-User MIMO Downlink

The capacity of multi-user MIMO downlinks is intimately connected with a result by Costa called "writing on dirty paper" [28], which is briefly summarized here. Suppose X represents a transmitted signal, and W and Z are additive noise terms, so that the received signal Y is

$$Y = X + W + Z. \tag{2.54}$$

Costa showed that if W is known deterministically to the transmitter, then the capacity of the communication channel is the same as a channel with only the second interference term:

$$Y = X + Z, \tag{2.55}$$

regardless of whether or not the receiver knows W and independent of the statistics of W. The concept of writing on dirty paper implies designing a code that avoids the known interference W. A tutorial example of this can be found in [29]. The dirty paper concept up to this point has had some application in problems such as data hiding, but has recently become useful in the downlink of multi-user MIMO channels, and has been shown to achieve the outer bound of the rate region of such channels.

Multi-user capacity can be taken to have different meanings. It is possible to consider the capacity of one particular user in the context of a system, or to consider the sum capacity of all users in the system. Under a single power constraint, it is possible to achieve a variety of different combinations of rates for different users by allocating resources differently to different users. The convex set of all achievable sets of rates for a multi-user channel is commonly referred to as a "rate region". Several researchers have recently attempted to characterize the rate region for the multi-user MIMO channel [25, 27, 67]. It has been shown that the maximum sum capacity, as well as other points on the boundary of the region, are achievable by using the dirty-paper coding approach. This involves ordering the users in some arbitrary order $1 \ldots K$, and then for each user j using dirty-paper codes to avoid the interference caused by transmitting to users $1 \ldots j-1$, since those signals are known completely. Actually designing this type of code is a very complex problem. One proposed approach can be found in [68].

This dissertation focuses on transmission methods that do not require new coding schemes, and thus the interference transmitted to other users than the intended user is treated as noise. In the information theory literature, a multi-user channel like this is referred to as a "degraded broadcast channel". The advantage of this approach is that it allows simpler structures at both the transmitters and receivers, and can be applied to existing communication systems and protocols because it does not require new types of codes. We now examine the capacity of the degraded broadcast channel. It should be noted that the following capacity results which treat interference as noise assume that the interference has a Gaussian distribution, which is the distribution which maximizes entropy and thus capacity [8]. This assumption may not always be realistic in practice, but in matrix channels where the final noise term at the output is the result of an additional matrix product,

the interference will be asymptotically Gaussian as the size of the receiver's array increases due to the law of large numbers.

Informed Transmitter, Single User Perspective

When the users' channels are known at the transmitter, SDMA can be employed to increase capacity. In particular, the capacity of the channel for user j is

$$C_j = \max_{\mathbf{M}_j} \log_2 \left| \mathbf{I} + (\sigma_n^2 \mathbf{I} + \mathbf{H}_j \tilde{\mathbf{M}}_j \tilde{\mathbf{M}}_j^* \mathbf{H}_j^*)^{-1} \mathbf{H}_j \mathbf{M}_j \mathbf{M}_j^* \mathbf{H}_j^* \right| , \tag{2.56}$$

where $\mathbf{R}_{n_j}$ is the covariance of the noise vector. The capacity is thus a function of not only what modulation matrix is chosen for the particular user of interest, but also those chosen for all other co-channel users as well. Viewing the problem entirely from the perspective of receiver j, capacity is maximized when

$$\mathbf{H}_j \tilde{\mathbf{M}}_j = \mathbf{0} , \tag{2.57}$$

or, in other words, when the transmit matrix $\tilde{\mathbf{M}}_j$ for all users other than j lies in the null space of $\mathbf{H}_j$. If this is done, then the capacity of user j is equal to the water-filling capacity of the channel matrix $\mathbf{H}_j$. Note that $n_T \geq n_R$ is a necessary condition for achieving (2.57), a requirement not imposed in the blind transmitter case.

Informed Transmitter, System Perspective

In general, the sum capacity of the channel from a system perspective (C_S) would be the sum over j of all terms like (2.56), maximized over all $\mathbf{M}_j$:

$$C_S = \max_{\mathbf{M}_j, j=1\ldots K} \sum_{j=1}^{K} \log_2 \left| \mathbf{I} + (\sigma_n^2 \mathbf{I} + \mathbf{H}_j \tilde{\mathbf{M}}_j \tilde{\mathbf{M}}_j^* \mathbf{H}_j^*)^{-1} \mathbf{H}_j \mathbf{M}_j \mathbf{M}_j^* \mathbf{H}_j^* \right| . \tag{2.58}$$

The set of $\mathbf{M}_j$ matrices that achieves the sum capacity of the system is the one that balances optimal power transmission to each user against the resulting inter-user interference. Using only beamforming (no dirty-paper coding), there is no simple way to find the optimal set of $\mathbf{M}_j$ matrices that achieves this.

A suboptimal approach would be to force all of the multi-user interference to zero: $\mathbf{H}_i\mathbf{M}_j = \mathbf{0}$ for $i \neq j$. The maximum information rate (C_{BD}) of such a scheme is:

$$C_{\text{BD}} = \max_{\mathbf{M}_S,\, \mathbf{H}_i\mathbf{M}_j=\mathbf{0},\, i\neq j} \log_2 |\mathbf{I} + 1/\sigma_n^2 \mathbf{H}_S\mathbf{M}_S\mathbf{M}_S^*\mathbf{H}_S^*| \tag{2.59}$$

$$= \max_{\mathbf{H}_i\mathbf{M}_j=\mathbf{0},\, i\neq j} \sum_{j=1}^{K} \log_2 |\mathbf{I} + 1/\sigma_n^2 \mathbf{H}_j\mathbf{M}_j\mathbf{M}_j^*\mathbf{H}_j^*| \leq C_S\,. \tag{2.60}$$

Equation (2.60) results because the product $\mathbf{H}_S\mathbf{M}_S$ takes on a block-diagonal structure, hence the subscript BD in equation (2.60). The throughput of the constrained solution C_{BD} is a lower bound for the channel capacity because the set of possible solutions for $\mathbf{M}_j$ has been reduced in size. However, at high SNR, the interference terms in equation (2.58) will be large relative to the noise, and the maximizing solution will force those terms toward zero. In the next chapter, we describe a block-diagonalization algorithm based on this idea.

Blind Transmitter

When the channel is unknown at the transmitter, it is not possible to communicate with multiple users simultaneously by means of SDMA without coordination between the user. Therefore, traditional multiplexing methods such as CDMA, TDMA, and FDMA are required. Because of bandwidth spreading, the capacity of a CDMA system requires additional considerations, and is not addressed here. In a TDMA or FDMA system, the capacity can be thought of as the average of the capacities of the $\mathbf{H}$ matrices for each user. Since the channel matrix is not known *a priori*, the best that can be done is to assume $\mathbf{M}_j\mathbf{M}_j^* = P/n_T\mathbf{I}$ for each user. Assuming equal time sharing for all users, the system capacity for the uninformed transmitter (C_{UT}) will then be

$$C_{\text{UT}} = \frac{1}{K}\sum_{j=1}^{K} \log_2\big|\mathbf{I} + \rho/n_T\mathbf{H}_j\mathbf{H}_j^*\big|. \tag{2.61}$$

where the SNR ρ is defined as the total available system power P divided by the mean noise power σ_n^2 at the receivers (we assume throughout the paper that the additive noise at each receiver has covariance $\sigma_n^2\mathbf{I}$).

2.3.3 Multi-User Transmission Algorithms

The problem of transmitting data over multi-user MIMO downlinks is relatively new, and is an active area of research. In spite of all of the recent work on discovering the rate region for the uplink and downlink channels, few approaches to achieving the bound have actually been proposed. Of the work has been done so far, solutions to the problem can be classified in two general categories: those that use coding methods to avoid inter-user interference, and those that use a linear processing approach.

The Coding Approach

The coding approach to multi-user multiplexing is to use the Costa dirty-paper concept to avoid inter-user interference. The dirty-paper approach requires exact knowledge of the interfering signal, but if one assumes that the channel is known, the interference due to signals intended for other users can be known exactly, and dirty-paper methods can be used to avoid the interference. This approach has been shown to achieve sum-capacity, and the other boundary points of the capacity region [24–27, 67]. However, these results, much like Shannon's original work, have really only established the existence of codes that achieve this rate, rather than propose any codes that might achieve them. Some more recent work such as [68, 69] has begun to address this issue, although with the primary focus on channels where the users have only one antenna.

The Linear Processing Approach

The signal processing approach to multi-user multiplexing assumes, as has been done extensively in this chapter, that each transmitter sends a linear combination of the symbols to be transmitted, and the receiver estimates the symbols also using linear processing. Thus rate and SINR constraints must be met by appropriately choosing the transmitter weights $\mathbf{M}_S$ and the receiver weights $\mathbf{W}_j$, while any channel coding in the symbols is assumed to be addressed separately. The advantage of this approach is that separating the spatial processing from all other parts of signal design (modulation, coding, etc.) makes it

possible for many of the ideas presented here to be used in conjunction with existing communication protocols, while achieving most of the available performance gain. The dirty-paper approach to multi-user transmission, while capable of higher capacity, combines all elements of signal design into one step and thus requires completely new commuication protocols. Dirty-paper concepts still require a significant amount of research before they are ready for widespread implementation. While they may prove to be the superior solution in the long term, the signal-processing approach is likely to be more readily usable for real-world implementation in the coming years. The linear processing approach has already been studied extensively for the case where the users each have only one antenna. In this section, we review some previous results.

The simplest approach to this is channel inversion, which was originally proposed in [70], and later mentioned in [25, 71]. If the transmitter knows $\mathbf{H}_S$, the channel can be diagonalized completely, eliminating all inter-user interference, by using the pseudo-inverse:

$$\mathbf{M}_S = \mathbf{H}_S^{\dagger} = \mathbf{H}_S^* \left(\mathbf{H}_S \mathbf{H}_S^*\right)^{-1} . \tag{2.62}$$

Compared to other alternatives, this is perhaps the most computationally inexpensive. It is proposed in [72] that this approach be extended to the multi-antenna receiver case. This provides a feasible but sub-optimal solution. Chapter 3 of this thesis provides a optimal solution to the multi-antenna case under the constraint that all inter-user interference be forced to zero.

Forcing all interference to zero is useful mainly because of its low computational cost. Its main drawback is susceptibility to noise in the channel estimate. If the channel information available to the transmitter is in fact perfect, the "zero-forcing" approach will work well. However, since most practical systems will be subject to noisy channel estimates and time variation in the channel, any channel information available to the transmitter will be to some degree corrupted.

A more robust approach to multiplexing users in the downlink would be to allow some inter-user interference, but at controlled levels that are close to the noise levels. This

approach, referred to here as "interference balancing", can potentially allow better solutions because it allows us to consider a larger set of possible solutions.

For the special case of one antenna per user, the simplest approach to this is a closed-form expression that is a generalization of channel inversion. Adding an additional term into the matrix inverse of equation (2.62), we obtain a structure that resembles an MMSE detector:

$$\mathbf{M}_S = \mathbf{H}_S^*(\mathbf{H}_S\mathbf{H}_S^* + \alpha\mathbf{I})^{-1}\,. \tag{2.63}$$

In [68], it is shown that the SINR for each user is optimized when α is chosen to be K/ρ, where ρ represents SNR, and is defined as total transmitted power over average noise power. This solution also has minimal computational cost, and has been shown in simulations to be an improvement on channel inversion. Disadvantages of this solution are the restriction to one antenna per user, and the assumption all users have the same SINR requirements.

For the case when users have differing SINR requirements, several iterative solutions to the interference balancing problem exist. An excellent summary of these can be found in [32]. Because it will be used later in this dissertation, one of these algorithms is detailed here. Specifically, we are interested in solving the optimization problem of minimizing the total transmitted power while satisfying the SINR requirement for each user. The solution to this problem can be found in [32], and was originally derived in [73, 74]. To express the problem in mathematical terms, let γ_j represent the SINR requirement for user j. Since we are dealing with the special case where $n_{Rj} = 1$ for all users, we modify the notation slightly and call the channel matrix $\mathbf{h}_j^*$, the transmit vector $\mathbf{m}_j$ and the transmitted data d_j. The received signal for user j is

$$x_j = \underbrace{\mathbf{h}_j^*\mathbf{m}_j d_j}_{\text{signal}} + \underbrace{\sum_{i\neq j}\mathbf{h}_j^*\mathbf{m}_i d_i}_{\text{interference}} + \underbrace{n_j}_{\text{noise}}\,. \tag{2.64}$$

Let $\mathbf{R}_j$ represent the covariance $E[\mathbf{h}_j^*\mathbf{h}_j]$. Then the SINR can be expressed as

$$\text{SINR}_j = \frac{\mathbf{m}_j^*\mathbf{R}_j\mathbf{m}_j}{\sum_{i\neq j}\mathbf{m}_i^*\mathbf{R}_j\mathbf{m}_i + \sigma_j^2}\,. \tag{2.65}$$

If we assume that $E[s_j s_j^*] = 1$, the optimization problem can be stated as

$$\begin{aligned} &\min \sum_{l=1}^{K} \mathbf{m}_l^* \mathbf{m}_l \\ &\text{such that} \quad \mathbf{m}_j^* \mathbf{R}_j \mathbf{m}_j - \gamma_j \sum_{i \neq j} \mathbf{m}_i^* \mathbf{R}_j \mathbf{m}_i \geq \gamma_j \sigma_j^2, \quad j = 1 \ldots K \,. \end{aligned} \tag{2.66}$$

Since the $\mathbf{m}_j$ vectors contain power control information, we separate the power by defining $\mathbf{m}_j$ as $\sqrt{\rho_j}\mathbf{u}_j$, the product of a unit vector and a scalar containing the power gain. If the beamforming vectors $\mathbf{u}_j$ are known, the values of ρ_j have a straightforward solution, but the challenge is that the $\mathbf{u}_j$ vectors are dependent on the ρ_j values. This can be solved by reformulating the problem. It is shown in [32] that if $\mathbf{m}_j$ are the optimal beamformers for the downlink and $\mathbf{u}_j$ are the solution to the problem

$$\begin{aligned} &\min_{\mathbf{u}_j, \|\mathbf{u}_j\|=1, \rho_j} \sum_{l=1}^{K} \rho_L \\ &\text{such that} \quad \frac{\mathbf{u}_j^* \mathbf{R}_j \mathbf{u}_j}{\mathbf{u}_j^* \left(\sum_{i \neq j} \rho_i \gamma_i \mathbf{R}_j + \mathbf{I} \right) \mathbf{u}_j} \geq 1 \,, \end{aligned} \tag{2.67}$$

then for all the beamformers, $\mathbf{m}_j = \sqrt{p_j}\mathbf{u}_j$ for some positive constants p_j. This problem now assumes the same form as the uplink beamforming problem, and is therefore referred to as a *virtual uplink problem*. Using this result, the following algorithm can be used to solve for the optimal $\mathbf{m}_j$ vectors.

Optimal Downlink Beamforming for Power Control

1. Initialize $\rho_j(1) = 1$ for $j = 1 \ldots K$

2. Repeat until convergence

   Beamformer Update: Find μ_j and $\mathbf{u}_j$ such that

   $$\mu_j = \max_{\|\mathbf{u}_j\|=1} \frac{\rho_j(n) \mathbf{u}_j^* \mathbf{R}_j \mathbf{u}_j}{\mathbf{u}_j^* \left(\sum_{i \neq j} \rho_i(n) \gamma_i \mathbf{R}_j + \mathbf{I} \right) \mathbf{u}_j} \,. \tag{2.68}$$

   Power control update: Let

   $$\rho_j(n+1) = \frac{\gamma_j}{\mu_j} \rho_j(n) \,. \tag{2.69}$$

3. After convergence

$$\boldsymbol{\eta} = \begin{bmatrix} \gamma_1 \sigma_1^2 & \dots & \gamma_K \sigma_K^2 \end{bmatrix}^T \tag{2.70}$$

$$[\mathbf{F}]_{i,j} = \begin{cases} \mathbf{u}_i^* \mathbf{R}_i \mathbf{u}_i & i = j \\ -\gamma_i \mathbf{u}_i^* \mathbf{R}_j \mathbf{u}_i & i \neq j \end{cases} \tag{2.71}$$

$$\mathbf{p} = \mathbf{F}^{-1} \boldsymbol{\eta} \tag{2.72}$$

$$\mathbf{m}_j = \sqrt{[\mathbf{p}]_j} \mathbf{u}_j \, . \tag{2.73}$$

The solution for μ_j and $\mathbf{u}_j$ in equation (2.68) are equivalent to the maximum eigenvalue and corresponding eigenvector of the generalized eigenvalue problem

$$\rho_j(n) \mathbf{R}_j \mathbf{u}_j = \mu_j \left(\sum_{i \neq j} \rho_i(n) \gamma_i \mathbf{R}_j + \mathbf{I} \right) \mathbf{u}_j \, . \tag{2.74}$$

An alternative solution to this problem, proposed in [32], uses semidefinite optimization tools to achieve the same result.

Several recent publications have also addressed this problem or closely related ones [30, 31, 75–85]. A variation on the problem using a 1-norm optimization is presented in [75]. The problem of maximizing throughput under a constraint on the total transmitted power is addressed in [76]. Closely related is the problem of maximizing the mean effective bandwidth, or the amount of network resource a user consumes for a given SINR [77]. The interference balancing approach to downlink beamforming does not always have a feasible solution. For example, it is possible for the matrix $\mathbf{F}$ in equation (2.71) to be singular. An analysis of conditions under which there exists a feasible solution to this class of problems is found in [78] as well as in [79], which also studies the similarities between the uplink and downlink problems. In cases where the desired set of rates does not have a feasible solution, a "rate balancing" approach could be considered, which seeks to maximize the ratio of achievable rate to desired rate [80, 81]. Each of these problems are similar, but have different constraints and optimization criteria. A comparison of these and the resulting solutions is found in [82]. One additional approach to the multi-user downlink is to integrate the dirty-paper coding and downlink beamforming methods [31,

83, 84]. A comparison of all of these approaches, including channel inversion, downlink beamforming, and dirty-paper methods, is found in [85].

Chapter 3

Closed-Form Solutions

In this chapter, we present three different non-iterative algorithms for choosing downlink transmit vectors for the case where the users in the network have multiple antennas. The first, *Block Diagonalization*, can be thought of as a generalization of channel inversion for situations with multiple antennas per user. The block-diagonalization 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 has a structure similar to that of the capacity-optimal dirty-paper methods, but without use of the dirty-paper codes. It can outperform block-diagonalization 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 [35]. 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 this chapter are that they provide efficient, closed-form solutions that yield a reasonable trade-off between performance and computational complexity.

Section 3.1 outlines the block-diagonalization algorithm for two cases: first where the transmitter has complete channel information, and second where it has incomplete or partial information. Section 3.2 describes the successive optimization algorithm for achieving power control with arbitrary rate points. Section 3.3 discusses coordinated transmit-receive processing, a framework for extending the first two algorithms to handle

larger channel geometries, and finally Section 3.4 presents simulation results comparing the algorithms under various conditions.

3.1 Block Diagonalization Algorithm

This section outlines a procedure for finding the optimal transmit vectors $\mathbf{M}_S$ such that all multi-user interference is zero. We assume the multi-user MIMO transmission model discussed in Chapter 2. Since the resulting product $\mathbf{H}_S\mathbf{M}_S$ will be block diagonal, the algorithm is referred to here as Block Diagonalization (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, as in equation (2.62). 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 3.4.

3.1.1 Block Diagonalization for Throughput Maximization

To eliminate all multi-user interference, we impose the constraint that $\mathbf{H}_j\mathbf{M}_j = \mathbf{0}$ for $i \neq j$. The achievable throughput of the resulting block-diagonal system is characterized by C_{BD} in equation (2.60). Given $\tilde{\mathbf{H}}_j$ defined as in (2.53), 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 accommodated 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 $n_T > \max\{\text{rank}(\tilde{\mathbf{H}}_1), \ldots, \text{rank}(\tilde{\mathbf{H}}_K)\}$. 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 SVD

$$\tilde{\mathbf{H}}_j = \tilde{\mathbf{U}}_j\tilde{\mathbf{\Sigma}}_j \left[\tilde{\mathbf{V}}_j^{(1)} \quad \tilde{\mathbf{V}}_j^{(0)}\right]^* , \tag{3.1}$$

where $\tilde{\mathbf{V}}_j^{(1)}$ holds the first $\tilde{L}_j$ right singular vectors, and $\tilde{\mathbf{V}}_j^{(0)}$ 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 $L_j + \tilde{L}_j - n_T \leq \bar{L}_j \leq \min\{L_j, \tilde{L}_j\}$ [86]. 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 $\tilde{\mathbf{H}}_j$. This condition is not as restrictive as it appears 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. Note that both the dimension and independence conditions allow certain cases that can not be handled by channel inversion, which 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:

$$\mathbf{H}'_S = \begin{bmatrix} \mathbf{H}_1 \tilde{\mathbf{V}}_1^{(0)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{H}_K \tilde{\mathbf{V}}_K^{(0)} \end{bmatrix}. \tag{3.2}$$

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

$$C_{\text{BD}} = \max_{\mathbf{M}'_S} \log_2 \left| \mathbf{I} + 1/\sigma_n^2 \mathbf{H}'_S \mathbf{M}'_S \mathbf{M}'^*_S \mathbf{H}'^*_S \right|. \tag{3.3}$$

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 [6]. 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

$$\mathbf{H}_j \tilde{\mathbf{V}}_j^{(0)} = \mathbf{U}_j \begin{bmatrix} \mathbf{\Sigma}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_j^{(1)} & \mathbf{V}_j^{(0)} \end{bmatrix}^* , \tag{3.4}$$

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:

$$\mathbf{M}_S = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)}\mathbf{V}_1^{(1)} & \tilde{\mathbf{V}}_2^{(0)}\mathbf{V}_2^{(1)} & \cdots & \tilde{\mathbf{V}}_K^{(0)}\mathbf{V}_K^{(1)} \end{bmatrix} \mathbf{\Lambda}^{1/2} , \tag{3.5}$$

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 (3.5), the capacity of the BD method in (2.60) becomes

$$C_{\text{BD}} = \max_{\mathbf{\Lambda}} \log_2 \left| \mathbf{I} + \mathbf{\Sigma}^2 \mathbf{\Lambda} / \sigma_n^2 \right| , \tag{3.6}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & & \\ & \ddots & \\ & & \mathbf{\Sigma}_K \end{bmatrix} . \tag{3.7}$$

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 Algorithm

1. For $j = 1, \cdots, K$:

   (i) Compute $\tilde{\mathbf{V}}_j^{(0)}$, the right null space of $\tilde{\mathbf{H}}_j$.

(ii) 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 a total power constraint P.

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

3.1.2 Block Diagonalization for Power Control

The problem with sum capacity maximization in a multi-user 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 multi-user 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 [30, 31]. 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:

$$2^{R_j} = \left| \mathbf{I} + \left(\sigma_n^2 \mathbf{I} + \sum_{i=1, i\neq j}^{K} \mathbf{H}_j \mathbf{M}_i \mathbf{M}_i^* \mathbf{H}_j^* \right)^{-1} \mathbf{H}_j \mathbf{M}_j \mathbf{M}_j^* \mathbf{H}_j^* \right| . \tag{3.8}$$

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 also be solved by water filling, as illustrated in Section 2.1.2. Extending this idea to the multi-user 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 (3.8) is possible. We discuss one based on block-diagonalization 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 block-diagonalization 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.

Power Control Block Diagonalization Algorithm

1. For each user $j = 1, \cdots, K$:

   (i) Compute $\tilde{\mathbf{V}}_j^{(0)}$, the right null space of $\tilde{\mathbf{H}}_j$.

   (ii) Compute the SVD

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

   (iii) 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 $\mathbf{\Lambda}_1, \cdots, \mathbf{\Lambda}_K$.

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

3.1.3 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, as discussed in Chapter 2, there are situations where it is possible to only obtain partial rather than full channel state information. In this section, we show how the block-diagonalization 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 [36]. One case where this model is applicable occurs when temporal averages are performed on the subspaces of $\mathbf{H}_j$ [57], 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

$$\mathbf{H}_j = \mathbf{\Phi}_{R,j}\mathbf{\Gamma}_j\mathbf{\Phi}_{T,j} \equiv \begin{bmatrix} \boldsymbol{\alpha}_{R,j}(\theta_{j,1}) & \dots & \boldsymbol{\alpha}_{R,j}(\theta_{j,L_j}) \end{bmatrix} \begin{bmatrix} \gamma_{j,1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \gamma_{j,L_j} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{T,j}^H(\phi_{j,1}) \\ \vdots \\ \boldsymbol{\alpha}_{T,j}^H(\phi_{j,L_j}) \end{bmatrix}, \tag{3.9}$$

where $\boldsymbol{\alpha}_{R,j}(\theta_{j,i})$ is the $n_{Rj} \times 1$ steering vector at receiver j for the i^{th} 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 $\mathbf{H}_i\mathbf{M}_j = 0, i \neq j$, 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 equation (2.50):

$$\tilde{\mathbf{B}}_j = \begin{bmatrix} \mathbf{B}_1^T & \dots & \mathbf{B}_{j-1}^T & \mathbf{B}_{j+1}^T & \dots & \mathbf{B}_K^T \end{bmatrix}^T. \tag{3.10}$$

Let the SVD of $\tilde{\mathbf{B}}_j$ be $\tilde{\mathbf{U}}_{B_j}\tilde{\mathbf{\Sigma}}_{B_j}\begin{bmatrix}\tilde{\mathbf{V}}_{B_j}^{(1)} & \tilde{\mathbf{V}}_{B_j}^{(0)}\end{bmatrix}^*$, 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

inter-user 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

$$C_{\text{BD}} = \max_{\mathbf{M}'_j, j=1,\cdots,K} \sum_{j=1}^{K} \log_2 |\mathbf{I} + 1/\sigma_n^2 \mathbf{A}_j \mathbf{B}_j \tilde{\mathbf{V}}_{B_j}^{(0)} \mathbf{M}'_j \mathbf{M}'^*_j \tilde{\mathbf{V}}_{B_j}^{(0)*} \mathbf{B}_j^* \mathbf{A}_j^*| \tag{3.11}$$

$$= \max_{\mathbf{M}'_j, j=1,\cdots,K} \sum_{j=1}^{K} \log_2 |\mathbf{I} + 1/\sigma_n^2 \mathbf{A}_j^* \mathbf{A}_j \mathbf{B}_j \tilde{\mathbf{V}}_{B_j}^{(0)} \mathbf{M}'_j \mathbf{M}'^*_j \tilde{\mathbf{V}}_{B_j}^{(0)*} \mathbf{B}_j^*| \tag{3.12}$$

$$\approx \max_{\mathbf{M}'_j, j=1,\cdots,K} \sum_{j=1}^{K} \left[\log_2 |1/\sigma_n^2 \mathbf{A}_j^* \mathbf{A}_j| + \log_2 |\mathbf{B}_j \tilde{\mathbf{V}}_{B_j}^{(0)} \mathbf{M}'_j \mathbf{M}'^*_j \tilde{\mathbf{V}}_{B_j}^{(0)*} \mathbf{B}_j^*| \right] \tag{3.13}$$

$$= \sum_{j=1}^{K} \log_2 |1/\sigma_n^2 \mathbf{A}_j^* \mathbf{A}_j| + \max_{\mathbf{M}'_j, j=1,\cdots,K} \sum_{j=1}^{K} \log_2 |\mathbf{M}'^*_j \tilde{\mathbf{V}}_{B_j}^{(0)*} \mathbf{B}_j^* \mathbf{B}_j \tilde{\mathbf{V}}_{B_j}^{(0)} \mathbf{M}'_j| \,. \tag{3.14}$$

Equation (3.13) is a "high SNR" approximation achieved by dropping the identity matrix in the previous equation, similar to the derivation of single-user capacity with partial channel information in Chapter 2. 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 (3.14) 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.

3.2 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 $1, \ldots, j-1$, 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 [24, 25, 27, 67] 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.

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

$$\mathbf{R}_{ni,j} = \sigma_n^2 \mathbf{I} + \sum_{i=1}^{j-1} \mathbf{H}_j \mathbf{M}_i \mathbf{M}_i^* \mathbf{H}_j^* \,. \tag{3.15}$$

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

$$\hat{\mathbf{H}}_j = \begin{bmatrix} \mathbf{H}_1^T & \mathbf{H}_2^T & \dots & \mathbf{H}_{j-1}^T \end{bmatrix}^T = \mathbf{U}_j \mathbf{\Lambda}_j \begin{bmatrix} \tilde{\mathbf{V}}_j^{(1)} & \tilde{\mathbf{V}}_j^{(0)} \end{bmatrix}^* \,. \tag{3.16}$$

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 block-diagonalization 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

$$2^{R_j} = \left| \mathbf{I} + \mathbf{M}_j'^* \tilde{\mathbf{V}}_j^{(0)*} \mathbf{H}_j^* \left(\sigma_n^2 \mathbf{I} + \sum_{i=1}^{j-1} \mathbf{H}_j \mathbf{M}_i \mathbf{M}_i^* \mathbf{H}_j^* \right)^{-1} \mathbf{H}_j \tilde{\mathbf{V}}_j^{(0)} \mathbf{M}'_j \right| . \tag{3.17}$$

such that $\mathrm{tr}\left(\mathbf{M}'_j \mathbf{M}_j'^*\right)$ 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:

$$\tilde{\mathbf{V}}_j^{(0)*} \mathbf{H}_j^* \left(\sigma_n^2 \mathbf{I} + \sum_{i=1}^{j-1} \mathbf{H}_j \mathbf{M}_i \mathbf{M}_i^* \mathbf{H}_j^* \right)^{-1} \mathbf{H}_j \tilde{\mathbf{V}}_j^{(0)} = \mathbf{Y}_j \mathbf{\Lambda}_{H,j} \mathbf{Y}_j^* \,. \tag{3.18}$$

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:

$$\mathbf{M}_j = \tilde{\mathbf{V}}_j^{(0)} \mathbf{Y}_j \mathbf{\Lambda}_{Z,j}^{1/2} \,, \tag{3.19}$$

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," the convex set of achievable rates for all users at a fixed total power level. To illustrate the properties of the two optimization algorithms, Figures 3.1 and 3.2 show two dimensional rate regions for a randomly chosen $\mathbf{H}$ matrix of dimension $\{2, 2\} \times 4$. Figure 3.2 uses the same $\mathbf{H}$ as Figure 3.1, 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 3 possible regions: the region resulting from BD, and two regions for SO, one where user one is optimized before user two (labelled "U1"), and one for the opposite case (labelled "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 3.1.2, 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, so 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 Figures 3.1 and 3.2, 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 Figure 3.1 and nonexistent in Figure 3.2. This isn't 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 Figure 3.2. In this case the BD solution

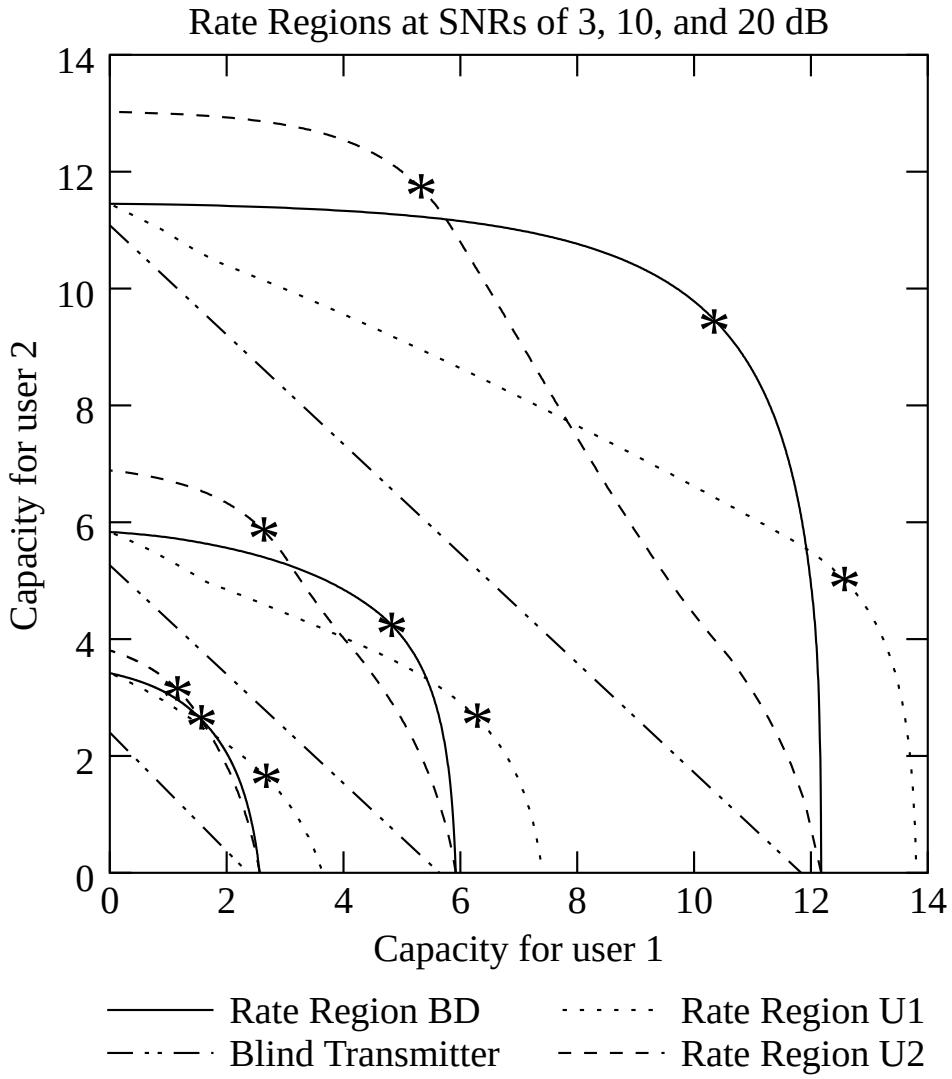

Figure 3.1: Rate Regions for a randomly generated $\mathbf{H}$ of dimension $\{2, 2\} \times 4$ at various power constraints.

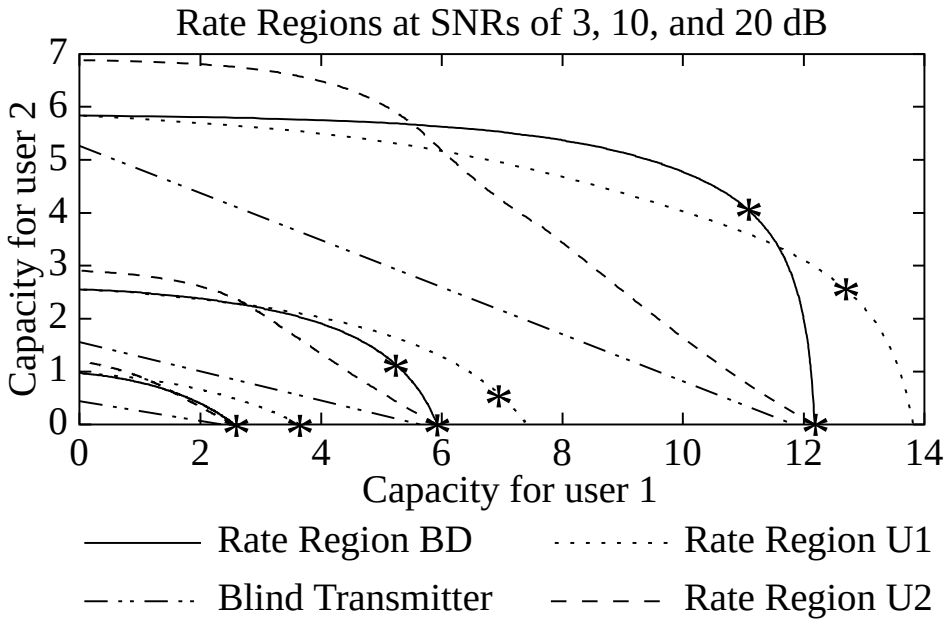

Figure 3.2: Rate regions for a "Near-Far" $\mathbf{H}$ of dimension $\{2, 2\} \times 4$ with 10 dB difference between users

results in a rate-region that is strongly biased toward user 1, but using SO with user 2 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 Section 3.4 lend support to this idea.

3.3 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 [35] for the power control problem. In [35], 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 sub-channel 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 block-wise optimization of the transmit vectors for cases where multiple data sub-channels 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 $\overline{\mathbf{H}}_S$:

$$\overline{\mathbf{H}}_S = \begin{bmatrix} \overline{\mathbf{H}}_1 \\ \vdots \\ \overline{\mathbf{H}}_K \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1^* \mathbf{H}_1 \\ \vdots \\ \mathbf{W}_K^* \mathbf{H}_K \end{bmatrix} . \tag{3.20}$$

The matrix $\overline{\mathbf{H}}_S$ has dimensions 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 inter-user 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 sub-channels 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 sub-channels to some or all users. If n_T is not sufficient to allocate a secondary sub-channel to all users, the question of which user(s) should be given additional sub-channels 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 sub-channels. Chapter 6 presents a detailed discussion of the resource allocation problem.

When the values of m_j have been determined, it is then necessary to determine the $\mathbf{W}_j$ matrices. The approach in [35] 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 block-diagonalization is given in the following algorithm description:

Coordinated Tx-Rx Block-Diagonalization Algorithm

1. For $j = 1, \cdots, K$:

   Compute the SVD $\mathbf{H}_j = \mathbf{U}_j \mathbf{\Sigma}_j \mathbf{V}_j^*$.

2. Determine m_j, the number of subchannels for each user.

3. For $j = 1, \cdots, K$:

   (i) Let $\mathbf{W}_j$ be the first m_j columns of $\mathbf{U}_j$.

   (ii) Calculate $\overline{\mathbf{H}}_j = \mathbf{W}_j^* \mathbf{H}_j$.

4. Apply the block-diagonalization 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. Also, when $m_j = 1$ for all users, the block-diagonalization 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 $\{4, 4\} \times 4$ channel for different numbers of sub-channels per user.

3.4 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 $\{1, 1, 1, 1\} \times 4$ and $\{2, 2\} \times 4$ 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 [72]. 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

$$\mathbf{M}_S = \frac{\sqrt{P}}{\|\mathbf{H}_S^\dagger\|_F} \mathbf{H}_S^\dagger \ . \tag{3.21}$$

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

$$R_{PI} = \log_2 \left| \mathbf{I} + \mathbf{H}_S \mathbf{M}_S \mathbf{M}_S^* \mathbf{H}_S^* \right| \tag{3.22}$$

$$= \log_2 \left| \mathbf{I}_{n_R} + (\rho / \|\mathbf{H}_S^\dagger\|_F^2) \mathbf{I}_L \right| \tag{3.23}$$

$$= L \log_2 \left[1 + \rho \left(\sum_{n=1}^{L} \sigma_{\mathbf{H}_S,n}^{-2} \right)^{-1} \right], \tag{3.24}$$

where L is the rank of $\mathbf{H}_S$ and $\sigma_{\mathbf{H}_S,n}$ is its n^{th} 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 block-diagonalization 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 block-diagonalization 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 (3.21), "Block Diag" is the sum capacity BD algorithm of Section 3.1.1, and "Blind Tx" is

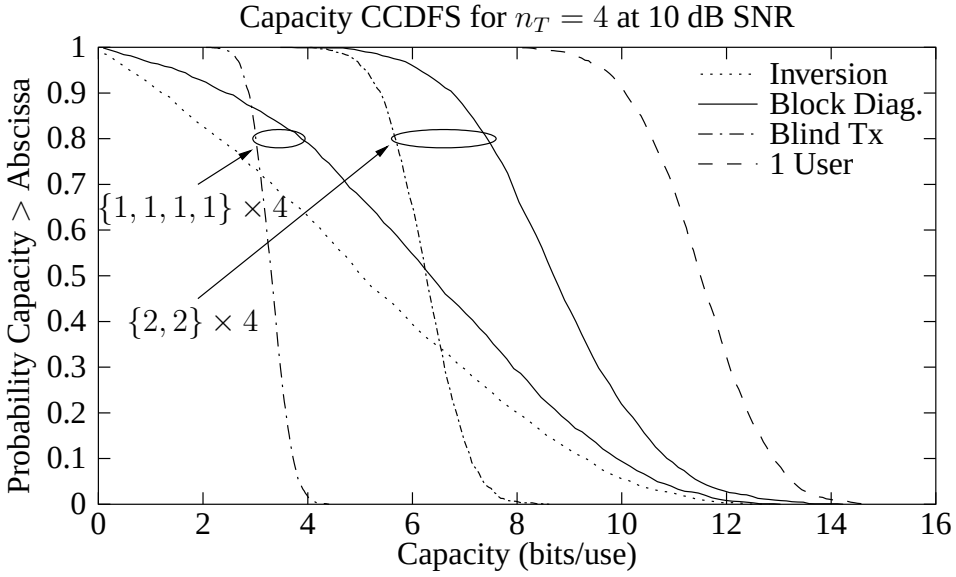

Figure 3.3: Complementary cumulative distribution functions of sum capacity for Gaussian channels for 4 transmitters.

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 $\mathbf{H}_S$.

Figure 3.3 compares the probability distributions of sum capacity for channels with dimensions $\{1,1,1,1\} \times 4$, $\{2,2\} \times 4$, and single-user 4×4. The SNR is 10 dB, and all channels are 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 Figure 3.3 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.

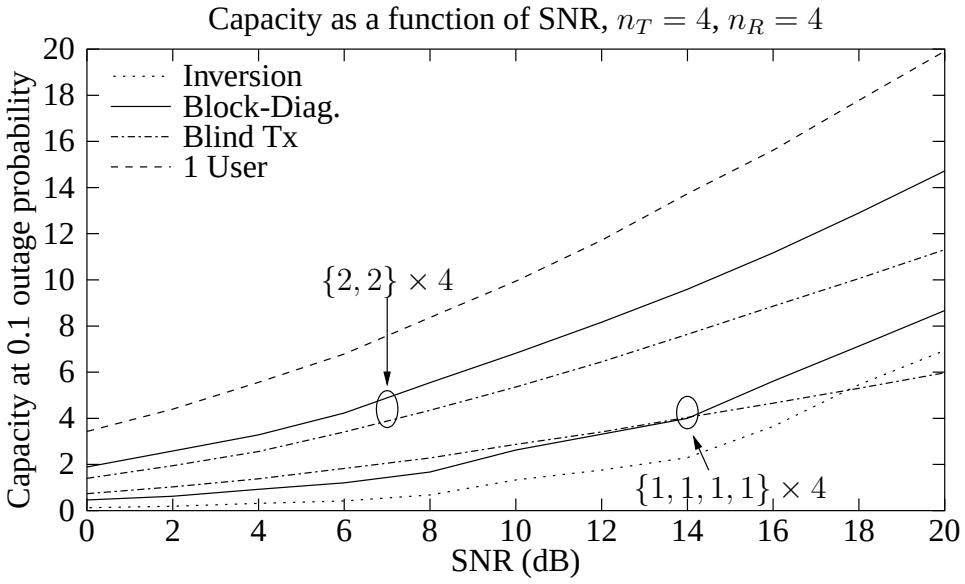

Figure 3.4: Capacity as a function of SNR at an outage probability of 0.1, for 4 transmitters.

Figure 3.4 fixes the outage probability at 0.1 and shows the same capacity curves as a function of SNR. For the $\{2,2\} \times 4$ channels, the increase in capacity is fairly constant at around 30%, and for the $\{1,1,1,1\} \times 4$ channel a significant capacity gain (also up to 30%) is only achieved at SNRs above 15 dB. Note how the availability of channel knowledge improves performance here for increasing SNR, unlike the single-user case.

Figure 3.5 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 block-diagonalization algorithms to optimally use the excess degrees of freedom available at the transmitter.

Figure 3.6 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

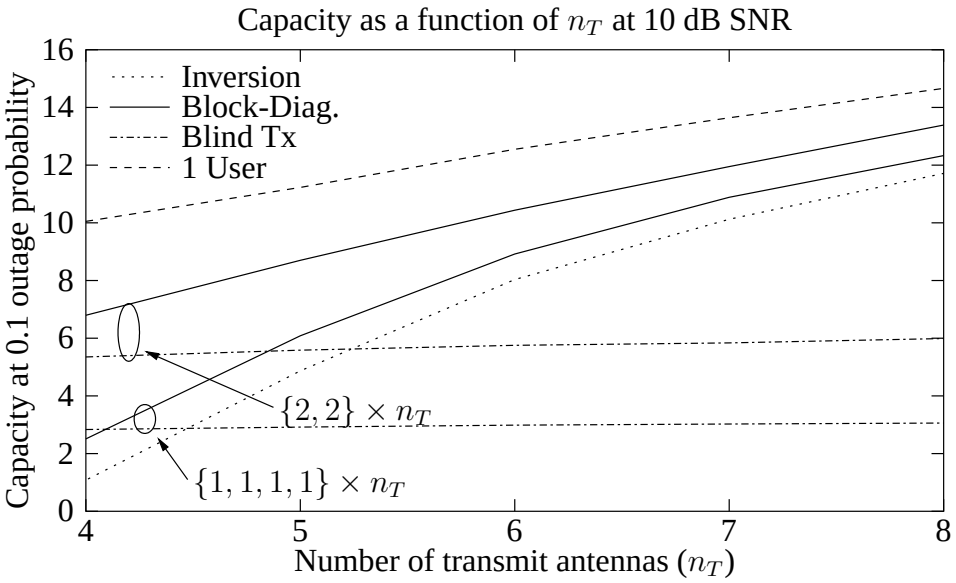

Figure 3.5: Capacity as a function of transmitter array size at an outage probability of 0.1 and an SNR of 10 dB.

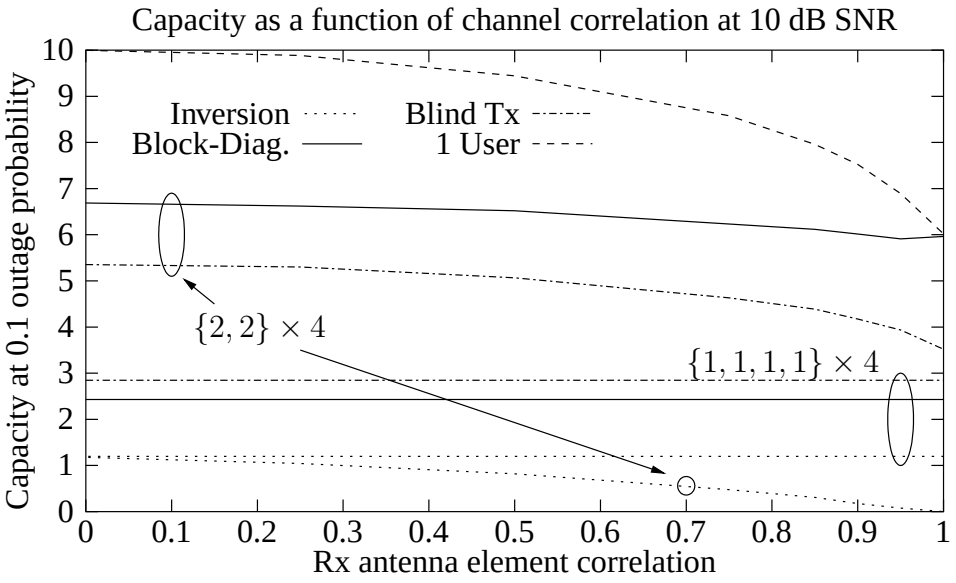

Figure 3.6: Capacity as a function of channel correlation between Rx antennas at an outage probability of 0.1 and an SNR of 10 dB.

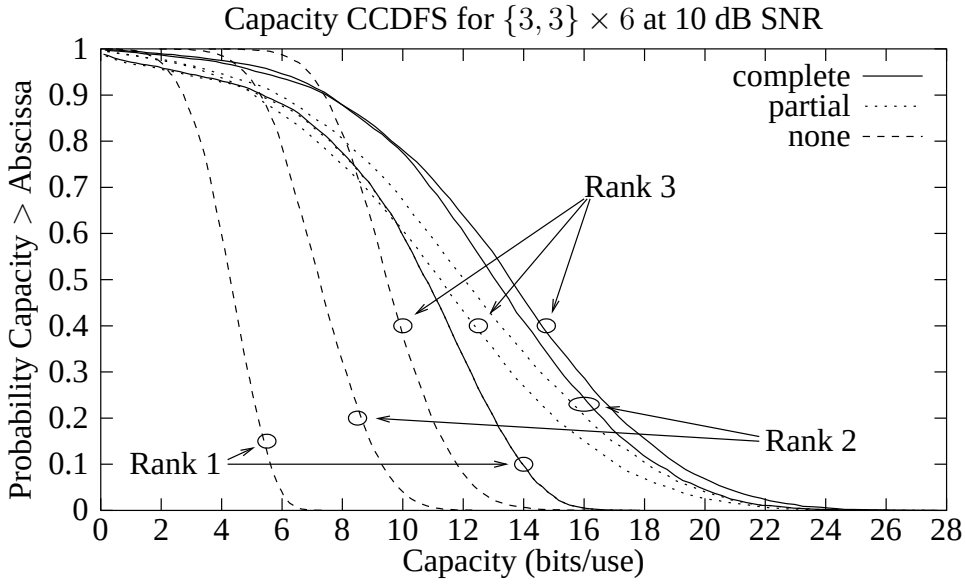

Figure 3.7: Capacity CCDFs for different cases of partial channel information.

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 block-diagonalization solution decreases slightly, but less than the other algorithms.

Figure 3.7 illustrates the performance of the block-diagonalization algorithm for the case of partial channel information. Channels were generated for this example using angle of arrival information, as described by equation (3.9). 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 block-diagonalization 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 IID complex Gaussian random variables. The plot in Figure 3.7 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 does not provide any increase in capacity.

Figure 3.8 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 3.2, "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 6 transmit antennas and 3 users. Figure 3.8(a) shows the results for the $\{2, 2, 2\} \times 6$ channel, and Figure 3.8(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 Section 3.2. 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.

Figure 3.9 compares some of the previous results with the performance of coordinated transmit-receive processing, using CCDFs similar to those in Figure 3.3. 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 1 or 2 sub-channels per user, labeled in the figure as "1 SC" or "2 SC" respectively. For the case of a single sub-channel 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 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 sub-channels 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

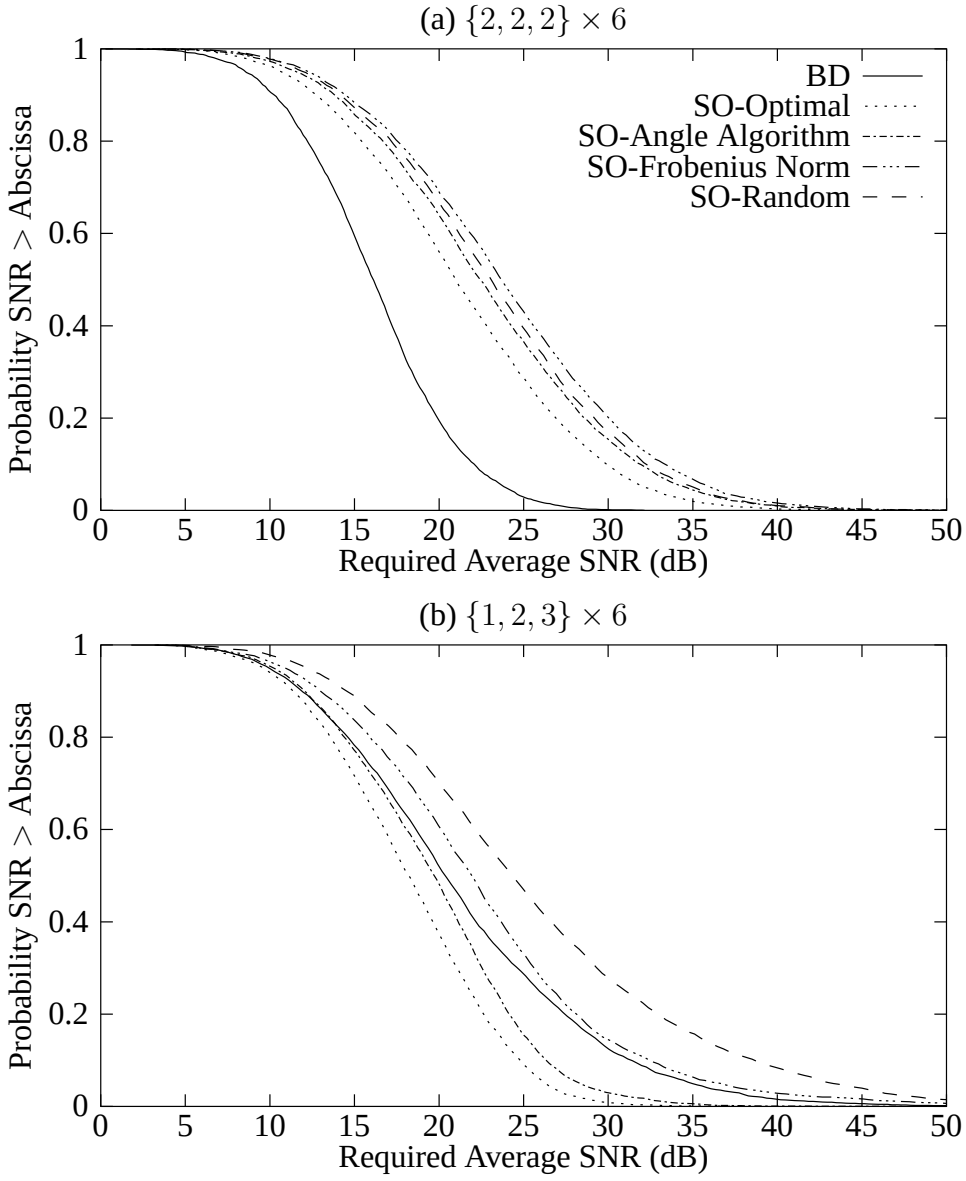

Figure 3.8: Performance of Successive Optimization compared with Block-Diagonalization for $n_T = 6$, random rate points in the interval $[2, 8]$, and random channel gains in the interval $[-6, 6]$ dB.

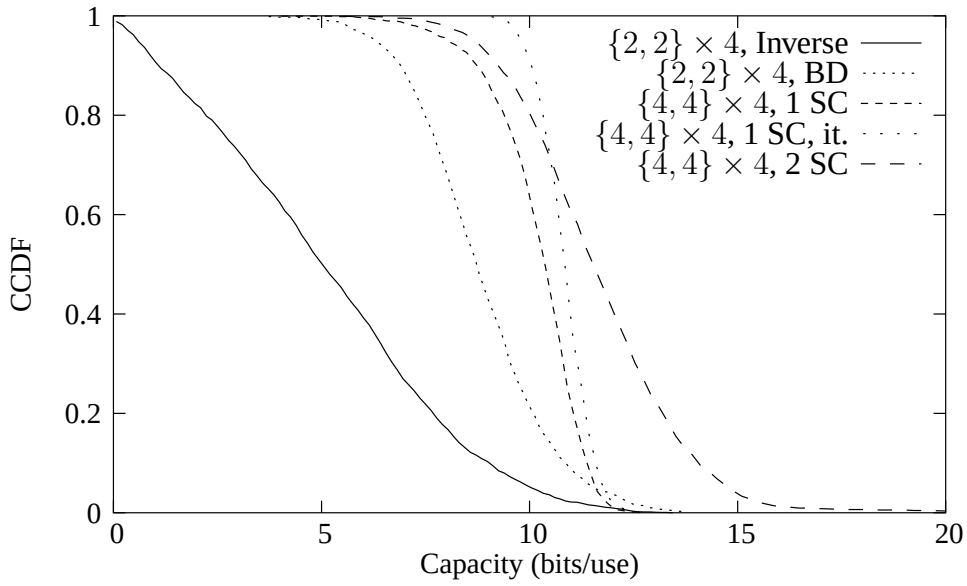

Figure 3.9: A comparison of probability densities of capacity for different channel geometries and channel decomposition algorithms at a system SNR of 10 dB.

here shows some small gains in performance, but it is inferior in some cases to the non-iterative two-sub-channel approach, illustrating the benefit of the "block optimization" that characterizes the BD and SO algorithms.

3.5 Conclusions

New approaches for optimizing information transfer in a multi-user channel have been presented here. They are sub-optimal 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 trade-off 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 which 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 can not 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 multi-user 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.

Chapter 4

Iterative Solutions

The closed-form solutions for downlink multi-user transmission presented in the last chapter are advantageous because of their simplicity, but the constraints required to make a closed-form solution can result in sub-optimal solutions. By relaxing the constraints and allowing a larger number of potential solutions, it is possible to achieve superior performance. These solutions, referred to here as "interference balancing", trade-off the inter-user interference against the additive noise in the system, and are computed iteratively. For receivers with only one antenna, this problem has been addressed in [30, 32, 73, 74]. Extensions for multiple antennas can be found in [35, 87, 88]. In [35], it is assumed that the transmitters and receivers have arrays, but only one data stream is transmitted to each user. In [87], a single-user MIMO channel with multiple data streams is considered, and it is assumed that the QoS requirement for each user is known in advance. Because the constraints on these solutions are less restrictive than the ZF constraint, they result in less transmitted power than ZF solutions optimized for minimum power, although the ZF solutions approach them in performance at high SNR. Also proposed recently is a non-iterative scheme where the transmitter uses a structure similar to an MMSE receiver [68]. This approach is similar to the iterative interference-balancing solutions because it takes noise into account, and it has been shown to have better performance than the ZF solutions. However, it assumes all users have equal SINR (QoS) requirements, which we do not assume here. In this chapter we consider the problem of a base station transmitting to a group of users, all of which may have multiple antennas. The transmission of multiple data streams to a particular user is allowed if the user's channel has sufficient rank, but the only QoS constraints are the the total rate (*i.e.*, sum of the rates of all data streams) transmitted to

that user and a specified bit error rate. We wish to achieve the QoS constraint for each user while minimizing the transmitted power.

To solve this problem we propose a new hybrid algorithm that combines the ZF and interference-balancing (IB) solutions. We first present extensions of the ZF and IB algorithms to more general scenarios than they were originally derived for (*e.g.*, multiple antennas and/or data streams per receiver, more receive than transmit antennas, imprecise channel estimates, *etc.*). We then describe a hybrid algorithm that begins by using the generalized ZF algorithm to generate initial estimates of the receivers and the sub-channel gains. Given the sub-channel gains, a bit-loading algorithm is then employed to determine the optimal power allocation. By using this information to initialize the IB algorithm, a solution is obtained that converges more quickly and uses less transmitted power than other approaches. We also show how the IB step can be modified to include noise statistics when the channel information available to the transmitter is imperfect, and we show some simplifications that arise in the special case where only one data steam is transmitted to each user, allowing a further reduction in computational cost.

The chapter will proceed as follows. Section 4.1 outlines the channel model and notation. Section 4.2 reviews the Iterative Zero-Forcing and Interference-Balancing algorithms [89]. Generalizations to the interference-balancing approach are discussed in Section 4.3. The hybrid Zero-Forcing/Interference-Balancing algorithm is then presented in Section 4.4, and Section 4.5 presents some simulation results that illustrate its performance.

4.1 Problem Definition

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. We assume the channel to be quasi-static, and that the transmitter and receiver perform linear pre- and post-processing:

$$\hat{\mathbf{d}} = \mathbf{W}^*(\mathbf{H}\mathbf{M}\mathbf{d} + \mathbf{n})\,,$$

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 $\hat{\mathbf{d}}$ by an $m \times n_R$ matrix $\mathbf{W}$. The * operator denotes Hermitian transpose.

Consider a multi-user downlink channel with K users and a single base station. The base has n_T antennas, and the j^{th} 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 4-antenna base and two 2-antenna users. The channel matrix from the base to the j^{th} user is denoted by $\mathbf{H}_j$, the associated modulation matrix by $\mathbf{M}_j$, and the transmitted data vector $\mathbf{d}_j$, which has dimension m_j. The signal at the j^{th} receiver is thus:

$$\mathbf{x}_j = \sum_{i=1}^{K} \mathbf{H}_j \mathbf{M}_i \mathbf{d}_i + \mathbf{n}_j \,, \tag{4.1}$$

and that receiver's estimate of the transmitted data is calculated using $\mathbf{W}_j$:

$$\hat{\mathbf{d}}_j = \mathbf{W}_j^* \left(\mathbf{H}_j \mathbf{M}_j \mathbf{d}_j + \sum_{i \neq j} \mathbf{H}_j \mathbf{M}_i \mathbf{d}_i + \mathbf{n}_j \right) . \tag{4.2}$$

If $\mathbf{H}_j$ has rank L_j, then $m_j \leq L_j$. The total number of data streams or subchannels allocated by the transmitter is $m = \sum_j m_j$. Any transmission method that attempts to cancel out all inter-user interference will have the requirement that $m \leq n_T$, since a beamformer using n_T antennas can null out at most $n_T - 1$ interfering signals. For interference-balancing methods, it is theoretically possible to accommodate values of $m > n_T$, but the existence of a good solution is unlikely except in certain special cases. Thus, we consider n_T to be a practical upper bound on m for all the algorithms considered here.

Clearly, if the number of users is larger than the number that can be supported using spatial multiplexing (SDMA), a likely scenario in cellular or wireless LAN systems, then other multiple access schemes must be used in conjunction with SDMA. An important question is how to allocate the users among the various available dimensions (space, time,

frequency, etc.). For example, if two users are located close to each other or have highly correlated spatial channels, it would make sense to assign them separate time or frequency slots, so that excess energy is not used to attempt to multiplex them spatially. Finding solutions to the resource allocation is discussed more extensively in Chapter 6, but it is introduced here in order to compare the approaches of choosing $m_j = 1$ versus using $m_j > 1$ in channels that can support both. For example, suppose a base has n_T antennas and there are n_T users, each having two antennas. The single-channel approach could be used to communicate with all n_T users simultaneously. On the other hand, the users could be split into two groups of $n_T/2$ users and time-multiplexed (assuming $n_T/2$ is an integer). The transmitter could use the multi-channel approach to transmit 2 data streams to each user, but since only half the time is available, the bit rate constraints would need to be doubled. The optimal strategy that minimizes transmit power, including how to group users if more than one multiple access method is used, can likely only be found by global search, but the cost function is expensive to compute. This question is studied in the simulation results section. The following sections discuss methods of jointly optimizing $\mathbf{M}_j$ and $\mathbf{W}_j$ for all users simultaneously.

4.2 Relevant Algorithms

4.2.1 Generalized Iterative Zero-Forcing

The zero-forcing approach to multi-user downlink processing [37–40] is useful because it decomposes the channel into what is effectively a set of orthogonal single-input single-output (SISO) channels. Given the set of SISO channels and the gains associated with each one, power can be allocated to the channels to solve either the power control or throughput maximization problems. These two problems are closely related and have well-known solutions for a collection of non-interfering channels. A disadvantage of zero-forcing solutions is their sensitivity to noise, but in the case of a ZF structure at the transmitter, the "noise" that causes problems is in the channel estimate rather than in the communication channel. We assume for now that the channel is known perfectly, and consider noisy estimates in Section 4.3. Another problem with the schemes in [37–40] is the restrictions they place on the size of the arrays and number of users that can be supported.

In this section we outline a generalized iterative ZF algorithm that relaxes some of these assumptions.

The fundamental idea of the ZF solution is that interference is removed by forcing $\mathbf{H}_i\mathbf{M}_j = \mathbf{0}$ for $i \neq j$. This results in the constraint that the total number of antennas at all receivers must not be larger than the number of transmit antennas. However, for cases where m_j is strictly less than n_{Rj}, the dimensionality of the problem can be reduced by viewing the "channel" for user j as not just $\mathbf{H}_j$, but $\mathbf{W}_j^*\mathbf{H}_j$, the transfer function from the transmitters to the output of the linear combiner at the receiver. The resulting structure is similar to the coordinated transmitter-receiver processing proposed in [35], but here we allow for $m_j > 1$.

Define $\overline{\mathbf{H}}_j = \mathbf{W}_j^*\mathbf{H}_j$ and

$$\tilde{\overline{\mathbf{H}}}_j = \left[\overline{\mathbf{H}}_1^T \quad \ldots \quad \overline{\mathbf{H}}_{j-1}^T \quad \overline{\mathbf{H}}_{j+1}^T \quad \ldots \quad \overline{\mathbf{H}}_K^T\right]^T . \tag{4.3}$$

The transmitter matrix $\mathbf{M}_j$ for user j will not interfere with the signal at the output of the receivers for other users if it lies in the null space of $\tilde{\overline{\mathbf{H}}}_j$. Define $\tilde{\overline{\mathbf{V}}}_j^{(0)}$ to be an orthonormal basis for the null space of $\tilde{\overline{\mathbf{H}}}_j$, and let

$$\overline{\mathbf{H}}_j\tilde{\overline{\mathbf{V}}}_j^{(0)} = \left[\mathbf{U}_j^{(1)} \quad \mathbf{U}_j^{(0)}\right]\boldsymbol{\Sigma}_j\left[\mathbf{V}_j^{(1)} \quad \mathbf{V}_j^{(0)}\right]^* \tag{4.4}$$

be the SVD of $\overline{\mathbf{H}}_j\tilde{\overline{\mathbf{V}}}_j^{(0)}$, where $\mathbf{U}_j^{(1)}$ and $\tilde{\mathbf{V}}_j^{(1)}$ correspond to the non-zero singular values of $\boldsymbol{\Sigma}_j$. For user j, capacity will be maximized (under the zero-interference constraint) when the transmitter uses $\mathbf{V}_j^{(1)}$ as its transmit vectors, the receiver uses $\mathbf{U}_j^{(1)}$ as its linear combiner weights, and the gains are allocated by water-filling on the singular values in $\boldsymbol{\Sigma}_j$ [6, 40].

Given an $\tilde{\overline{\mathbf{H}}}_j$ matrix, optimizing the transmitter and receiver for user j is straightforward, but the difficulty is that the other users' receivers ($\mathbf{W}_j$) needed to form $\tilde{\overline{\mathbf{H}}}_j$ must be known. An iterative solution can be found by assuming an initial set of $\mathbf{W}_j$ matrices, from which $\tilde{\overline{\mathbf{H}}}_j$ can be computed. This information is then used to compute $\mathbf{M}_j$ and a new $\mathbf{W}_j$ matrix. Since $\tilde{\overline{\mathbf{H}}}_j$ is now changed for all the other users, an iterative solution results in which the optimal transmitter and receiver matrices are successively recomputed until convergence. A good candidate for the initial $\mathbf{W}_j$ is to use the dominant left singular vectors of each $\mathbf{H}_j$ matrix.

To quantify algorithm convergence, define the "system" matrices

$$\overline{\mathbf{H}}_S = \left[\overline{\mathbf{H}}_1^T \quad \dots \quad \overline{\mathbf{H}}_K^T\right]^T \tag{4.5}$$

$$\mathbf{M}_S = \left[\mathbf{M}_1 \quad \dots \quad \mathbf{M}_K\right], \tag{4.6}$$

and let $\mathbf{S} = \overline{\mathbf{H}}_S \mathbf{M}_S$. Since the goal is that $\overline{\mathbf{H}}_S \mathbf{M}_S$ be (essentially) diagonal, [89] compares the maximum off-diagonal element of $\mathbf{S}$ to a certain threshold. However, since this ignores the magnitude of the diagonal elements of $\mathbf{S}$, a better alternative is to use the signal-to-interference ratio of each created channel, using the minimum value as the convergence metric:

$$\min_{i=1,\dots,K} \frac{[\mathbf{S}]_{i,i}}{\sum_{j \neq i} [\mathbf{S}]_{i,j}}, \tag{4.7}$$

where the notation $[\mathbf{S}]_{i,j}$ refers to the element in row i and column j of $\mathbf{S}$. In our simulation results, the generalized iterative ZF algorithm nearly always converges, although there are rare cases where the algorithm reaches an equilibrium point without satisfying the convergence threshold. As a result the algorithm is stopped after a fixed number of iterations. Since the main use of this method is as an initialization step in the hybrid algorithm in Section 4.4, this is not critical.

There are two special cases of the generalized ZF algorithm worth mentioning. The first arises when $\sum n_{Rj} \leq n_T$ and $m_j = n_{Rj}$ for all j. In this case, the zero-forcing solution exists after the first iteration. We compare this to the "block-diagonalization" algorithm [37–40]. If we apply the generalized solution, we have $\mathbf{H}_j = \mathbf{U}_j \mathbf{\Sigma}_j \mathbf{V}_j^*$, which results in $\overline{\mathbf{H}}_j = \mathbf{U}_j^* \mathbf{H}_j = \mathbf{\Sigma}_j \mathbf{V}_j^*$. However, the null space of $\tilde{\overline{\mathbf{H}}}_j$ in this case is the same as the null space of $\tilde{\mathbf{H}}_j$, which is defined in [39] as:

$$\tilde{\mathbf{H}}_j = \left[\mathbf{H}_1^T \quad \dots \quad \mathbf{H}_{j-1}^T \quad \mathbf{H}_{j+1}^T \quad \dots \quad \mathbf{H}_K^T\right]^T, \tag{4.8}$$

so the receiver structure can be ignored, and the two solutions are therefore equivalent.

The second special case to consider is when $m_j = 1$ for all j. The receiver structure derived from the ZF solution in this case is equivalent to a maximal ratio combiner

(MRC). This single-channel-per-user approach is the same as the combined transmitter-receiver processing scenario in [35]. In that case, the transmitter used interference balancing, and an MMSE structure was assumed at each receiver, since it is better able to cope with the resulting structured interference than an MRC receiver.

4.2.2 Interference Balancing

Zero-Forcing algorithms have the advantage of allowing the channel decomposition and power distribution problems to be solved separately, making them easily adapted to handling either rate maximization or power control. On the other hand, interference-balancing algorithms, such as those in [30, 32, 33, 35, 90], require an SINR for each sub-channel to be specified, so they are useful for the power control problem but not easily extended to rate maximization. However, they have several advantages that make them worth considering. First, they can more readily accommodate noisy channel estimates. Second, they offer a better solution to the power control problem because the zero-forcing constraint is relaxed, allowing a larger set of possible solutions. This section and the next present a review of interference-balancing algorithms and a description of how they can be extended to the case of multiple antennas and multiple sub-channels per user.

We begin by considering the special case when all users have one antenna ($n_{Rj} = 1$). The channel matrices are row vectors, which we will denote as $\mathbf{h}_j^T$. When the channels are all known to the transmitter, the SINR at the j^{th} receiver (γ_j) is a function of the channels and the transmit vectors $\mathbf{m}_j$:

$$\gamma_j = \frac{|\mathbf{h}_j^T \mathbf{m}_j|^2}{\sum_{i \neq j} |\mathbf{h}_j^T \mathbf{m}_i|^2 + \sigma_j^2} \,. \tag{4.9}$$

Since the QoS requirement for each user can often be directly mapped to an equivalent SINR (γ_j) given the available signal and code designs, the goal is to find an optimal set of transmit vectors $\mathbf{m}_j$ that satisfy all of the γ_j requirements with minimum power. Define $\mathbf{R}_j = \mathbf{h}_j^{T*} \mathbf{h}_j^T$ and represent the transmit vectors as $\mathbf{m}_j = \sqrt{\lambda_j}\mathbf{u}_j$, the product of a real

scalar and a unit-norm vector. We want to minimize the total transmitted power ($\sum \lambda_j$), while satisfying (4.9) at the specified SINR or higher:

$$\lambda_j \mathbf{u}_j^* \mathbf{R}_j \mathbf{u}_j - \sum_{j \neq k} \gamma_j \lambda_k \mathbf{u}_k \mathbf{R}_j \mathbf{u}_k \geq \gamma_j \sigma_j^2 \,. \tag{4.10}$$

There are multiple solutions to this optimization problem in the literature [32, 73, 74].

4.3 Generalized Interference-Balancing

To begin, we generalize to the case where $\mathbf{H}_j$ is a matrix (i.e., multiple antennas at each receiver) and a beamformer $\mathbf{w}_j$ is used at the receiver (i.e., one data stream per user). We also assume that the estimate of $\mathbf{H}_j$ is not perfect. If $\mathbf{H}_j$ is obtained by feedback from the receiver, the primary causes of channel estimation error will be the finite length of the training signals, and time variation in the channel. It is reasonable to assume that the statistics of the estimate $\hat{\mathbf{H}}_j$ are known. In the literature, this problem has been studied for some cases when $n_{Rj} = 1$. In [91, 92], it is assumed that either the mean or covariance of $\mathbf{h}_j$ are known, but not both. In [32, 90], robust beamforming vectors are designed to take into account best and worst cases of the covariance matrices $\mathbf{R}_j = E[\mathbf{h}_j^{T*} \mathbf{h}_j^T]$.

Here we take a slightly different approach, assuming that we have an estimate of the channel $\hat{\mathbf{H}}$, which we describe as the sum of the true channel and an error term:

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{N}. \tag{4.11}$$

We assume that the error term is Gaussian with zero mean, and a covariance that is characterized by two matrices, a row covariance $\mathbf{R}_R$ and a column covariance $\mathbf{R}_C$, whose elements are defined as:

$$[\mathbf{R}_R]_{i,j} = E\left\{ [\mathbf{N}]_{i,l} [\mathbf{N}]_{j,l}^* \right\} \tag{4.12}$$

$$[\mathbf{R}_C]_{i,j} = E\left\{ [\mathbf{N}]_{k,i} [\mathbf{N}]_{k,j}^* \right\}. \tag{4.13}$$

We assume a separable covariance function, meaning that $E\left\{ [\mathbf{N}]_{i,j} [\mathbf{N}]_{k,l}^* \right\} = [\mathbf{R}_R]_{i,k} [\mathbf{R}_R]_{j,l}$. Since $\mathbf{R}_R$ and $\mathbf{R}_C$ are covariance matrices, there exist square roots $\mathbf{R}'_R$ and $\mathbf{R}'_C$ such that $\mathbf{R}_R = \mathbf{R}'_R \mathbf{R}'^*_R$ and $\mathbf{R}_C = \mathbf{R}'_C \mathbf{R}'^*_C$. Thus, the estimation error matrix can be represented

in terms of a matrix $\mathbf{X}$ with independent, zero-mean, unit-variance complex Gaussian elements:

$$\mathbf{N} = \mathbf{R}'_R \mathbf{X} \mathbf{R}'^*_C \,. \tag{4.14}$$

The above model does not account for all possible cases, but it is appropriate for many situations encountered in practice. One example is when $\mathbf{H}$ is estimated from training data and sent through a feedback channel to the transmitter. Let $\mathbf{T}$ be the matrix of training data, such that the received training signal $\mathbf{Q}$ is

$$\mathbf{Q} = \mathbf{H}\mathbf{T} + \mathbf{Y} \,, \tag{4.15}$$

where $\mathbf{Y}$ represents an additive noise term. Then the least squares estimate of $\mathbf{H}$ is $\hat{\mathbf{H}} = \mathbf{Q}\mathbf{T}^*(\mathbf{T}\mathbf{T}^*)^{-1}$, and the error in the estimate is $\mathbf{Y}\mathbf{T}^*(\mathbf{T}\mathbf{T}^*)^{-1}$. If the elements of $\mathbf{Y}$ are uncorrelated, then $\mathbf{R}_R = \mathbf{I}_{n_R}$, and $\mathbf{R}_C = \mathbf{T}\mathbf{T}^*$ (since $\mathbf{T}$ is a design parameter, it can be chosen such that $\mathbf{T}\mathbf{T}^* = \mathbf{I}_{n_T}$). One situation that is more likely to produce colored estimates of $\mathbf{H}$ is when significant time variation in the channel is present. If the statistics of the time variations are separable in time and space (not an unreasonable assumption), then (4.14) will hold in this case as well.

Let d_j be the symbol transmitted to user j. If user j uses a predetermined unit-norm beamforming vector $\mathbf{w}_j$, then the estimate of the transmitted symbol at the receiver will be:

$$\hat{d_j} = \mathbf{w}_j^* \left(\sum_{i=1}^{K} \mathbf{H}_j \mathbf{m}_i d_i + \mathbf{n}_j \right) \tag{4.16}$$

$$= \underbrace{\mathbf{w}_j^* \left(\hat{\mathbf{H}}_j + \mathbf{N}_j \right) \mathbf{m}_j d_j}_{\text{signal } (\hat{d}_{j(\text{sig})})} + \underbrace{\sum_{i \neq j} \mathbf{w}_j^* \left(\hat{\mathbf{H}}_j + \mathbf{N}_j \right) \mathbf{m}_i d_i + \mathbf{w}_j^* \mathbf{n}_j}_{\text{noise } (\hat{d}_{j(\text{noise})})} \tag{4.17}$$

The transmitter views the "channel" as being the transfer function from the output of the transmitting array to the output of the receiver's beamformer $\mathbf{w}_j$, and the covariance of this channel is:

$$\mathbf{R}_j = E\left[\mathbf{H}_j^* \mathbf{w}_j \mathbf{w}_j^* \mathbf{H}_j\right] = \hat{\mathbf{H}}_j^* \mathbf{w}_j \mathbf{w}_j^* \hat{\mathbf{H}}_j + E\left[\mathbf{N}_j^* \mathbf{w}_j \mathbf{w}_j^* \mathbf{N}_j\right] \tag{4.18}$$

$$\begin{aligned} E\left[\mathbf{N}_j^* \mathbf{w}_j \mathbf{w}_j^* \mathbf{N}_j\right] &= E\left[\mathbf{R}_{C_j}'^* \mathbf{X}_j^* \mathbf{R}_{R_j}'^* \mathbf{w}_j \mathbf{w}_j^* \mathbf{R}_{R_j}' \mathbf{X}_j \mathbf{R}_{C_j}'\right] \\ &= \mathbf{R}_{C_j} \operatorname{tr}(\mathbf{R}_{R_j} \mathbf{w}_j \mathbf{w}_j^*) = (\mathbf{w}_j^* \mathbf{R}_{R_j} \mathbf{w}_j) \mathbf{R}_{C_j} \end{aligned} \tag{4.19}$$

$$\mathbf{R}_j = \hat{\mathbf{H}}_j^* \mathbf{w}_j \mathbf{w}_j^* \hat{\mathbf{H}}_j + (\mathbf{w}_j^* \mathbf{R}_{R_j} \mathbf{w}_j) \mathbf{R}_{C_j} \,. \tag{4.20}$$

Given $\mathbf{R}_j$, and assuming the noise vector $\mathbf{n}$ has covariance $\sigma_n^2 \mathbf{I}$, the SINR for user j becomes:

$$\gamma_j = \frac{E[\hat{d}_{j(\text{sig})}^2]}{E[\hat{d}_{j(\text{noise})}^2]} = \frac{\lambda_j \mathbf{u}_j^* \mathbf{R}_j \mathbf{u}_j}{\sigma_n^2 + \sum_{i \neq j} \lambda_i \mathbf{u}_i^* \mathbf{R}_j \mathbf{u}_i} \,. \tag{4.21}$$

The result is that the problem is now identical to that posed in equation (4.10), except that the definition of $\mathbf{R}_j$ now includes information regarding channel estimation error and user j's beamformer. Solving this system of equations would require of course that the receiver weight vectors $\mathbf{w}_j$ be known in advance. Since the transmitter knows what the statistics of the received signal at each user will be, it can predict $\mathbf{w}_j$ for some receiver structures (eg. MMSE, maximal ratio combiner, etc.). If the transmitter first guesses an initial set of $\mathbf{w}_j$ vectors, it is then possible to calculate all $\mathbf{R}_j$ and solve (4.10) for the $\mathbf{m}_j$ vectors. The transmitter's estimate of each $\mathbf{w}_j$ can then be updated according to the known receiver structure. Repeated alternating recalculation of $\mathbf{w}_j$ and $\mathbf{m}_j$ will reduce the required transmitted power until it converges to a minimum [33, 35, 87]. There are situations where equation (4.10) does not have a solution (see [32, 73, 74]). Care must be taken in choosing the initial set of $\mathbf{w}_j$ vectors, because it is possible to choose an initial vector that does not lead to a solution, even when one exists. Furthermore, a poorly chosen initialization point can also increase the number of iterations required for convergence. We propose using $\mathbf{w}_j$ vectors taken from either the multi-user zero-forcing solution, or the left singular vectors of the $\mathbf{H}_j$ matrices.

The above method can be generalized one step further to accommodate the transmission of multiple data streams per user, provided the channel dimensions and rank

allow it. Assume that the SINR requirements for each sub-channel have been specified; *i.e.*, let $\gamma_{j,k}$ represent the SINR required for the k^{th} data stream and let $\mathbf{w}_{j,k}$ be the corresponding column of $\mathbf{W}_j$ for user j. If we define

$$\mathbf{R}_{j,k} = \hat{\mathbf{H}}_j^* \mathbf{w}_{j,k} \mathbf{w}_{j,k}^* \hat{\mathbf{H}}_j + (\mathbf{w}_{j,k}^* \mathbf{R}_{R_j} \mathbf{w}_{j,k}) \mathbf{R}_{C_j} \,, \tag{4.22}$$

the SINR is thus:

$$\gamma_{j,k} = \frac{\lambda_{j,k} \mathbf{u}_{j,k}^* \mathbf{R}_{j,k} \mathbf{u}_{j,k}}{\sigma_n^2 + \sum_{j \neq l, k \neq m} \lambda_{l,m} \mathbf{u}_{l,m}^* \mathbf{R}_{j,k} \mathbf{u}_{l,m}}. \tag{4.23}$$

Given a set of SINR requirements $\gamma_{j,k}$, solving the resulting set of inequalities like equation (4.10) is straightforward. The challenge is determining the optimal SINR requirements. A simple solution would be to use equal power for all channels, but this could easily result in a situation where sub-channels with low gain have unusually high amounts of power forced into them. In iterative interference-balancing algorithms, this results in more interference for other channels to deal with, and therefore solutions that require higher total power, as will be seen in the simulation results.

If we assume as we did in the derivation of the generalized ZF method that m_j, the number of sub-channels for user j, is determined in advance, an alternative approach is to estimate the best power distribution by using the first m_j singular values of $\mathbf{H}_j$ and water-filling. This approach, referred to here as the "SVD initialization," will likely result in a power distribution that is superior to equal power, but will still be sub-optimal because it does not take into account the interaction with the channels of other users. The presence of high correlation between two users' channels in certain dimensions (e.g., due to a strong common scatterer) can result in the algorithm allocating high transmit power to those dimensions, and lead to increased interference. In such cases, a better solution would distribute more power to spatial dimensions that are not common to both users.

4.4 Hybrid Zero-Forcing/Interference-Balancing Algorithm

We propose as an alternative approach a hybrid between the generalized ZF solution and existing interference-balancing solutions. The general idea is to begin by finding the zero-forcing solution, use the resulting sub-channel gains for determining the

power distribution, and the resulting $\mathbf{W}_j$ matrices as a starting point for an interference-balancing solution. This approach has two advantages. The first is that the resulting power distribution takes into account the interaction of all users' channels, and the second is faster convergence of the interference-balancing step due to the fact that the ZF step provides an excellent initialization.

The problem with using either the SVD or ZF initializations to determine power allocation is that the subsequent interference-balancing step will change both the transmitter and receiver vectors, resulting in different gains for each sub-channel, and making any previous power distribution sub-optimal. One possible solution is to add an additional outer iteration loop that re-allocates power based on the new sub-channel gains and repeats the entire process. However, such an approach would likely be computationally prohibitive. This problem is particularly challenging if the classic water-filling solution is used, which assumes signal constellations with infinite granularity. If we focus instead on more practical applications which will likely have a predefined discrete set of available constellations and code designs, we are then interested in finding the optimal power distribution using the available signal designs, a problem referred to in the literature as "bit-loading" [93–96]. The consequence of this approach for the hybrid algorithm is that there are now a finite number of possible solutions. Thus, solutions based on an initial estimate of the sub-channel gains are likely to be close to or even equal to the optimal power distribution, making the approach of estimating the power distribution without an outer iteration loop feasible.

We briefly explain the bit-loading algorithm used here. Bit-loading is a well-studied problem that is most well known for its application to multi-carrier modulation schemes, but that has recently been applied to the single-user MIMO channel [97, 98]. The bit-loading problem is illustrated in Figure 4.1, where bits must be allocated to a set of parallel channels with different noise powers similar to the water-filling problem illustrated in Figure 2.1. The difference is that this is an optimization problem over a discrete set of possible solutions. Solutions have been proposed for both power control and rate optimization. We are interested in power control, which in this case means determining the optimal bit distribution over the available sub-channels, given the sub-channel gains

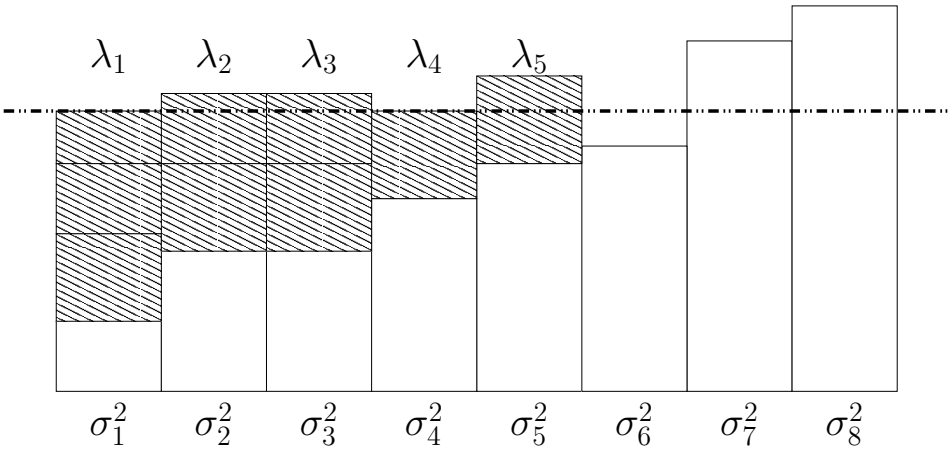

Figure 4.1: An illustration of the bit-loading problem.

and a total required transmission rate and bit error rate. This is equivalent to the "Margin Maximization Problem" in [95]. The original algorithm for this problem is described in [93, 94], and requires $\mathcal{O}(L \log L)$ computations for L available channels. More recently, fast algorithms have been developed that are able to find the optimal distribution in $\mathcal{O}(L)$ computations [95, 96], but we use the original algorithm here because of its simplicity and the small number of sub-channels we are dealing with. A brief summary of the algorithm is as follows: let $P_{n,l}$ be the power required to transmit n bits through the l^{th} sub-channel given the SINR of the sub-channel, and define $P_{0,l} = 0$. Initialize all channels to zero bits, and repeat the following steps until the desired total rate is achieved:

1. Compute $\Delta P_l = P_{N+1,l} - P_{N,l}$, where N is the number of bits currently assigned to sub-channel l.

2. Let $k = \arg\min_{l=1\ldots L} \Delta P_l$, increment N for channel l, and recompute ΔP_k.

Note that this method could also work for a code with non-integer rate R, as long as there exist codes of rate $2R$, $3R$, etc., up to the maximum possible rate.

A problem not yet addressed is how to determine values for m_j. If we assume that $n_T \geq n_{Rj}$, then m_j will be limited by the minimum value of n_{Rj} and $\mathrm{rank}(\mathbf{H}_j)$. If it is known that users are likely to have limited channel rank or receiver array size, it may make sense to assign $m_j = 1$ for all users in advance. We show in the next section how

this special case can result in significant computational savings, but it will likely be sub-optimal if some users have channel rank greater than one. The best choice of m_j in this case becomes a function of the other users with which user j is sharing a channel, and is difficult to find without an exhaustive search. An alternative solution, which we use here, is to determine how many sub-channels user j would use in isolation by bit-loading using the singular values of $\mathbf{H}_j$ as the sub-channel gains, and letting m_j represent the number of sub-channels with non-zero power assigned to them. We evaluate the performance of assigning single versus multiple sub-channels per user later in the simulation results. The following is a complete outline of the proposed hybrid algorithm:

Hybrid Zero-Forcing/Interference-Balancing Algorithm

1. Determine m_j, the number of sub-channels per user, by bit-loading using the singular values of $\mathbf{H}_j$ and the rate constraint R_j.

2. Given R_j and m_j for each user, compute the Generalized Iterative Zero-Forcing solution and use bit-loading to determine the sub-channel power distribution subject to meeting the rate constraints. Let $\gamma_{j,1} \ldots \gamma_{j,m_j}$ represent the resulting sub-channel SINRs for user j and let $\mathbf{w}_{j,l}$ be the l^{th} column of $\mathbf{W}_j$.

3. Define $\mathbf{R}_{j,k} = \hat{\mathbf{H}}_j^* \mathbf{w}_{j,k} \mathbf{w}_{j,k}^* \hat{\mathbf{H}}_j + (\mathbf{w}_{j,k}^* \mathbf{R}_{R_j} \mathbf{w}_{j,k}) \mathbf{R}_{C_j}$. Find the unit vectors $\mathbf{u}_{j,k}$ and power coefficients $\sqrt{\lambda_{j,k}}$ such that $\sum \lambda_{j,k}$ is minimized and

$$\lambda_{j,k} \mathbf{u}_{j,k}^* \mathbf{R}_{j,k} \mathbf{u}_{j,k} - \sum_{j \neq l, k \neq m} \gamma_{j,k} \lambda_{l,m} \mathbf{u}_{l,m}^* \mathbf{R}_{j,k} \mathbf{u}_{l,m} \geq \gamma_{j,k} \sigma_{j,k}^2 \,,$$

   using, for example, an available algorithm from [32].

4. Repeat until convergence (optional):

   (i) Recalculate the predicted receiver weights according to the algorithm used at the receiver. For MMSE, this is

$$\mathbf{w}_{j,l} = \left[\mathbf{H}_j \left(\sum_{j,k} \lambda_{j,k} \mathbf{u}_{j,k} \mathbf{u}_{j,k}^* \right) \mathbf{H}_j^* + \sigma_j^2 \mathbf{I} \right]^{-1} \mathbf{H}_j \mathbf{u}_{j,l},$$

assuming that the noise at the receiver is spatially white with variance σ_j^2. If the receiver has an MRC structure, the weights are

$$\mathbf{w}_{j,l} = \mathbf{H}_j \mathbf{u}_{j,l}.$$

In either case, the weight vectors must be normalized so that $\mathbf{w}_{j,l}^* \mathbf{w}_{j,l} = 1$.

(ii) Recalculate $\mathbf{u}_{j,k}$ and $\lambda_{j,k}$ for all j and k (repeat step 3).

The reason that step 4 in the above algorithm is optional is that after step 3 is completed, a solution exists that satisfies all constraints, although sub-optimally. Step 4 generally accounts for most of the computational cost of the algorithm, and as will be seen in the next section, only a relatively small gain in performance. At each step, the required power to achieve the solution will be reduced slightly, so convergence is determined by comparing the required power to that of the previous iteration. If P_n represents the total transmitted power at iteration n, the algorithm is considered to have converged when

$$10 \log_{10} \left(\frac{P_{n-1}}{P_n} \right) < \epsilon \text{ dB}, \tag{4.24}$$

for some ϵ. Note that this implies a minimum of 2 iterations.

4.4.1 The Single-Channel Case

An important special case of the Hybrid ZF/IB algorithm is when $m_j = 1$ for all users. This can be a natural consequence of channel dimensions or rank, but it may also be by design in order to simplify the system. There are some natural simplifications that arise in this situation, which can substantially reduce the computational cost of the hybrid algorithm.

The problem of transmitting a single data stream to each user with arrays of arbitrary dimension is discussed in [35], and the SINR is characterized by equation (4.21). The hybrid solution uses the generalized ZF algorithm to estimate a good power distribution among each user's sub-channel and reduce the number of iterations in the interference-balancing step. With only one sub-channel per user, there is no need for sub-channel power distribution, but reducing convergence time is still important.

Normally, the generalized zero-forcing method in Section 4.2.1 uses a lengthy block-diagonalization procedure, but in the case of $m_j = 1$, one can simply use a pseudo-inverse: $\mathbf{M}_S = \overline{\mathbf{H}}_S^{\dagger}$ (the zero-forcing solution in [25] is computed this way). The difference between this solution and the block-diagonalization procedure is that in this case the columns of $\mathbf{M}_S$ do not have unit length. Given the SVD $\mathbf{H}_j = \mathbf{U}_j\mathbf{B}_j\mathbf{V}_j^*$, the zero-forcing approach will choose $\mathbf{m}_j$ as the first column of $\mathbf{V}_j^*$, since $m_j = 1$. The received signal $\mathbf{H}_j\mathbf{m}_j d_j + \mathbf{n}_j$ simplifies to $\mathbf{u}_j\beta_j d_j + \mathbf{n}_j$, where $\mathbf{u}_j$ is the first column of $\mathbf{U}_j$ and β_j is the first singular value of $\mathbf{B}_j$. The optimal $\mathbf{w}_j$ is $\mathbf{u}_j$, which is equivalent to a maximal ratio combiner (note that this result does not hold when $m_j > 1$). Thus is it possible to define $\mathbf{w}_j = \mathbf{H}_j\mathbf{m}_j$ and avoid computing an SVD to find $\mathbf{w}_j$. However, in order to repeat this process iteratively without causing the rows of $\overline{\mathbf{H}}_S$ to diverge, it is necessary to normalize $\mathbf{w}_j$ at each step. In addition to providing a good initialization point, another reason for using the generalized zero-forcing solution with the hybrid algorithm is its use in predicting a good power distribution among the sub-channels for a particular user. For a single channel per user, the SINR requirement is known without this step, so the first row of $\mathbf{H}_j$ is used as the initial estimate of $\overline{\mathbf{H}}_j$. These steps are illustrated in the single-channel version of the hybrid algorithm below.

Fast Single-Channel SINR Balancing Algorithm

1. Initialization: For $j = 1, \ldots, K$, set $\mathbf{w}_j = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T$
2. Repeat until convergence:
   (i) $\overline{\mathbf{H}}_S = \begin{bmatrix} \mathbf{w}_1^*\mathbf{H}_1 \\ \vdots \\ \mathbf{w}_K^*\mathbf{H}_K \end{bmatrix}$.
   (ii) $\mathbf{M}_S = \overline{\mathbf{H}}_S^{\dagger}$.
   (iii) For $j = 1, \ldots, K$, set $\mathbf{w}_j = \mathbf{H}_j\mathbf{m}_j / \sqrt{\mathbf{m}_j^*\mathbf{H}_j^*\mathbf{H}_j\mathbf{m}_j}$.

3. Let $\mathbf{R}_j = \mathbf{H}_j^* \mathbf{w}_j \mathbf{w}_j^* \mathbf{H}_j$. Find the unit vectors $\mathbf{u}_j$ and power coefficients $\sqrt{\lambda_j}$ such that $\sum \lambda_j$ is minimized and

$$\lambda_j \mathbf{u}_j^* \mathbf{R}_j \mathbf{u}_j - \sum_{j \neq k} \gamma_j \lambda_k \mathbf{u}_k^* \mathbf{R}_j \mathbf{u}_k \geq \gamma_j \sigma_j^2,$$

and then let $\mathbf{m}_j = \lambda_j \mathbf{u}_j$.

4. Repeat until convergence:

   (i) Recalculate predicted receiver weights, and normalize so that $\mathbf{w}_j^* \mathbf{w}_j = 1$.

   (ii) Recalculate all $\mathbf{m}_j$ (repeat step 3).

Step 2 does not necessarily converge to a diagonal solution for $\overline{\mathbf{H}}_S \mathbf{M}_S$, but it generally converges to a diagonally dominant solution in just a few iterations. As before, after Step 3 is completed, a feasible solution exists, so Step 4 is optional. In addition to the cost of iterating to find the minimum power solution, a further problem is the unpredictability of how many iterations will be required. For situations where computational cost is critical, one could use a "fixed-cost" approach, where Step 2 is repeated a fixed number of times and Step 4 is omitted. This results in a feasible solution that comes at a known cost.

4.5 Simulation Results

For simulation results involving bit-loading, we assume that the required rate for each user is an integer number of bits/use. The available set of signals are the QAM constellations from 1-8 bits/symbol (including BPSK and QPSK as special cases). The power requirement is based on the upper bound on the symbol error rate from [99] (equation 5-2-80):

$$P_M \leq 2 \ \mathrm{erfc}\left(\sqrt{\frac{3k}{2(M-1)}\gamma_b}\right), \tag{4.25}$$

where k is the number of bits, $M = 2^k$, and γ_b is the average SNR per bit. The required bit error rate is fixed at 10^{-5}, assuming no additional coding. Except where otherwise noted, the channels used in the simulations were randomly generated from an uncorrelated complex Gaussian distribution.

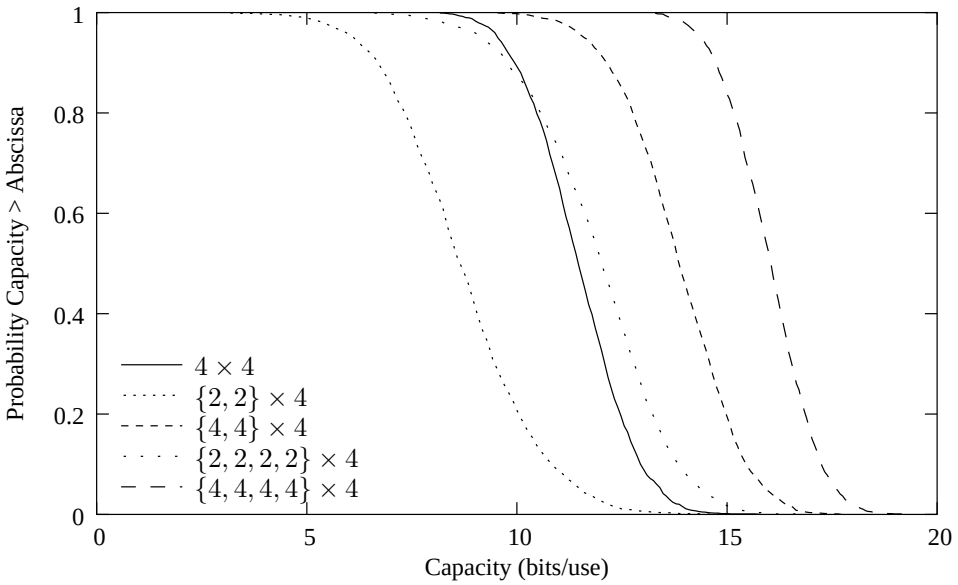

Figure 4.2: A performance comparison of the generalized iterative ZF algorithm for various antenna configurations at SNR=10 dB.

4.5.1 Zero-Forcing Performance

Figure 4.2 compares the performance of the generalized ZF algorithm for various antenna configurations. The system SNR is fixed at 10 dB (total transmitted power relative to mean noise power). The capacity numbers refer to the maximum sum capacity achievable under the constraints, and does not assume the use of bit-loading. For comparison, we include results for the single-user 4×4 MIMO channel. It is interesting to note that there is a performance gain from the $\{2, 2, 2, 2\} \times 4$ to the $\{4, 4\} \times 4$ channel, though both have the same total number of receivers. This illustrates the gain achievable using the block-wise processing that results from allowing two sub-channels per user.

4.5.2 Hybrid ZF/IB Performance

In this example, the Hybrid ZF/IB algorithm is compared with three other algorithms. The first is the ZF algorithm alone without the interference-balancing step. For an accurate comparison, we use bit-loading to determine the power distribution among the sub-channels instead of the water-filling approach used in [40]. The second is interference balancing with the SVD initialization described in Section 4.4, and is referred to in

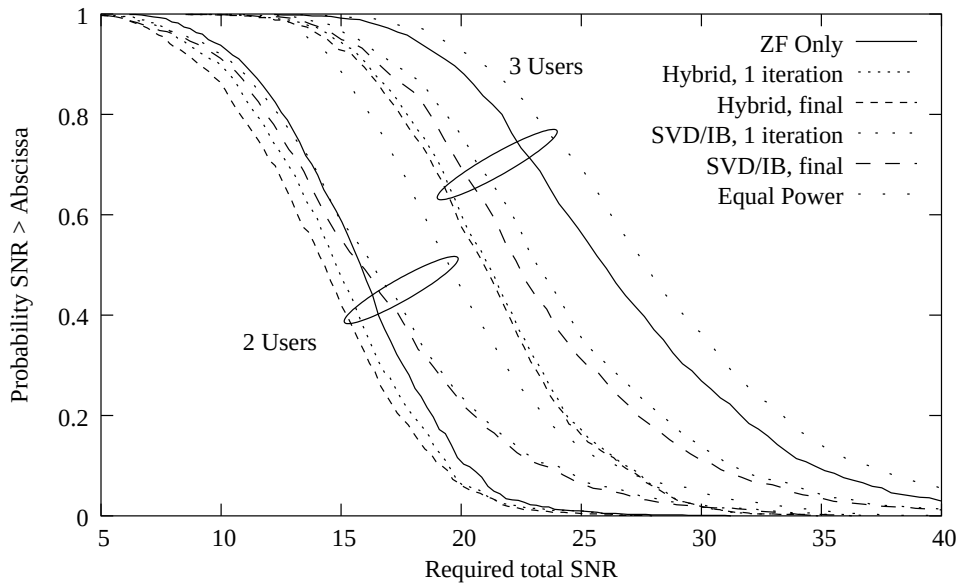

Figure 4.3: A comparison of the power minimization capability of the algorithms for $\{2, 2\} \times 4$ and $\{1, 2, 3\} \times 6$ channels.

the plots as "SVD/IB". The third algorithm uses interference balancing with equal power distribution among the sub-channels, and the SVD initialization. We also include results for the Hybrid solution and SVD initialization after only one iteration of the interference-balancing (IB) step, which yields a feasible solution for the QoS requirements without achieving optimality.

We consider two scenarios. The first involves a $\{2, 2\} \times 4$ channel where both channels are uncorrelated complex Gaussian, with zero mean and unit variance. In the second, a $\{1, 2, 3\} \times 6$ channel is assumed, where the channels are also uncorrelated and Gaussian, but with attenuations varying uniformly from 0-10 dB. In both cases, the rate requirements for the users are integers chosen randomly in the range of 2-8 bits/use per user. These same channel dimensions, attenuations, and transmission rates are used in Figures 4.3-4.6. Figure 4.3 shows the total SNR required to transmit at the requested QoS for both scenarios. The hybrid algorithm achieves a consistent performance improvement over all other solutions, up to 5 dB when compared with equal power distribution. The ZF solution performs well in the two user-case, but much worse in the three-user case. This is likely due to the variable channel attenuations, since the results in [40] show that the ZF

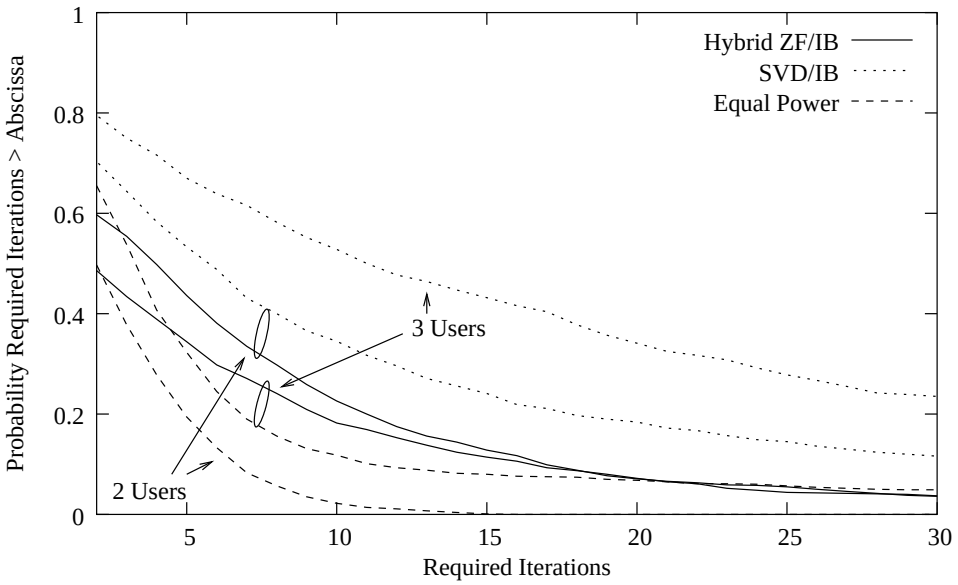

Figure 4.4: A comparison of the number of iterations required for convergence for $\{2, 2\}\times 4$ and $\{1, 2, 3\} \times 6$ channels.

solution works best when the users have similar power levels. The excellent performance of the hybrid solution after only one iteration makes the "fixed-cost" approach an attractive alternative.

Figure 4.4 illustrates the number of iterations required for the scenarios considered in Figure 4.3 for the hybrid algorithm, SVD initialization, and equal power distribution. In this plot and other subsequent plots illustrating convergence speed, convergence was defined as occurring when two iterations produce a reduction in transmitted power of less than 0.01 dB. The hybrid algorithm converges consistently faster than SVD/IB for both cases, with a larger improvement in the three-user scenario, while IB with an equal power distribution converges faster in most cases.

4.5.3 A comparison of MMSE and MRC receivers

The hybrid zero-forcing/interference-balancing algorithm can be configured for any receiver structure that is a function of information available to the transmitter. Figure 4.5 compares the performance of the algorithm with MMSE and MRC receivers, using the same two-user and three-user setup as in the previous example. The MMSE receiver

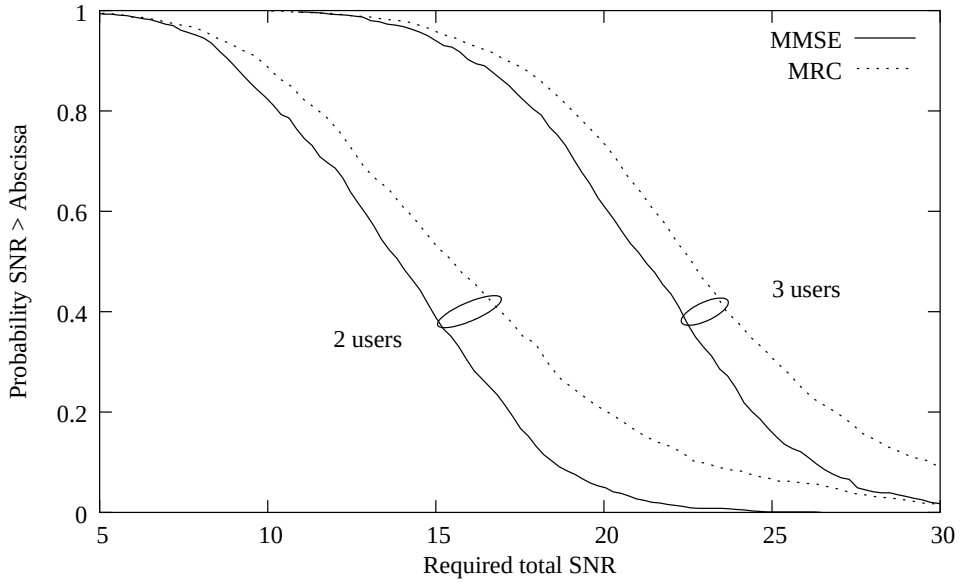

Figure 4.5: Performance comparison of MMSE and MRC receivers for $\{2, 2\} \times 4$ and $\{1, 2, 3\} \times 6$ channels.

performs better in both cases, with an improvement of 1-3 dB in average transmitted power. The larger improvements come in cases with difficult channels that lead to higher power requirements. For some applications, this increase in required power may be acceptable in order to reduce computation at the receiver.

Figure 4.6 shows the number of iterations required for the second stage of the algorithm to converge for both MMSE and MRC receivers. In most cases, assuming an MMSE receiver leads to faster convergence. The relative computational cost per iteration of MMSE and MRC at the transmitter is a function of array size, but these results indicate that while MRC will result in a reduced computational cost at the receiver, the cost at the transmitter can potentially be higher, due to an increase in the number of iterations required for convergence.

4.5.4 Single-Channel Performance

In order to examine the performance of the single-beam algorithm, we use a $\{2, 2, 2, 2\} \times 4$ channel and compare four scenarios. First, the conventional single-beam algorithm was run until convergence for the SVD initialization and for the pseudo-inverse

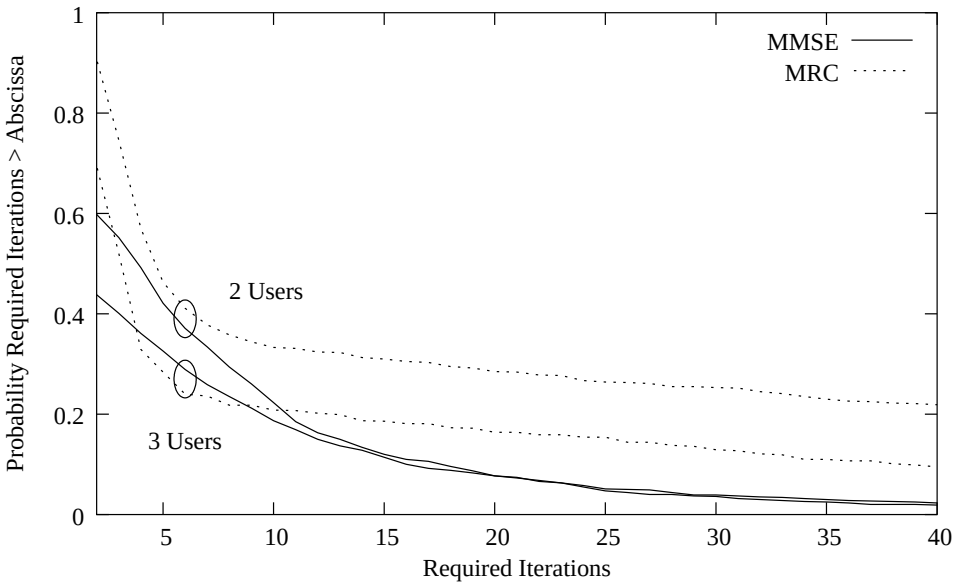

Figure 4.6: Comparison of the required iterations for convergence for MMSE and MRC receivers for $\{2, 2\} \times 4$ and $\{1, 2, 3\} \times 6$ channels.

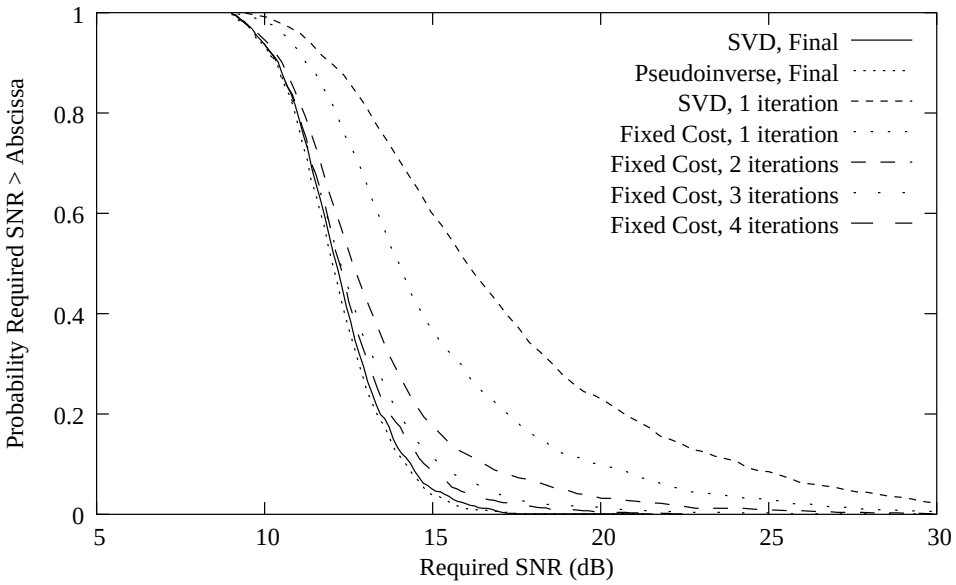

Figure 4.7: Performance of the hybrid algorithm with different initialization methods and fixed numbers of iterations.

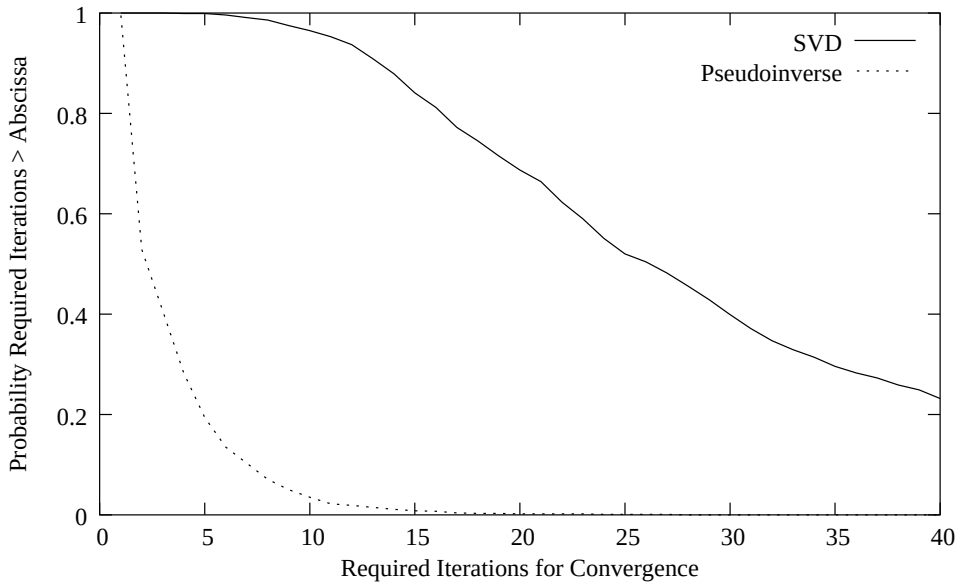

Figure 4.8: A comparison of the number of iterations until convergence for the hybrid algorithm for one data channel per user vs. the SVD initialization

.

initialization (Steps 1 and 2 of the algorithm in Section 4.4.1). We also tested the SVD initialization after a single iteration, and examined the "fixed cost" approach, in which the iterations in Step 4 are omitted, and the number of iterations in Step 2 are fixed. If Step 2 is allowed to converge, it does so typically in around 5 iterations, and in virtually all cases less than 10 iterations. While these iterations require much less computation than those of Step 4, it may still be desirable to keep computational cost fixed by repeating step 2 a predetermined number of times, rather than allowing it to repeat until convergence. Perfect convergence is not necessary since the purpose of Step 2 is only to find an intelligent guess at an initial set of $\mathbf{w}_j$ vectors. Here we fixed the first step at 5 iterations. In Figure 4.7, we compare the fixed cost approach for 1, 2, 3, and 4 iterations with the SVD initialization and the final convergence point of both initializations. The final convergence points are virtually identical, verifying that they result in the same solution. All of the fixed-cost algorithms are substantially better than the SVD approach after 1 iteration, and it appears that 4 iterations are sufficient to reach near-optimal performance almost all the time.

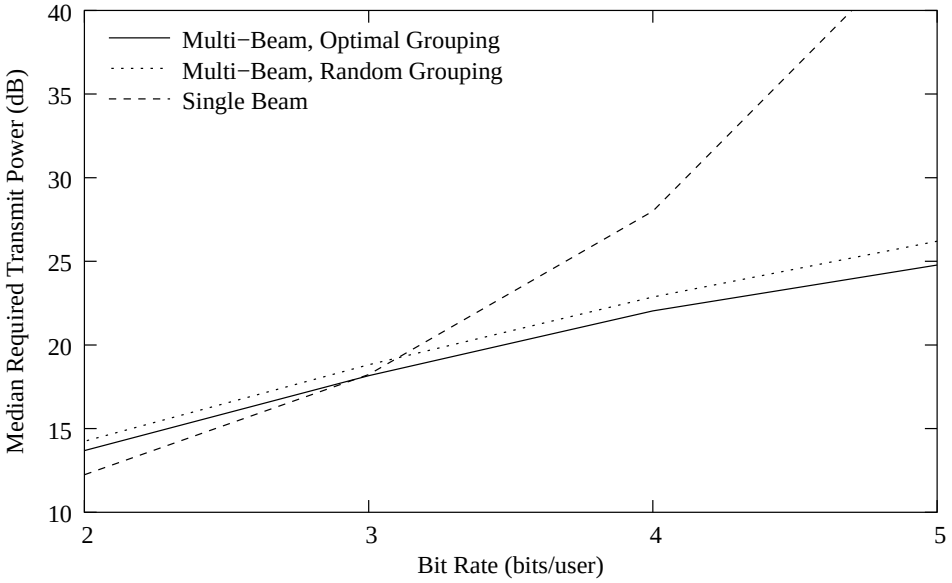

Figure 4.9: A comparison of channel allocation schemes as a function of required transmission rate

The required number of iterations for convergence for the two initialization approaches are compared in Figure 4.8. The cost difference is quite dramatic–even more so than with the general ZF/IB algorithm and multiple sub-channels per user.

4.5.5 Multi-Channel/Single-Channel Comparison

In this section, we compare the multi-channel and single-channel approaches using a $\{2, 2, 2, 2\} \times 4$ channel. The users can all be accommodated simultaneously with $m_j = 1$, or they can be handled as two $\{2, 2\} \times 4$ channels with $m_j = 2$. For the $m_j = 2$ case, three distinct groupings are possible, and the performance of all three were evaluated. Figures 4.9 and 4.10 show results derived by choosing the best of the three, and a second curve representing the performance of one of the groupings chosen at random. The performance metric in this example is the median power required to transmit at the desired rates to all groups of users. The plotted performance numbers are derived by generating 1000 random channels and computing the median required transmit power over all implementations. All channels are assumed to be simple Rayleigh fading channels. Figure 4.9 compares the channel allocation strategies as a function of the transmission rate, which is

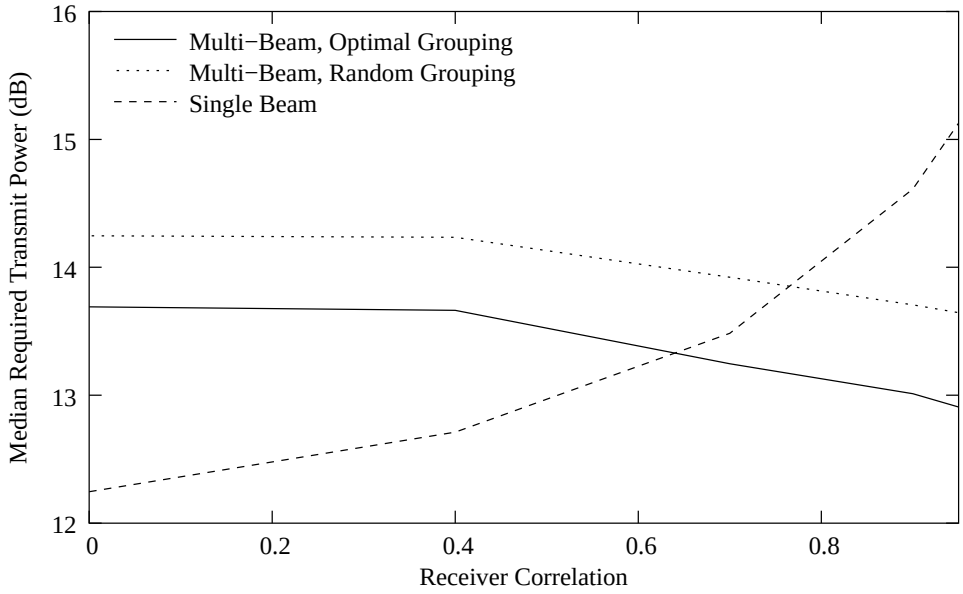

Figure 4.10: A comparison of channel allocation schemes as a function of correlation between receiver antennas.

assumed to be equal for all users. For this figure, all channels are uncorrelated. The performance gap between the optimal and randomly chosen grouping for the multi-channel approach is less than 0.5 dB at low transmission rates, and almost 1 dB at 5 bits per user. The single-channel approach outperforms the multi-channel approach only at the lowest rate of 2 bits/user. Thus, while the single-channel approach is attractive because of its reduced computational cost, the use of multiple sub-channels per user comes at a significantly lower cost in transmitted power as the transmission rate increases.

Figure 4.10 compares the two channel allocation approaches as a function of channel correlation when all users have a transmission rate of 2 bits/user. The $\mathbf{H}_j$ matrices are independent of each other, and the columns of $\mathbf{H}_j$ are independent, while the rows have covariance matrix $\mathbf{R}$, where $E\{[\mathbf{R}]_{i,k}[\mathbf{R}]^*_{j,k}\} = \alpha^{|i-j|}$, and α is the correlation parameter on the horizontal axis of Figure 4.10. In completely uncorrelated channels, the single-beam approach is about 1.5 dB better, while it is over 2 dB worse in highly correlated channels. For large α, it is expected that the gain of the secondary sub-channels would be reduced. However, this example shows that the greater flexibility of using the secondary sub-channels results in better performance as the channel becomes more correlated.

4.6 Conclusion

We have presented a new hybrid algorithm for designing transmit vectors in multi-user MIMO downlink channels. By combining the properties of Generalized Iterative Zero-Forcing with bit-loading, a better initialization point for iterative interference-balancing algorithms is obtained that improves both the performance and the number of iterations required for convergence. We have also generalized the iterative solution to include imperfect channel information with known mean and covariance at the transmitter. The simulation results have revealed that the hybrid approach requires less transmitted power than all other approaches considered here, and is particularly better than distributing the power equally among the sub-channels. In terms of computational cost, the hybrid algorithm does not converge as quickly as interference balancing with an equal power distribution, but more quickly than other schemes that attempt to find a better initial power distribution. The use of MRC rather than MMSE receivers increases the required power slightly, and may increase computational cost at the transmitter, but comes with the benefit of reduced cost of the receiver.

Chapter 5

Spatial Multiplexing Algorithms Applied to Channel Measurements

The performance of all of the algorithms proposed in the last two chapters has been tested mainly on randomly generated channels with Gaussian gains that are uncorrelated from element-to-element of a user's array, and from user-to-user. This assumption, which is frequently used throughout the recent literature on MIMO SDMA algorithms, is an ideal scenario, since the algorithms rely on the users' channels being uncorrelated. It is clear that if two users are located too closely together, or if there is insufficient multipath scattering, these assumptions will be violated. Thus, an important question to address is what propagation characteristics allow one to achieve effective SDMA. In this chapter, we study aspects of this question for realistic indoor environments using data from two sources. The first is measurement data from an indoor MIMO channel sounding experiment [100], and the second is a statistical channel model that has been designed to simulate typical indoor channel conditions [47]. Channel sounding measurements have been used to study the multipath richness of the single-user MIMO channel [101–110], but to date there has not been a similar study for multi-user MIMO channels.

This chapter focuses on two important issues. The first deals with how closely two users can be located in space before a significant reduction in SDMA performance (measured in terms of either capacity, total transmitted power, or SINR) is observed. The second is relevant when users in the network are mobile, and addresses the question of how far a receiver terminal can move before updated CSI is required. The results show that, although the absolute performance of the algorithms tested is somewhat different depending on whether actual measurement data or synthetic models are used, the answers to these questions is not.

Of the many different transmission schemes discussed in the previous chapters, three representative ones are used as test cases. First, Block Diagonalization is used to estimate the channel capacity. While the capacity of the BD algorithm is really only a non-tight lower bound on the actual channel capacity, the relative values for different user locations are adequate for predicting capacity as a function of inter-user and inter-antenna separation. We also investigate a few cases where the dimensions do not support Block Diagonalization, so Generalized Iterative Zero-Forcing is used. In order to compare the performance of transmit vectors resulting in orthogonoal vs. non-orthogonal sub-channels, we also make use of the Hybrid Zero-Forcing/Interference-Balancing algorithm.

In the next section, we describe the experimental system that was used to collect the channel measurements, and Section 5.2 presents a corresponding statistical channel model for purposes of comparson. The results of the study are then presented in Section 5.3, which addresses the issue of inter-user channel correlation, and Section 5.4, which investigates the effects of user mobility.

5.1 Experimental Channel Measurements

A narrowband custom-made MIMO channel probing system designed and built at Brigham Young University (BYU) was used to collect channel measurements. Some details of the measurement system are discussed here, but a more complete description of the system, including diagrams, can be found in [105]. For the measurements used here, the system was equipped with ten monopoles forming a uniform circular array at both transmit and receive. The system was operated at a carrier frequency of 2.43 GHz, and the elements were positioned in a circle with a radius of 0.86 wavelengths so that the separation between adjacent elements was approximately one-half wavelength. The resulting 10×10 MIMO channel was sampled every 2.5 ms with a measurement bandwidth of 25 kHz.

The measurements used in this study were collected on the fourth floor of the Clyde Engineering Building on the BYU campus, which is constructed with steel-reinforced concrete structural walls and cinder-block partition walls. The measurements were taken with the transmitter in a fixed location and the receiver moving along a long corridor as illustrated in Figure 5.1. All channels in this scenario are non-line-of-sight

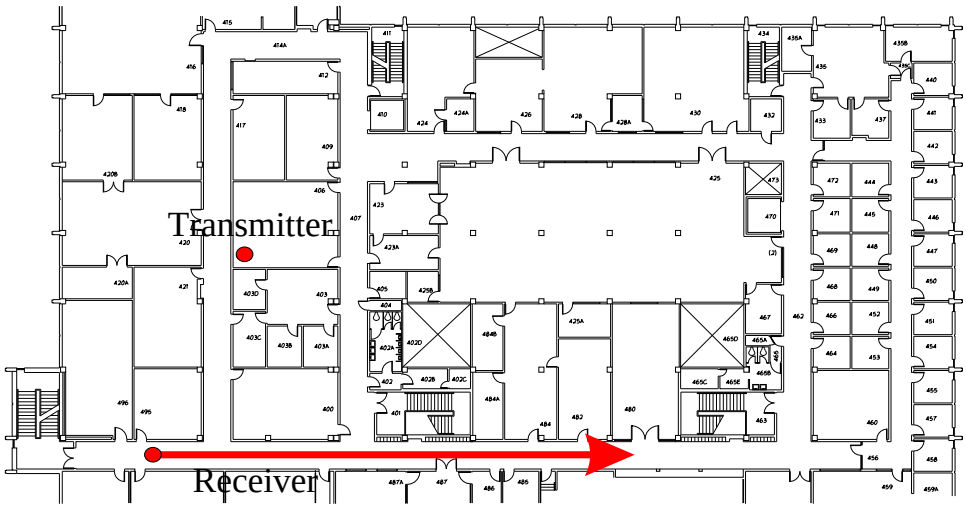

Figure 5.1: Map of the location of the measurement data.

(NLOS), which generally allows for better multipath diversity but with reduced gain compared to the line-of-sight (LOS) case. A total of 10000 samples of the channel were taken over a length of 42.6 meters, which corresponds to 29 samples per wavelength. For a more detailed description of the measurement scenario and the equipment, see [100]. Multi-user scenarios were created by using multiple points along the path as channels for different users. To test the effects of a particular separation distance d, channel measurements at points separated by distance d were compared along the entire length of the measurement set to calculate average and worst-case performance values.

Most of the test cases considered scenarios with fewer antennas than the original 10×10 data set. Appropriate antenna subsets were selected as follows. On the transmit side, antennas with maximal separation were chosen to mimic a fixed basestation that uses the entire 0.86 wavelength array aperture. For example, the solid circles in Figure 5.2 indicate how a subset of four antennas would be chosen from among the ten possible antennas for the transmitter. A mobile receiver, on the other hand, would be expected to have limited size, and thus only adjacent antennas were used in forming the subarrays for the end users.

An important issue that arises in MIMO channel data sets is how the various channels are normalized prior to processing. There are two common approaches. The first is to normalize over the entire data set, so that the ensemble of all of the measured

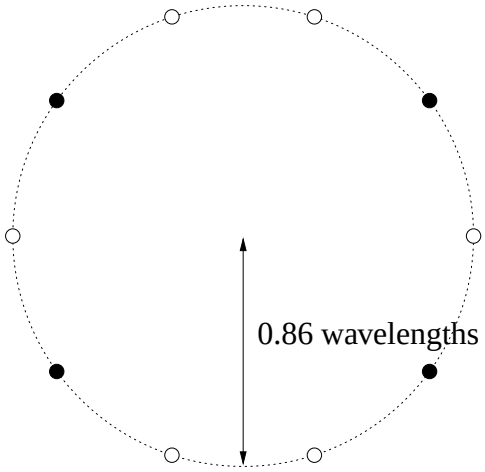

Figure 5.2: Measurement data array structure. The darkened points indicate the location of transmit antennas for test cases where only four antennas are used.

channels has a given average size (usually measured in terms of the Frobenius norm). This approach preserves relative power relationships between different channel samples, but is subject to large fluctuations due to multipath fading and variations in the number of walls between transmitter and receiver. We refer to this normalization approach as Global Channel Normalization (GCN). The second approach is to scale each individual channel sample to have the same Frobenius norm, and is referred to as Local Channel Normalization (LCN). This approach allows for more consistent comparisons with simulated data and makes the results less dependent on the specific physical environment. For the sake of comparison, both methods were used in the results that follow.

5.2 Statistical Model

Channel measurements are of great value in accurately predicting algorithm performance, but obtaining large quantities of measurement data can be prohibitive. When simulating communication systems, it is useful to be able to test them over a very broad range of channels, rather than on a data set from one specific location or group of locations. For this reason, statistical models are useful. Assuming that a model accurately reflects the channel conditions likely to be encountered, it is a relatively simple matter to generate large quantities of channels for simulation purposes. To complement the measured channels

used in testing various SDMA algorithms, we also employ randomly generated data from a realistic statistical channel model described below.

The model used in this paper is the "double bounce" indoor channel model proposed in [47], which is a modified version of the models originally developed in [111] and [46], which is sometimes referred to as the Saleh-Valenzuela with Angle-of-Arrival (SVA) model. When applied to MIMO channels, a good statistical match has been observed with measured capacities in [112] and for a similar model in [102]. The basic properties of the SVA model are as follows:

- All arrivals are members of clusters. The time of arrival of each cluster and the arrivals within each cluster are characterized by a Poisson distribution, with separate delay parameters for the clusters and arrivals within clusters, denoted by Λ and λ, respectively. For example, the PDF of T_j, the arrival time of cluster j, is:

$$f(T_j - T_{j-1}) = \Lambda e^{-\Lambda(T_j - T_{j-1})}.$$

- The amplitudes of all arrivals have a Rayleigh distribution, the mean of which (denoted by σ_{kl} for arrival k in cluster l) is determined by an arrival's position within a cluster and the position of the beginning of the cluster. The mean amplitudes decay exponentially with two different decay constants: Γ for the cluster decay rate, and γ for the arrival decay rate. The mean amplitude is determined by the relation

$$\sigma_{kl}^2 = \sigma_0^2 e^{-T_l/\Gamma - t_{kl}/\gamma},$$

  where σ_0^2 is found from the separation distance between the transmitter and receiver, and t_{kl} is the time delay from the beginning of cluster l to arrival k.

- The mean cluster angles-of-arrival are uniformly distributed over $360°$. The angles-of-arrival of the rays within a cluster are taken from a Laplacian distribution, which is specified by a standard deviation σ.

- In [46], numerical values for the two rate-of-arrival parameters, the two amplitude decay parameters, and the angle-spread parameter were calculated to match measurements from two different buildings, one of which was the Clyde Building.

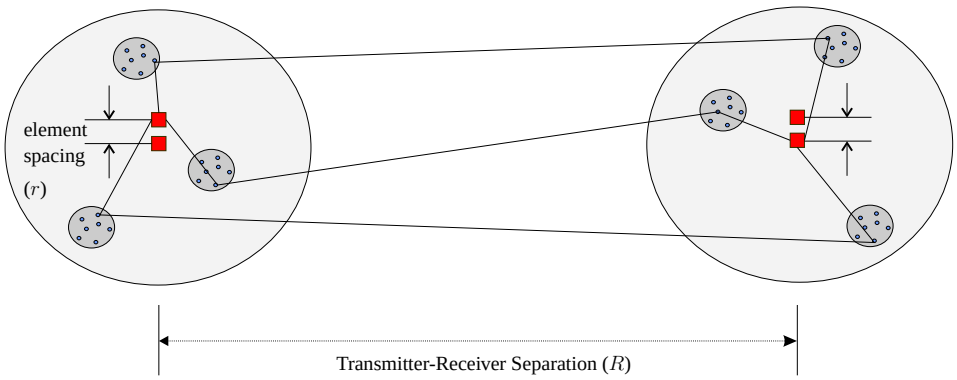

Figure 5.3: An illustration of the double-bounce model with three clusters.

For MIMO channel simulations, it is also necessary to characterize the angle-of-departure at the transmitter. While this quantity was not explicitly measured in [46], it is reasonable to assume that for the indoor environment studied here, the departure angle statistics are similar to and statistically independent of the arrival angle statistics [47]. As shown in [112], this model matches measured MIMO channels well, but it lacks a mechanism for characterizing channel changes when either the transmitter or receiver moves an arbitrary distance. As discussed below, one approach to modeling motion is to base the multipath arrivals on the physical locations of scatterers in the channel, rather than simply time and angle-of-arrival [113, 114].

To incorporate motion in our channel simulations, the double-bounce model of [47] takes a modified version of the statistical model described above and uses it to generate scatterer locations. To do this, a random set of clusters are generated with their associated times, amplitudes, and angles as seen by the transmitter and receiver. It then assumes that each of the waves detected by the receiver are due to two bounces, one local to the transmitter and one local to the receiver, and calculates the location of the scatterers that produce the desired time and angle-of-arrival. This is illustrated in Figure 5.3. Given the location of each scatterer, it is then possible to allow a limited amount of mobility on the part of either the transmitter or receiver or both, which is adequate for the distances considered here. The clusters are distributed uniformly in angle, and the angles of arrival within a cluster have a Gaussian distribution with respect to the angular center of the cluster. Note that while the SVA model originally proposed a Laplacian distribution for this, the

variance of the angular distribution is likely to have a much more noticeable effect on the end results than the actual distribution. In order to simplify the model, the specific arrival times of each ray within a cluster are ignored, and are determined by the propagation distance specified by the angles-of-departure and arrival. This eliminates the need for the λ parameter. Likewise, all members of a cluster are assumed to have equal mean amplitude, eliminating the γ parameter. The original model in [46] assumes that the arrivals and clusters continue to be generated until some noise floor is reached. Here, we simplify things further by fixing the number of clusters and arrivals per cluster in advance.

The double-bounce model of [47] also allows for different types of antennas and polarizations. However, since the measured channel data used for comparison were all collected using monopole antennas, the same assumptions were used in the synthetic data, and the polarization features of the model were not used. Another recently proposed model [114] has some similarities to the double-bounce model, but has been adapted specifically for the case of outdoor cellular channels.

5.2.1 Model Parameters

In this section we briefly discuss the choice of the model parameters that were used in the simulation. Since the measurement data we are using as a reference are narrowband, we conducted only narrowband tests on the simulated channels, so the time of arrival component was ignored in synthesizing the MIMO array responses from the path data. Thus, the main use of the time of arrival parameter Λ is in predicting the total delay spread, which influences the maximum distance at which a scatterer may be located from the transmitter or receiver. We now analyze the relationship between the amplitudes of each cluster with other clusters as a function of the model parameters.

The mean amplitude of the arrivals in a cluster decays exponentially over time, and the actual amplitudes are Rayleigh distributed with respect to the mean. Since the phase is uniformly distributed over the interval $[0, 2\pi]$, each arrival is effectively a Gaussian random variable with zero mean and variance σ_n^2 for the n^{th} cluster. Each σ_{n+1}^2 is related to

the previous σ_n^2 by the same distribution. The relationship between the variance for arrivals n and $n+1$ is

$$\sigma_{n+1}^2 = \sigma_n^2 e^{-\Delta T/\Gamma}\,, \tag{5.1}$$

where $\Delta T = T_{n+1} - T_n$ is the difference in time of arrival between clusters n and $n+1$. The statistical model characterizes this time difference using an exponential distribution with parameter Λ:

$$f(\Delta T) = \Lambda e^{-\Lambda \Delta T}\,. \tag{5.2}$$

This results in the following distribution for σ_{n+1}^2 with respect to σ_n^2:

$$u = \sigma_{n+1}^2 = W(t) = \sigma_n^2 e^{-t/\Gamma} \tag{5.3}$$

$$t = W^{-1}(u) = -\Gamma \ln \frac{u}{\sigma_n^2} = \ln\left(\frac{u}{\sigma_n^2}\right)^{-\Gamma} \tag{5.4}$$

$$|J| = \left|\frac{\partial}{\partial u}[W^{-1}(u)]\right| = \frac{\partial}{\partial u}\left(-\Gamma \ln \frac{u}{\sigma_n^2}\right) = \frac{\Gamma \sigma_n^2}{u} \tag{5.5}$$

$$f(\sigma_{n+1}^2) = f(t = W^{-1}(u))|J| \tag{5.6}$$

$$f(\sigma_{n+1}^2|\sigma_n^2) = \Lambda\Gamma\left(\frac{\sigma_{n+1}^2}{\sigma_n^2}\right)^{\Lambda\Gamma - 1}\,. \tag{5.7}$$

It can be shown that the mean of this distribution is[1]:

$$E(\sigma_{n+1}^2|\sigma_n^2) = \sigma_n^2 \Lambda\Gamma \int_0^1 u^{\Lambda\Gamma} du = \sigma_n^2 \frac{\Lambda\Gamma}{\Lambda\Gamma + 1} u^{\Lambda\Gamma+1}\Big|_0^1 \tag{5.8}$$

$$= \frac{\Lambda\Gamma}{\Lambda\Gamma + 1}\sigma_n^2\,. \tag{5.9}$$

With this information, it is possible to predict the relationships between cluster amplitudes without time-of-arrival information. The original measurement data used to derive the model parameters for the Clyde Building at BYU yielded estimates of $1/\Lambda = 17$ ns and $\Gamma = 34$ ns, resulting in a product $\Lambda\Gamma \approx 2.0$ (note that this product is unitless). The expected amplitude of each cluster with respect to the previous one is thus 0.67.

In order to verify that reasonable values for the model parameters have been chosen, the singular values of the measured and simulated channel matrices were compared. In a channel with several clusters uniformly distributed in angle, there is likely to be

[1]For simulation purposes, this distribution can be generated easily from a uniform distribution. If w is uniformly distributed on $(0,1)$, $u = w^{1/\nu}$ has the probability density $f(u) = \nu u^{\nu-1}$.

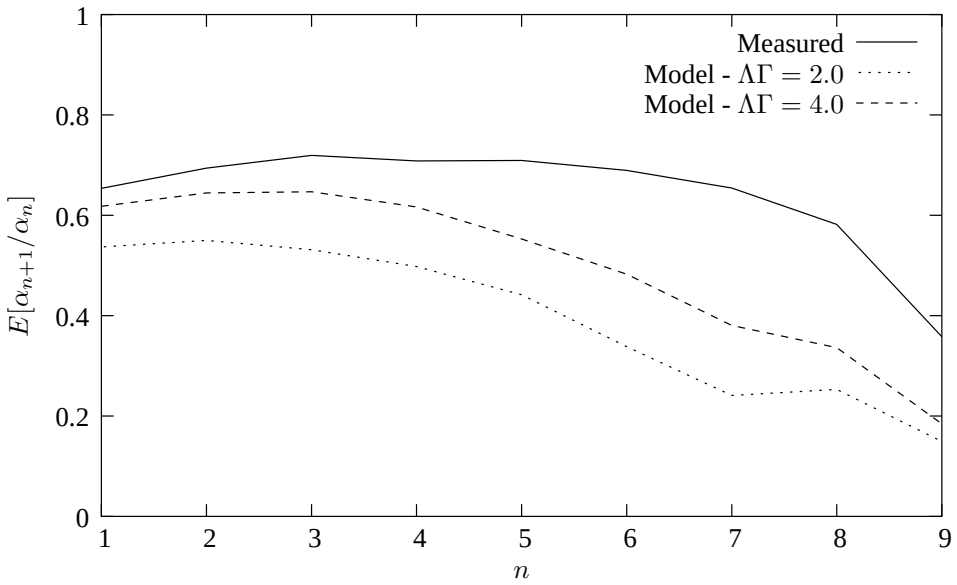

Figure 5.4: A comparison of the mean singular values of the **H** matrices generated from measurements and models.

some correlation between the amplitude spread of the clusters and the eigenvalue spread of the associated **H** matrix. To compare the properties of the measured and synthetic **H** matrices, we calculated the singular values of each measured **H** matrix, and a corresponding number of model-generated **H** matrices. Let α_n represent the n^{th} singular value of **H**. We calculated α_n/α_{n+1} for $n = 1 \ldots n_T - 1$, and plotted the mean values of these ratios in Figure 5.4. The figure compares the case of $\Lambda\Gamma = 2.0$, which was chosen based on previous measurements at higher frequencies, to a higher value of $\Lambda\Gamma = 4.0$, which reduces the decay of secondary multipath components and more closely matches the singular values of the measured matrices. The larger discrepancy between the curves for the smaller singular values can be attributed to the fact that the statistical model assumed a limited number of clusters, and to the presence of some limited diffuse scattering in the measured data. This is not a significant issue since small subarrays are used in the simulation results, and only the first few singular values are important.

Table 5.1 lists all of the parameters that were used in the simulations presented in the next two sections.

Table 5.1: Model parameters used to generate synthetic channels.

Parameter	Value
Frequency	2.43 GHz
Cluster Rate of Arrival ($1/\Lambda$)	17 ns
Cluster Decay Constant (Γ)	100 ns
Number of Clusters	7
Scatterers per Cluster	7
Angular Azimuth Spread in Clusters (σ)	25°
Tx-Rx Separation Distance (R)	10 m
Channel sample spacing	0.05 wavelengths
Total Samples per Channel Realization	200
Number of Channel Samples	100

5.3 Effects of Inter-User Separation

In the following, the notation $\{n_{R1}, n_{R2}\} \times n_T$ is used to indicate a scenario involving two receivers with n_{R1} and n_{R2} antennas, and a transmitter with n_T antennas. In studying the effects of inter-user separation on SDMA performance, we tested two channel geometries: $\{2, 2\} \times 4$ and $\{4, 4\} \times 10$. Figure 5.5 shows the results of implementing power minimization via Block-Diagonalization (BD) for both channel geometries with the two users separated by distances of 0.5, 1.5, and 5 wavelengths. The channels in the data set were scaled using GCN in this example. The power allocation was computed such that a total of 3 bits/symbol could be transmitted to each user at a symbol error rate of 10^{-5} while minimizing transmit power, and the data in the plot is the CCDF of the total transmitted power. The power fluctuations in the data set are evident in the irregular shape of the CCDFs. Due to the higher received power at certain locations along the measurement track, the measurement-based channel occasionally outperforms an IID channel. Results for local normalization (LCN) are shown in Figure 5.6. In this case, the measured channels never outperform the IID channel. As expected, capacity increases with user separation. In the case of the $\{2, 2\} \times 4$ channel, the capacities of the measured data approach the IID capacity, while in the case of the $\{4, 4\} \times 10$ channel, they do not. This implies that the number of usable multipath components of the channel is less than four. This is not

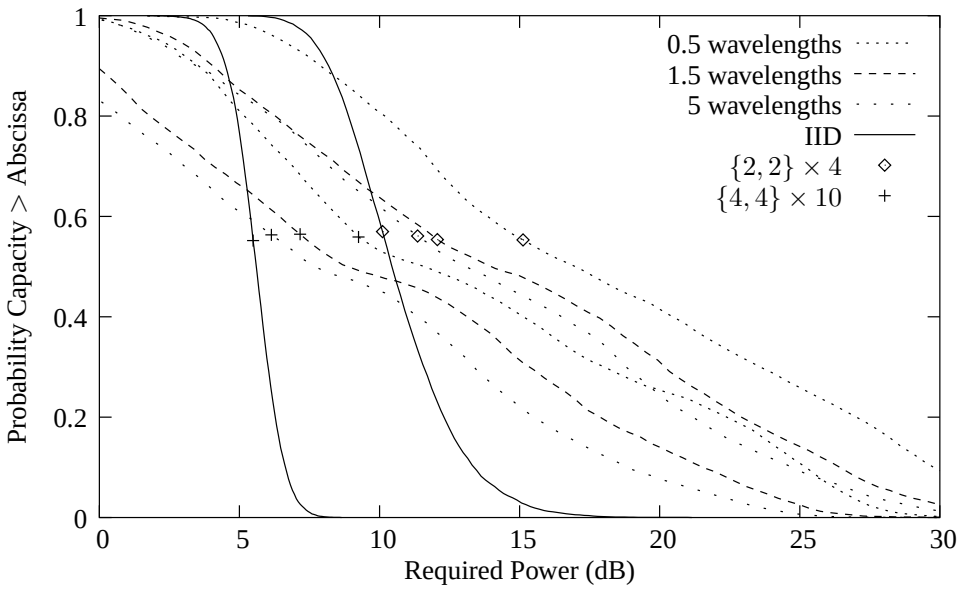

Figure 5.5: CCDFs of required power as a function of separation distance for a two-user MU-MIMO system using channel measurements with global normalization.

unexpected, since even in cases where the channel has full rank, the amplitude decay of the secondary multipath components will typically make the capacity somewhat less than that of IID channels.

Figure 5.7 illustrates the results for both sum capacity at 10 dB SNR and the required transmitted power for 3 bits/symbol per user at a symbol error rate of 10^{-5} as a function of separation distance from 0.5 to 10 wavelengths. The capacity numbers are generated using LCN and represent a 10% outage probability. The transmit power results are based on GCN and represent the median of all realizations in the data set. While it is evident from Figure 5.6 that the $\{4,4\} \times 10$ channel does not quite reach the IID channel capacity, it is clear that in both scenarios, 5 wavelengths of separation yields the maximum possible channel decorrelation. This 5 wavelength lower bound was also observed in other simulations not shown here involving a 10 element transmitter with 10 single-antenna users. At the carrier frequency of 2.43 GHz this, represents a distance of approximately 60 cm. This is a promising result since it is unlikely for two mobile users to operate wireless devices at distances much closer than this.

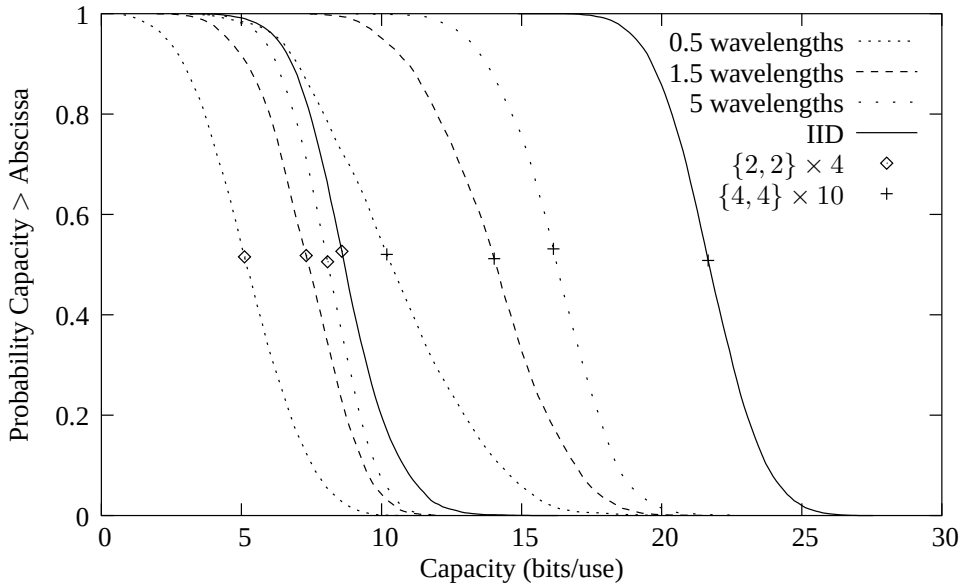

Figure 5.6: CCDFs of capacity as a function of separation distance for a two-user MU-MIMO system using channel measurements with local normalization.

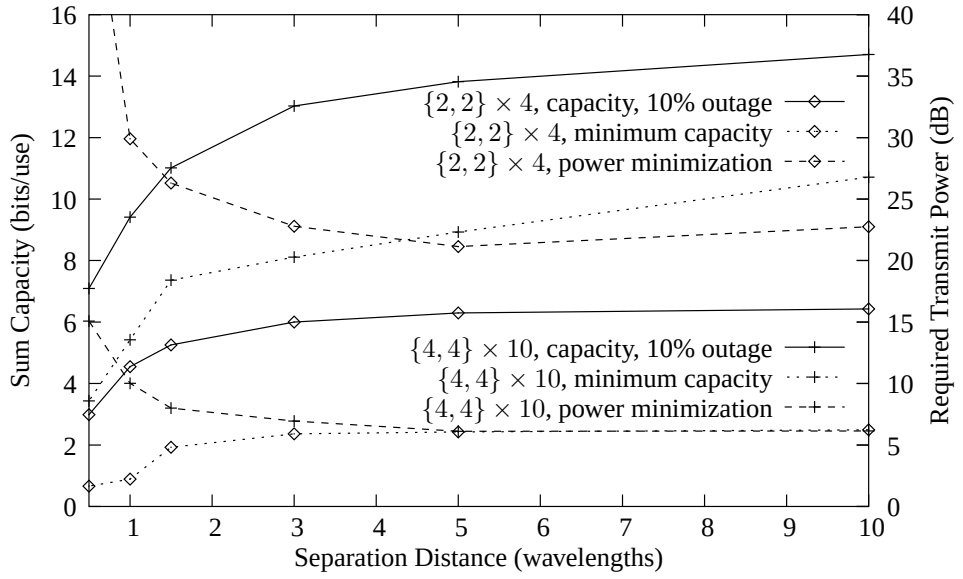

Figure 5.7: Performance as a function of separation distance using data from Figures 5.5 and 5.6.

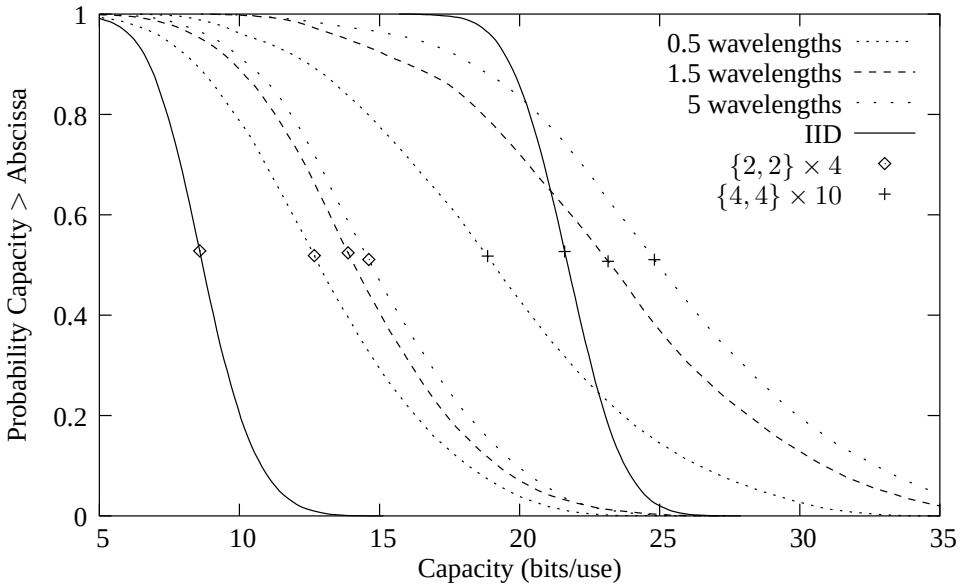

Figure 5.8: CCDFs of capacity as a function of separation distance for a two-user MU-MIMO system using statistical channel model data with local normalization.

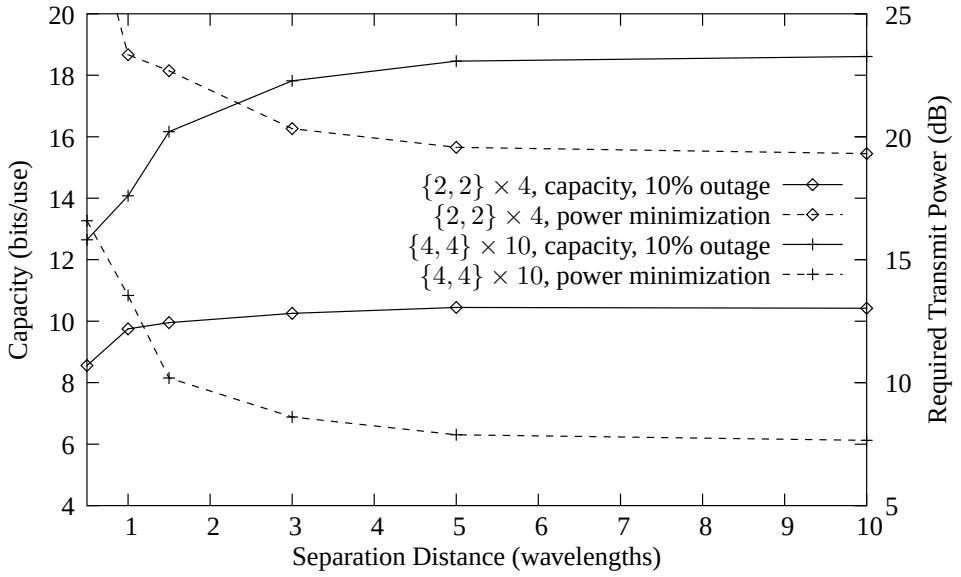

Figure 5.9: Capacity as a function of separation distance for a two-user MU-MIMO system using statistical channel model data.

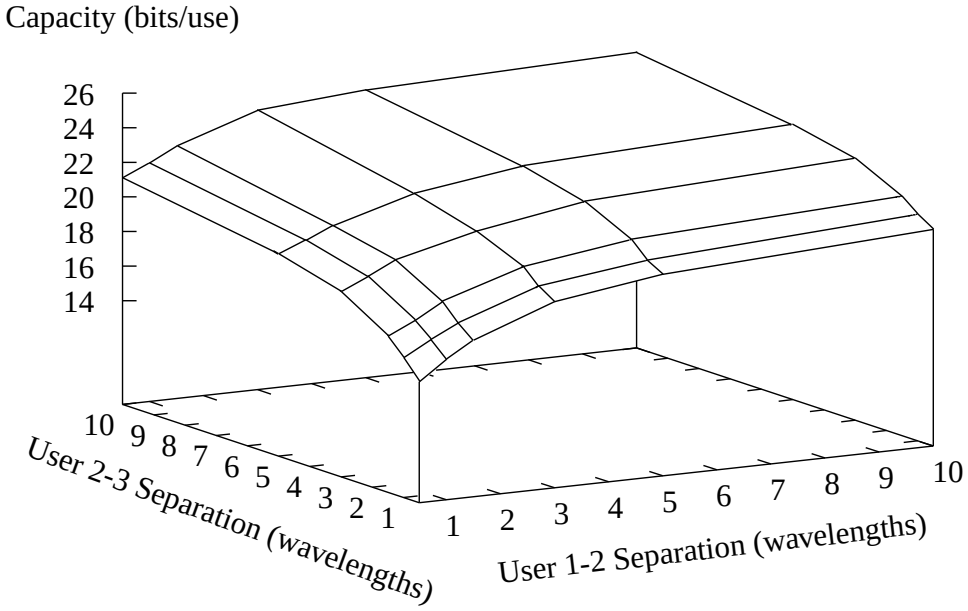

Figure 5.10: Capacity as a function of separation distance for a three-user $\{10, 10, 10\} \times 10$ MU-MIMO system using channel measurements with local normalization.

Figures 5.8 and 5.9 are the same as Figures 5.6 and 5.7, except that they are derived from synthetic channels based on the statistical model described in the previous section. While the maximum capacity available from the synthetic channels is somewhat less than from the measurements, it also exhibits the same trend of near-maximum performance at a user separation of about 5 wavelengths.

Figures 5.10 and 5.11 examine the performance of the coordinated zero-forcing algorithm. In Figure 5.10, we examined the performance of 3-user $\{10, 10, 10\} \times 10$ channels. Similarly to the 2-user cases studied here, we generated test cases by moving the three users along the entire length of the data set at fixed separation distances. Since there are two separation parameters in this case, we tested all possible combinations of the two separation distances, resulting in the 3-dimensional plot of Figure 5.10. In all cases, m_j was fixed at 3 sub-channels for all users. The results show the same trend as the 2-user channels, although there is a slightly larger increase in capacity when the minimum user separation is increased from 5 to 10 wavelengths.

Figure 5.11 illustrates the capacity of a $\{5, 5\} \times 10$ channel when coordinated zero-forcing is used, as a function of both separation distance and the number of data

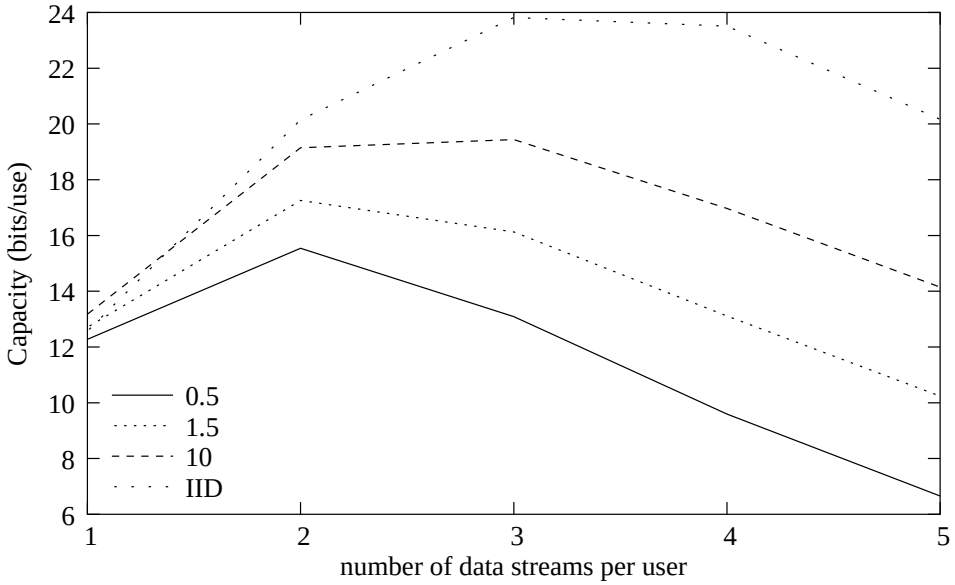

Figure 5.11: Capacity as a function of separation distance and the number of data streams in use for a $\{5, 5\} \times 10$ MU-MIMO system using channel measurements with local normalization.

streams (m_j) allocated to the two users. The IID case is also included as a reference. For these dimensions, choosing $m_j = 5$ results in the coordinated zero-forcing and block-diagonalization solutions being equivalent. Interestingly, both the measurement-based and IID generated channels benefit from reducing the number of data streams per user to 3. This can be attributed to the interaction between the two users' channels. For example, using only three sub-channels for one user allows more degrees of freedom for the other user to optimize its spatial channel allocation. The performance at a separation of 10 wavelengths in comparison to the IID channels suggests that there are 2-3 dominant paths that contribute most of the capacity of the channel. It is also interesting to note that for only one sub-channel per user, the performance varies very little from 0.5 to 10 wavelengths of separation, due to the fact that there is sufficient multipath richness to provide adequate performance for each user even when the two arrays are virtually superimposed. This consistent performance may make setting $m_j = 1$ for all users all the time an attractive design choice for some systems because this special case also provides reductions in computational complexity [42].

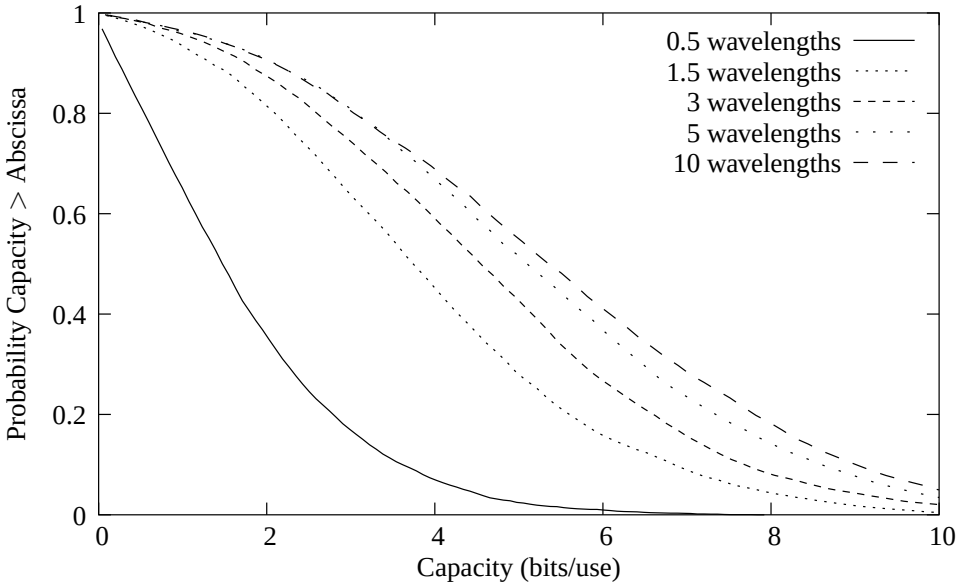

Figure 5.12: Capacity of a 10-user, single-antenna channel with 10-element base station, derived from channel measurements.

Figure 5.12 illustrates the capacity as a function of separation distance for a 10-element transmitter with 10 users all having only one antenna.

5.4 Effects of User Motion

The second application of the channel data was to measure the effects of channel latency due to user motion. The envisioned scenario in this case is that the transmitter is obtaining its estimates of the channel via feedback from the mobile users. By the time the transmitter has received and processed a channel estimate, the receiver may have moved slightly, causing a change in CSI. A similar scenario is considered in [115]. The error due to user motion is in addition to estimation error that is already present due to additive noise in the signal received by the mobile from the base. Minimizing CSI error is critical in multi-user scenarios since in accurate channel estimates reduce the effectiveness of interference reduction in SDMA.

The effect of CSI error due to user motion was quantified by measuring the mean SINR degradation of the sub-channels. Let $\mathbf{H}_S$ represent the true channel matrix, and $\hat{\mathbf{H}}_S$ be the information available to the transmitter that is corrupted by user motion and

estimation errors. If $\mathbf{m}_{i,j}$ is the transmit beamformer for sub-channel j of user i, and $\mathbf{w}_{i,j}$ is the corresponding receiver beamformer, then the SINR for the sub-channel at the output of the receiver is

$$\text{SINR}_{i,j} = \frac{\mathbf{m}_{i,j}^* \mathbf{H}_i^* \mathbf{w}_{i,j} \mathbf{w}_{i,j}^* \mathbf{H}_i \mathbf{m}_{i,j}}{\sum_{i \neq k, j \neq l} \mathbf{m}_{k,l}^* \mathbf{H}_i^* \mathbf{w}_{i,j} \mathbf{w}_{i,j}^* \mathbf{H}_i \mathbf{m}_{k,l} + \sigma^2}, \tag{5.10}$$

where σ^2 is the noise power. Since $\mathbf{m}_{i,j}$ is designed based on corrupted channel information, the true SINR will be different than what was intended. This SINR degradation is measured by comparing $\text{SINR}_{i,j}$ computed using the true channel $\mathbf{H}_S$ to the SINR resulting from the available channel information $\hat{\mathbf{H}}_S$. The corrupted CSI was generated by taking a given measured or synthetic channel at one location, displacing it by a distance d to obtain $\mathbf{H}_{S,d}$, and then adding an estimation error term $\mathbf{N}$ composed of uncorrelated Gaussian elements: $\hat{\mathbf{H}}_S = \mathbf{H}_{S,d} + \mathbf{N}$. The relative size of these quantities is specified using the "signal to estimation error" ratio $\|\mathbf{H}_{S,d}\|_F^2 / \|\mathbf{N}\|_F^2$. A total of 1000 test cases were generated by randomly selecting locations for the two users in the data set. All channels were of dimension $\{2, 2\} \times 4$. Since the locations are selected randomly, it is very unlikely that the two users ever had a very small separation distance, so their channels can be considered to be statistically independent. GCN was used to include the effects of power fluctuations at different points in the measurement set.

For each realization, the channel was decomposed into orthogonal sub-channels using both block-diagonalization and the hybrid algorithm. Each user was moved in a randomly chosen direction and the resulting SINR degradation was averaged over 10 samples. Note that any direction of motion was possible with the statistical model, while motion was restricted to a straight-line path in the measured data. Figure 5.13 illustrates the SINR loss as a function of the estimation error and the distance between where the channel was measured and where it was used. Each point on the curves represents the median SINR loss over 1000 trials. The median is used rather than the mean in this case since it is not altered by the conversion into log space. Data for both orthogonal multiplexing (BD algorithm) and non-orthogonal multiplexing (hybrid algorithm) are included. The three curves for each case are for relative estimation error powers of 5, 10, and 15 dB below the power

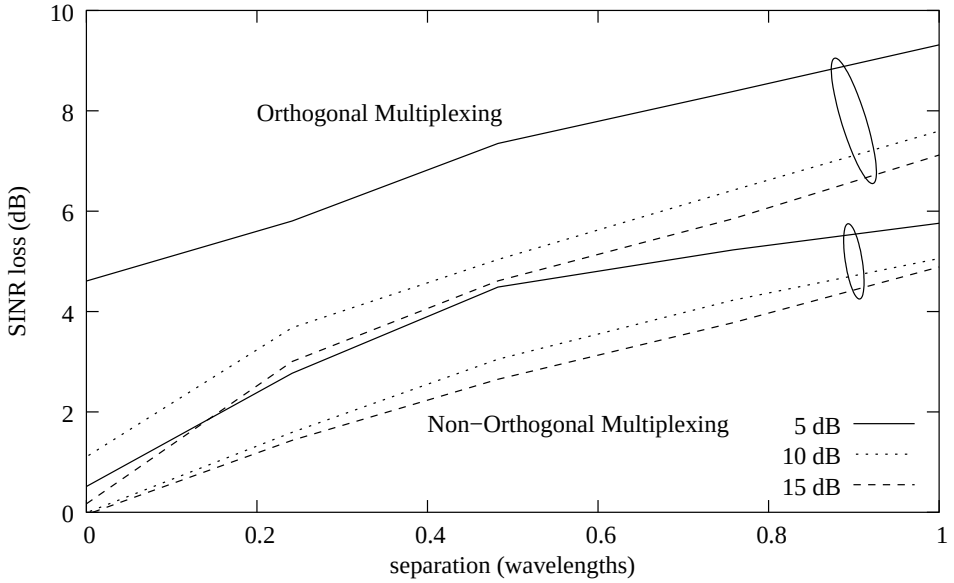

Figure 5.13: Median sub-channel SINR loss as a function of channel estimation error from channel measurements.

of $\mathbf{H}_{S,D}$. Above a 15 dB, there is no noticeable improvement in performance, and below 5 dB, the performance degradation increases very quickly, and so is not included here.

Figure 5.14 shows a plot similar to Figure 5.13, but based on synthetic data where the two users' channels were generated independently of each other. The plot generally confirms the findings of the previous one, except that the performance reduction as the separation increases toward 1 wavelength is not quite as great, and the degradation is non-monotonic. One possible explanation of this behavior is correlation in the fading characteristics. Channels with a higher degree of spatial correlation will have more accurate channel estimates than channels with a lower degree of spatial correlation at the same measurement distance. Thus, a local maximum in the channel degradation would correspond to a local minimum in correlation. In Figure 5.14, the local maxima approximately correspond to the nulls in correlation that result from Jakes' correlation model, which results from the structure imposed by the channel model.

Figures 5.13 and 5.14 also provide useful insights into the trade-offs of choosing orthogonal versus non-orthogonal multiplexing algorithms. In general, the non-orthogonal algorithms are capable of achieving better performance because they consider a larger set

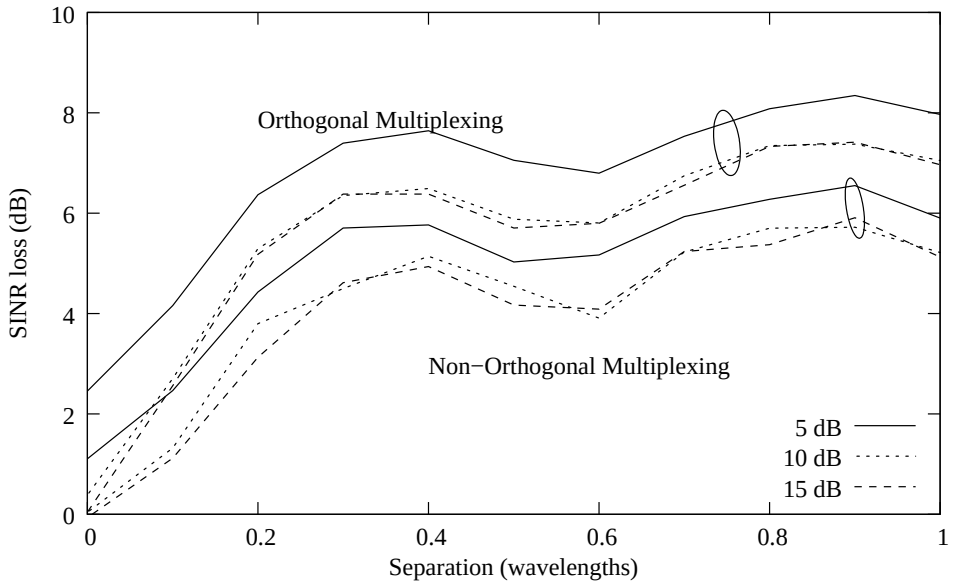

Figure 5.14: Median sub-channel SINR loss as a function of channel estimation error from statistical channel model data.

of possible solutions. Because we are considering only degradation rather than raw performance, this is not apparent from these plots, but it has been demonstrated elsewhere [42]. However, these plots demonstrate that non-orthogonal algorithms, in addition to having better overall performance, are more robust to channel estimation error. This is a consequence of the fact that they are already designed to tolerate a certain amount of inter-user interference, while the orthogonal methods are not.

Assuming an SINR loss of 3 dB due to channel mismatch can be tolerated, the above results show that motion on the order of one-half wavelength will invalidate the original channel estimate. At 2.43 GHz, this corresponds to a distance of only 6 cm. Assuming that the maximum speed of a mobile user indoors is around 5 m/s (walking speed), the required channel update rate would be on the order of 10 ms. Recent results [116] indicate that this rate could be significantly reduced through the use of MIMO channel prediction.

The synthetic and measured data considered in this paper have been strictly for the indoor environment, which has been observed to have somewhat more favorable characteristics for MIMO transmission than outdoor channels [117]. Users in outdoor environments will have higher mobility, but there is typically less multipath scattering and

also the possibility of LOS propagation, so a user may be able to move a greater distance before a significant change in the channel occurs. Thus, it is possible that the algorithms tested on the indoor data presented here may still be useful in outdoor channels, but further tests are necessary to validate such a claim.

5.5 Conclusion

Algorithms for spatial multiplexing in multi-user MIMO systems can substantially increase the capacity of a wireless network, assuming that accurate channel state information is available, and that the channels for different users are uncorrelated. This chapter has presented performance results for several multi-user MIMO downlink algorithms when tested against measured and synthetic indoor channels to determine their sensitivity to user motion and correlation among the users' channels. The analysis indicates that most of the potential throughput can be achieved at user separations as little as 5 wavelengths, and that even at shorter distances the performance can still be quite acceptable. This is a very promising result for future multi-user MIMO systems. Measurement of the effects of channel estimation error reveals that motion on the order of 0.5 wavelengths between the channel measurement and the use of the measurement by the transmitter can provide acceptable performance if the expected losses are built into the design requirements.

We have also compared orthogonal multiplexing and a more computationally expensive non-orthogonal multiplexing algorithm in the presence of channel estimation error. The non-orthogonal algorithm, which has previously been shown to require less power to achieve the same data rates as the orthogonal algorithm, was observed here to also be more robust to errors in the channel information. This robustness comes from the fact that the transmit vectors are already designed to tolerate a certain amount of inter-user interference. Results in previous chapters have demonstrated the trade-off between cost and performance for these two algorithms, and the results here show a similar trade-off in robustness to channel error.

Chapter 6

Channel Allocation Strategies

In previous chapters we have explored some of the various algorithms that have been proposed for multiplexing multiple users when the base and mobile stations all have multiple antennas. Transmission schemes based on linear pre-processing at the transmitter, such as those presented here, require to at least some degree linear independence between users' channels. While transmission schemes based on dirty-paper coding do not rely so heavily on linear independence, they still suffer from diminished capacity when there is heavy correlation between users. This is illustrated by the fact that even a single user MIMO channel, which represents the upper bound on sum capacity for all configurations, has reduced capacity when there is high correlation in the channel matrix. If two users' channels are not completely statistically correlated, they will never be completely linearly dependent, but even a moderate degree of correlation will at best reduce the system capacity and at worst make the problem numerically ill-conditioned. The last chapter was devoted to investigating how much separation must exist between users to assure adequate decorrelation. The results were quite promising, but there are other propagation environments (the outdoor environment, for example), where it is much more likely that two users have a highly correlated channel. In this situation, the only solution is intelligent channel allocation that avoids attempting to spatially multiplex two or more users whose channels are highly correlated. In this chapter we consider the problem of channel allocation in the context of MIMO systems that perform spatial multiplexing. Channel allocation for multi-user mobile wireless networks is a well-studied problem, but MIMO SDMA adds a new dimension to the problem and will require new allocation methods. There has been some research on channel allocation algorithms that take into account the use of adaptive

arrays at the base station, but these have assumed that no array processing is used at the mobile devices. Allocation for MIMO downlinks when the users have arrays has recently been considered in [118], but in that case the base station used antenna selection diversity to optimize performance, without any beamforming. Here we consider allocation when the base station uses multi-user downlink beamforming. We begin by reviewing the general allocation problem and past work in this area, and then propose an algorithm for allocation in multi-user MIMO systems.

6.0.1 The Channel Allocation Problem

Traditionally, the channel allocation problem arises in the context of cellular systems, where the available bandwidth is divided into some set of channels, whether they exist in time domain (TDMA), frequency domain (FDMA), or code domain (CDMA), or some combination of these. The pool of available channels must be allocated to meet the following goals:

- Avoid interference between cells. For example, using the same channel in two adjacent cells can result in a user receiving in-band interference from the other cell.
- Minimize "call blocking", where a user requests a channel and there are none available.
- Maximize total network capacity.

The many algorithms that have been proposed for channel allocation can generally be categorized into two categories. The Fixed Channel Allocation (FCA) approach permanently assigns a set of channels to each cell. When all channels are in use, new calls are blocked. Dynamic Channel Allocation (DCA) keeps all channels in a central pool and allows them to be allocated to particular cells as needed. This allows greater flexibility in dealing with fluctuations in the concentrations of users. Generally, FCA algorithms have been shown to be less efficient in the total traffic they can handle, but they tend to have a lower blocking probability under heavy loads. Since each of these approaches is more advantageous in certain situations, there are many hybrid schemes that have been proposed

which try to combine the desirable attributes of both approaches. A survey of the various FCA, DCA, and hybrid schemes is found in [119].

Another way in which the channel allocation algorithms can be categorized is centralized vs. distributed. In centralized allocation schemes, whether fixed or dynamic, the allocation is performed centrally for an entire network. In distributed algorithms, the implementation is spread across the network, so that each base station makes individual channel allocation decisions, based on some sort of algorithm that hopefully results in intelligent channel use across a network [120]. The motivation for this is generally to minimize the cost of computing allocation information at a single location and communicating with each of the member nodes.

The use of antenna arrays in base stations adds additional complexity to the allocation problem. Array processing gives the base station the capability of directional transmission and reception, using either a fixed-beam or an adaptive approach. There are several examples in the literature of channel allocation that assumes the use of adaptive arrays at the base station. One approach to this problem is to view the array processing only as a means of reducing inter-cell interference [121–125]. Alternatively, using the right processing methods can enable the base station to share a single channel with multiple users, thus improving the total network capacity [126–128]. Here we consider the problem of channel allocation for the multi-user MIMO downlink, where the transmitter is using algorithms such as those discussed in previous chapters to communicate with users who also have arrays. This significantly increases the complexity of the problem, so for simplicity we ignore larger network issues here and consider a system only a single base station. The uplink channel was recently considered in [129], where allocation is accomplished by intelligent scheduling in a TDMA system with time variation in the transfer functions from each of the users, similar to the schemes discussed in [130, 131].

6.0.2 Problem Definition

We consider the problem of spatial channel allocation for a single base station when the number of users is greater than the number that can be supported simultaneously on a single channel. For a system with K users and n_T transmitters, n_T is the upper bound

on the number of users that can be simultaneously supported. When $K > n_T$, there are more users than be supported by only SDMA, so we assume that SMDA is used in conjunction with other multiple access methods (i.e. TDMA, FDMA, CDMA) to increase system capacity. We assume for simplicity that the other multiple access methods do not cause inter-user interference for users that are not sharing a channel. This includes TDMA, and orthogonal FDMA and CDMA methods. Since all of these directly or indirectly separate users by temporal processing, we will use the term "time-domain channels" to refer to the different time, frequency, or code slots available to the base station transmitter.

Given these conditions, the problem we wish to solve is to minimize total transmitted power by choosing which users should share time-timain channels. In the following section, we propose a solution, referred to as the "Compatibility Optimization" algorithm (COA). Then, Section 6.2 discusses the application of the COA in situations where multiple sub-channels are being transmitted to a particular user, and simulation results follow in Section 6.3.

6.1 Compatibility Optimization Algorithm

Let K be the total number of users in a system, and G the number of groups, which will correspond to the number of time or frequency slots available for multiplexing. The groups are represented as $\mathcal{G}_1, \mathcal{G}_2 \ldots \mathcal{G}_G$. The number of users in $\mathcal{G}_j$ is q_j. In order to reduce the complexity of the problem, we impose the restriction that $q_{\text{min}} \leq q_j \leq q_{\text{max}}$, where the bounds are defined as:

$$q_{\text{max}} = \left\lceil \frac{K}{G} \right\rceil \tag{6.1}$$

$$q_{\text{min}} = \left\lfloor \frac{K}{G} \right\rfloor . \tag{6.2}$$

When K/G is an integer, $q_{\text{max}} = q_{\text{min}}$; otherwise $q_{\text{min}} = q_{\text{max}} - 1$. While it is theoretically possible for the globally optimal allocation to have minimum and maximum group sizes that are outside these bounds, the probability of this is likely quite low for most cases, and the size restriction allows us to manage the complexity of the problem.

Under these constraints, there will be

$$G_{\text{max}} = G(1 - q_{\text{max}}) + K \tag{6.3}$$

groups containing q_{max} users, and

$$G_{\text{min}} = G - G_{\text{max}} \tag{6.4}$$

groups containing q_{min} users. Given these parameters, there exists a total of

$$\frac{K!}{G_{\text{max}}! G_{\text{min}}! \prod_{i=1}^{G} q_i!}$$

possible unique groupings. For example, if $K = 4$, and $G = 2$, then $q_1 = q_2 = 2$, and there are a total of 3 unique permutations.

Let $P_{\mathcal{G}_j}$ be the required transmitted power for the users in $\mathcal{G}_j$. The globally optimal grouping is the one that minimizes

$$\sum_{j=1}^{G} P_{\mathcal{G}_j}.$$

The problem is that $P_{\mathcal{G}_j}$ is a function of both the chosen beamforming strategy and the channels of all users in $\mathcal{G}_j$. An individual user's contribution to the required power for any group is a function of its channel's interaction with other users' channels. As a result, the only way to minimize the total transmitted power is to compute $P_{\mathcal{G}_j}$ for all possible groupings. For simple cases such as the example above which has 3 unique groupings, it may be realistic to do this, but for any larger number of users, the cost quickly grows too large.

As an alternative, we propose a two-step process for assigning users to groups. The first step is to compute some metric $\xi_{i,j}$ for each pair of users which characterizes the cost of putting users i and j together in the same group. For K users, $K(K-1)/2$ metric calculations are required. After computing $\xi_{i,j}$ for all possible combinations, a Compatibility Optimization Algorithm (COA) is used to find the grouping is that optimizes

$$\sum_{g=1}^{G} \left(\sum_{i,j \in \mathcal{G}_g} \xi_{i,j} \right). \tag{6.5}$$

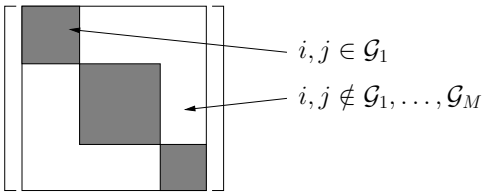

Figure 6.1: An illustration of the minimum-sum grouping algorithm.

The idea of a "compatibility metric" has been proposed previously for systems where all users have one antenna. One proposal is a normalized inner product of two users' channels [132]. This is attractive because it requires minimal computation, but it does not simplify to a single constant when the users have multiple antennas. In [126], where the compatibility metric is referred to as a "spatial compatibility check", several different metrics are proposed. The best performing of these is a normalized sum of the interference and noise transmitted to a user. Since computing this as a single scalar also presents problems for multiple-antenna users, a alternative would be to use another metric the authors proposed: the minimum angle between the subspaces spanned by $\mathbf{H}_i$ and $\mathbf{H}_j$. However, this did not perform as well in empirical tests as the alternative we propose here: a scaled Frobenius norm.

$$\xi_{i,j} = \frac{\|\mathbf{H}_i \mathbf{H}_j^*\|_F^2}{n_{Ri} n_{Rj}}.$$

This compatibility metric is effectively an estimate of the total correlation between the two users' channels. The number is scaled by the channel dimensions so that the metric is not unfairly biased against putting users together that have large numbers of receivers, when in reality the additional receivers are an asset that increases SINR when using joint beamforming algorithms.

When the pairwise compatibility metrics have been computed for all cases, it is necessary to find the grouping that minimizes the sum in (6.5). The problem can be visualized as a matrix $\mathbf{G}$ whose elements $[\mathbf{G}]_{i,j} = \xi_{i,j}$, where $\xi_{i,i} = 0$. The matrix $\mathbf{G}$ is illustrated in Figure 6.1. The sum of the shaded areas is equivalent to twice the sum of (6.5).

The problem can be characterized in mathematical terms by defining a set of matrices $\mathbf{F}_1 \ldots \mathbf{F}$, where $\mathbf{F}_j$ is a square matrix of dimension q_j (the size of group G_j), consisting of all ones. Define the block-diagonal matrix:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & & & \mathbf{0} \\ & \mathbf{F}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{F}_G \end{bmatrix}. \tag{6.6}$$

Let $\mathbf{G}_0$ be the matrix of compatibility metrics in arbitrary order. Then the optimal grouping can be characterized as

$$\mathbf{P}_{\text{opt}} = \arg \min_{\mathbf{P} \in \mathbf{P}^{(K)}} \mathbf{1}^T \mathbf{P} (\mathbf{G} \odot \mathbf{F}) \mathbf{P}^T \mathbf{1} \tag{6.7}$$

$$\mathbf{G}_{\text{opt}} = \mathbf{P}_{\text{opt}} \mathbf{G}_0 \mathbf{P}_{\text{opt}}^T, \tag{6.8}$$

where $\mathbf{P}^{(K)}$ represents the set of all permutation matrices of dimension $K \times K$ and $\mathbf{1}$ represents a vector consisting of all ones, also of dimension K. An algorithm for finding $\mathbf{P}_{\text{opt}}$ is outlined below.

Compatibility Optimization Algorithm

1. Initialization

   (i) Given G and K, compute q_{max}, q_{min}, G_{max}, and G_{min}.

   (ii) Generate $\mathbf{F}$ for the channel dimensions.

   (iii) Determine an arbitrary $\mathbf{G}_0$ and compute $s_0 = \mathbf{1}^T (\mathbf{G}_0 \odot \mathbf{F}) \mathbf{1}$.

   (iv) Generate the set of $K(K-1)/2$ pairs of integers taken from the set $\{1 \ldots K\}$.

   (v) Let $\mathbf{P}_{\text{opt}} = \mathbf{I}_K$.

2. For each pair $i, j \in \{1 \ldots K\}$

   (i) Let $\mathbf{P}_{i,j}$ be the permutation matrix generated by switching columns i and j of $\mathbf{I}_K$.

(ii) If $\mathbf{1}^T\mathbf{P}_{i,j}\mathbf{P}_{\text{opt}}(\mathbf{G}\odot\mathbf{F})\mathbf{P}_{\text{opt}}^T\mathbf{P}_{i,j}^T\mathbf{1} < 0$, let $\mathbf{P}_{\text{opt}} = \mathbf{P}_{i,j}\mathbf{P}_{\text{opt}}$.

The result of this procedure is that the permutation matrix $\mathbf{P}_{\text{opt}}$ combined with the initial ordering of the users in $\mathbf{G}_0$ reveals the grouping that minimizes the sum of the compatibility metrics. If some metric is used that should be maximized rather than minimized, the only modification required is to change the direction of inequality in step 2(ii).

In computing the inequality in step 2(ii), it is important to note that the only difference between $\mathbf{P}_{i,j}\mathbf{P}_{\text{opt}}(\mathbf{G}\odot\mathbf{F})\mathbf{P}_{\text{opt}}^T\mathbf{P}_{i,j}^T$ and $\mathbf{P}_{\text{opt}}(\mathbf{G}\odot\mathbf{F})\mathbf{P}_{\text{opt}}^T$ is the switching of two rows and columns. As a result, an efficient implementation of the inequality can be devised that only computes the sums of the rows and columns in question. The Compatibility Optimization Algorithm requires all $K(K-1)/2$ steps to be completed to find the global minimum. As noted before, there are $K(K-1)/2$ compatibility metrics to be computed to generate the matrix $\mathbf{G}_0$. So, both of the two steps have $\mathcal{O}(K^2)$ complexity.

After a grouping has been selected, it is still necessary to implement a spatial multiplexing algorithm to generate transmit vectors for each group. The Compatibility Optimization Algorithm is independent of which algorithm is chosen, but the compatibility metric is not necessarily. The correlation metric has been suggested because it worked best with the cases that were tested in the simulations section that follows, but it is possible that for a particular choice of multiplexing algorithm, an alternative metric may yield better performance.

6.2 Sub-Channel Allocation

Up to this point, our discussion of channel allocation has focused on the problem of choosing the best grouping of users without regard to the number of spatial "subchannels" (SSC) being used by each user. One of the fundamental trade-offs in the design of any MIMO system is whether to maximize transmission rate, which is achieved by parallel data transmission schemes, or to maximize diversity, using space-time codes [7]. For channels where the transmitter knows the transfer function in advance, one approach to parallel data transmission is to compute the singular value decomposition of the channel matrix, using the right and left singular vectors as transmit and receive beamformers to

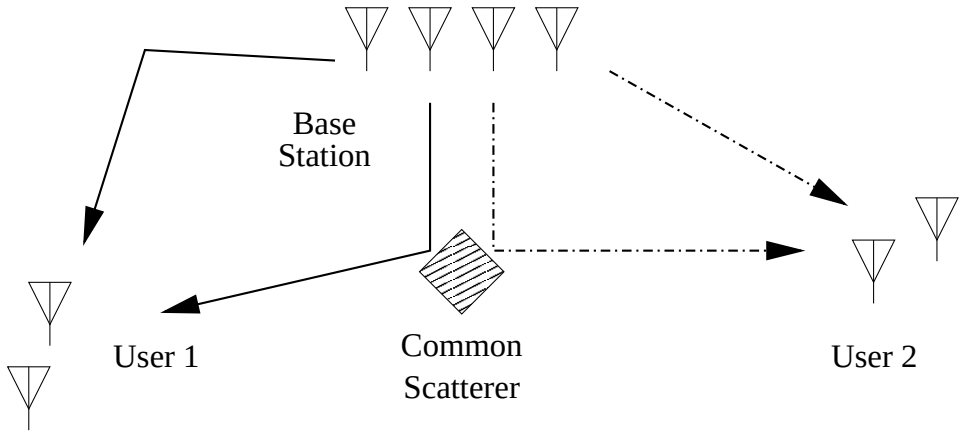

Figure 6.2: An illustration of a base station transmitting over multiple sub-channels to two users.

create independent, orthogonal SSC [6]. We now consider the question of optimal allocation in a multi-user MIMO channel where multiple SSC are used. A similar problem to this is the problem of SSC allocation in OFDM systems [133], where sub-carriers must be allocated to satisfy QoS requirements for each user while assigning multiple users to sub-carriers in a way that economizes resources.

The problem of optimal grouping when using a multi-stream transmission scheme (such as Block-Diagonalization or Generalized Iterative Zero-Forcing) raises interesting questions about channel allocation. This is not really an issue when the system is nearly fully loaded ($K \rightarrow n_T G$), since there is no excess capacity in the system to allow for the transmission of multiple SSC to individual users. On the other hand, when $K \ll n_T G$, the demands on system resources can be reduced by allocating free SSC to users whose channel characteristics will support multiple SSC.

The simplest way to handle the allocation of multiple SSC is to group the users using the algorithm outlined in the last section, and then determine within each group how many SSC are allocated to each user. However, suppose users i and j have been allocated to the same group, and both users have sufficient channel rank to support transmission of data over two SSC. While the two users may have sufficient linear independence to be grouped together when only transmitting one data sub-stream to each user, it my be possible that only some of the SSC of one user are "spatially compatible" with the SSC of

the other user. One example of this is illustrated in Figure 6.2, where users 1 and 2 share a common scatterer in their multipath channels. In this situation, it is possible that better system resource usage could be achieved by allowing the two SSC of users 1 and 2 to be allocated to different groups.

To accomplish this, we propose an independent sub-channel approach to allocation, where the SSC of each user are treated as completely independent channels in the allocation algorithm. Let m_j be the number of sub-channels allocated to user j, and let $m = \sum m_j$. So, for a system of K users, the m total sub-channels are instead treated as independent channels in the allocation algorithm, increasing the complexity of the allocation algorithm to $\mathcal{O}(m^2)$. This requires a different compatibility metric, since the correlation metric reflects only the compatibility between the entire $\mathbf{H}_i$ and $\mathbf{H}_j$, rather than their sub-channels. We propose computing the SVD of $\mathbf{H}_j$. We find the water-filling solution that minimizes transmitted power given the singular values of $\mathbf{H}_j$ and the desired transmission rate, and choose m_j as the number of channels with non-zero power. Then, we use the left singular vectors of $\mathbf{H}_j$ as a means of estimating spatial compatibility between sub-channels of users. If $\mathbf{w}_{i,k}$ is the k^{th} left singular vector of $\mathbf{H}_i$, then let

$$\mathbf{w}_{i,k}^* \mathbf{H}_i \mathbf{H}_j^* \mathbf{w}_{j,l} \tag{6.9}$$

be the spatial compatibility of sub-channel k of user i and sub-channel l of user k. Note that the orthogonality of the left singular vectors of a particular $\mathbf{H}_j$ results in a compatibility metric of zero when $i = j$. This will bias the COA toward putting the sub-channels of the same user in the same group. However, in simulations it was frequently observed that sub-channels of the same user were often allocated to different groups in spite of this, due to one of the sub-channels having sufficiently bad compatibility with other potential members of the group.

6.3 Simulation Results

Figure 6.3 compares the performance of different channel allocation schemes. In this case, there are a total of four antennas at the base station and eight users with the number of antennas at each user chosen randomly from one to three. A single-beam

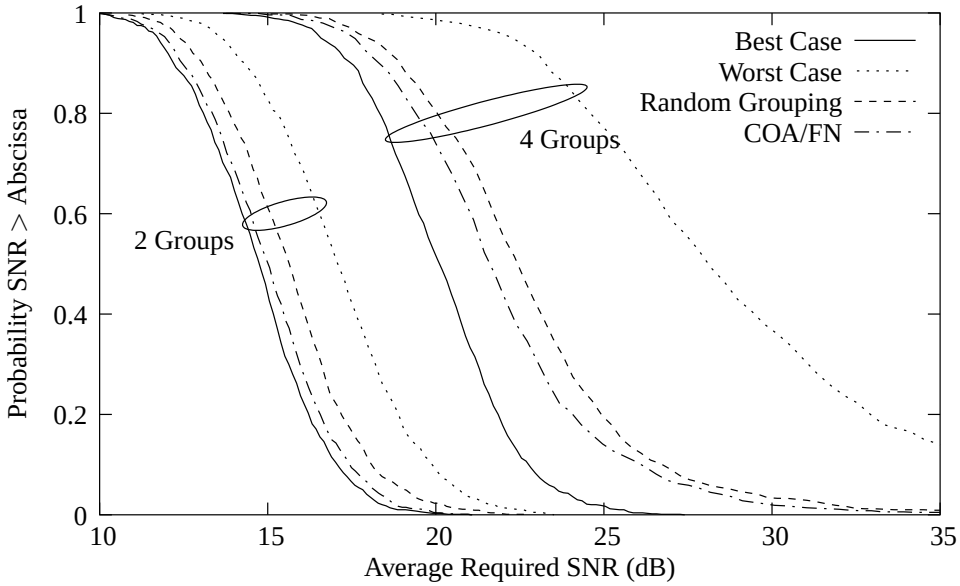

Figure 6.3: A comparison of different grouping algorithms for eight users allocated with a set of either two or four available channels.

algorithm is used to transmit to each user, and the transmission rate for each user is a randomly chosen integer in the interval $[1, 4]$ bits per user. The number of bits per channel was mapped to a SINR requirement (γ_b) using the upper bound on the symbol error rate for uncoded QAM constellations from [99] (equation 5-2-80):

$$P_M \leq 2 \ \mathrm{erfc}\left(\sqrt{\frac{3k}{2(M-1)}\gamma_b}\right), \tag{6.10}$$

where k is the number of bits, $M = 2^k$, and γ_b is the average SNR per bit. The required bit error rate was fixed at 10^{-5}. We consider two different channel configurations, where the eight users are allocated among either two or four available channels. For two channels, there are a total of 35 unique groupings, and for four channels there are a total of 105 unique groupings. The required power to satisfy the requested rate was calculated for each of the unique groupings in order to determine the best and worst cases. The Compatibility Optimization Algorithm is used in conjunction with the Frobenius norm metric, and is labeled "COA/FN".

The results in Figure 6.3 reveal that the Compatibility Optimization Algorithm has a cost between the best possible grouping and randomly chosen grouping for both

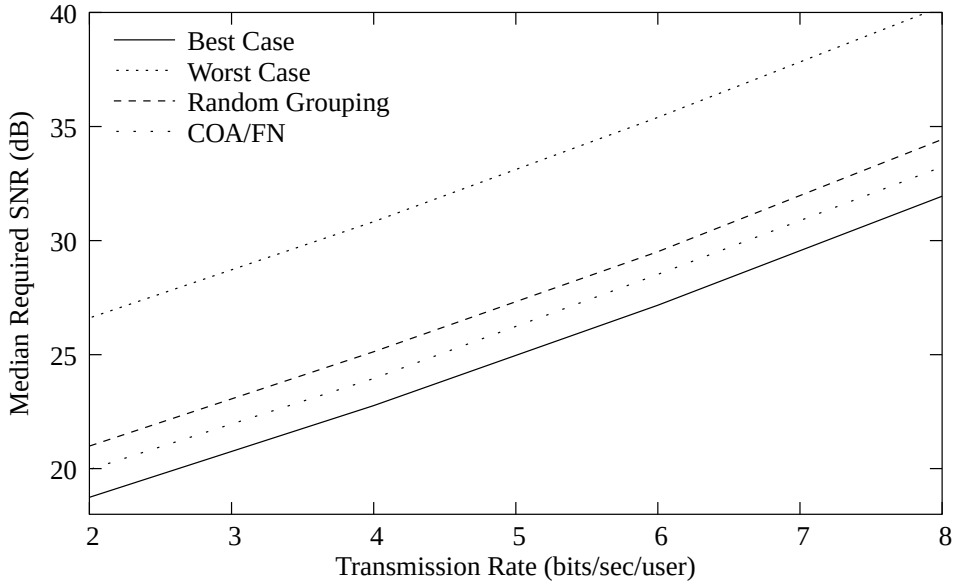

Figure 6.4: A comparison of grouping algorithms for different transmission rates, with eight users allocated to two groups.

group sizes. It should also be noted that the randomly chosen grouping has surprisingly good performance when compared to the worst case.

Figures 6.4 and 6.5 compare the grouping algorithms for different transmission rates, and different user array sizes.

Figure 6.6 shows the effects of channel correlation when the receivers have antennas that are closely spaced. The $\mathbf{H}_j$ matrices are independent of each other, and the columns of $\mathbf{H}_j$ are independent, while the rows have covariance matrix $\mathbf{R}$, where $E\{[\mathbf{R}]_{i,k}[\mathbf{R}]^*_{j,k}\} = \alpha^{|i-j|}$, and α is the correlation parameter on the horizontal axis of Figure 6.6. The values in the figure are the median required SNR for two groups. As the correlation increases, the worst case scenario becomes even worse relative to the other cases, while the CO algorithm performs consistently well in all cases.

Figure 6.7 compares three different approaches to sub-channel allocation as a function of system loading. In this case the total number of groups is fixed at four, and the number of transmitters at the base station is also four, resulting in a maximum supportable user load of 16. The number of users was varied from 4, which means only one user per

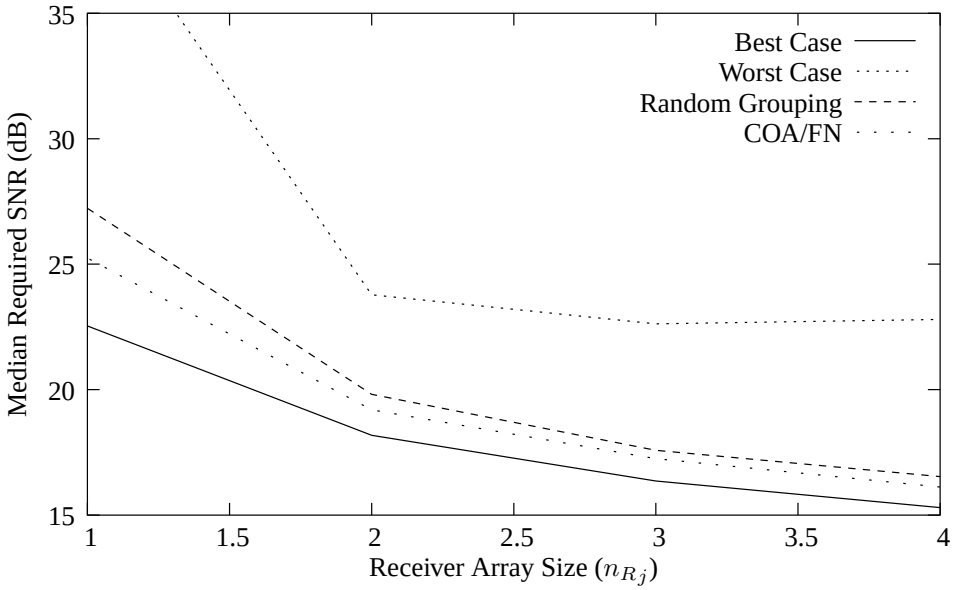

Figure 6.5: A comparison of grouping algorithms for different array sizes at the receivers, with eight users allocated to two groups.

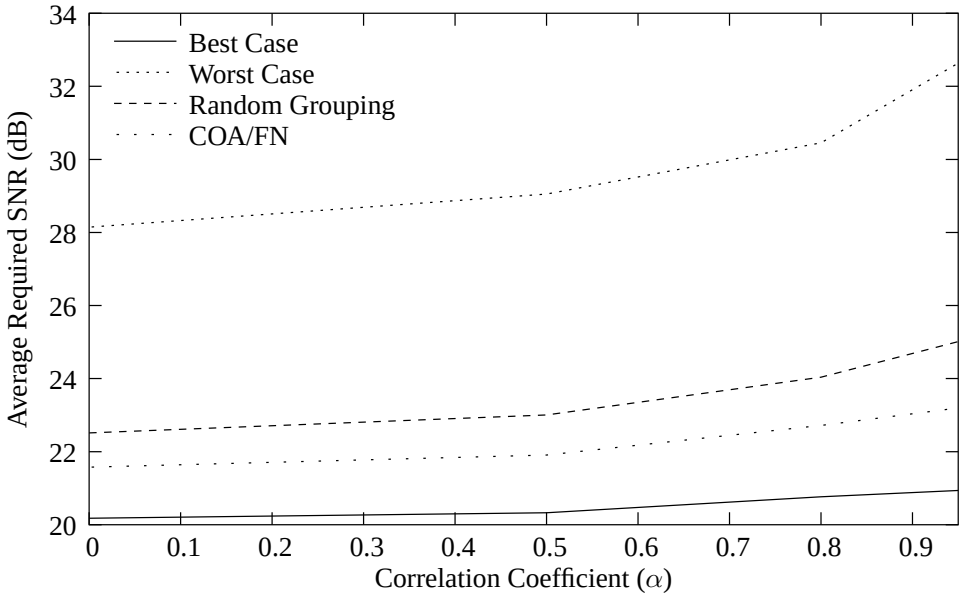

Figure 6.6: A comparison of grouping algorithms at different channel correlations for eight users allocated to two groups.

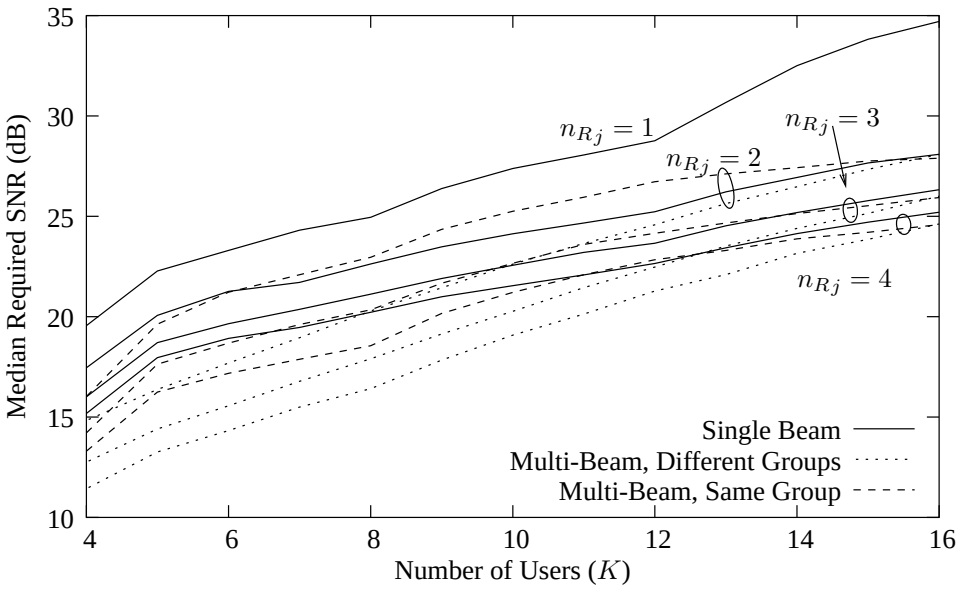

Figure 6.7: A comparison of sub-channel grouping algorithms.

group for the single-beam approach, up to 16, representing a fully loaded system. In addition to the single-beam algorithm, the multi-beam algorithm was tested for two different cases: one using the same grouping as the single-beam algorithm, but allowing additional sub-channels in cases where it was supported, and the second treating all sub-channels as independent channels and performing an independent channel allocation using the metric suggested in the last section. We include results in the plot for values of n_{Rj} from one to four antennas. When $n_{Rj} = 1$ for all users, only one sub-channel per user is available, so the results are the same in all cases.

The results show that the best algorithm is a function of how loaded the system is. Surprisingly, there is minimal advantage gained from allowing multiple sub-channels and requiring them to be in the same group. On the other hand, while all of the results are very close as the system becomes fully loaded, the independent sub-channel approach becomes increasingly superior as the load decreases, by as much as 4 dB. This illustrates the fact that the optimal solution to the channel allocation problem will not necessarily be one that requires all sub-channels for a particular user to be used simultaneously.

6.4 Conclusion

We have illustrated the problem of channel allocation in multi-user MIMO channels when there are more users than can be supported using only SDMA methods, and have proposed an algorithm for assigning users to groups. Simulation results have demonstrated that the Compatibility Optimization Algorithm presented here will choose groupings that are somewhat better than random groupings, but not necessarily the global optimum. The optimal solution to this problem can only be found by global search, so it is not a realistic option for any significant number of users. The COA is therefore a reasonable alternative when performance is important.

We have also studied the problem of channel allocation when multiplexing algorithms are used that allow multiple sub-channels for individual users. The simulation results show that the best possible solution is often achieved when a user's sub-channels are allowed to be allocated to different groups, being treated as independent channels.

An important future step in the study of channel allocation for SDMA systems is to study the problem from the context of an entire network. This is a considerably more complex problem which raises a number of interesting questions:

- In addition to *a priori* channel information, is it feasible for a base station to have advance interference information (*e.g.* interference from other base stations–co-channel interference within one's own cell is controllable)? How would this information affect network performance?

- When a mobile station has multiple antennas, and thus the possibility for directivity, does this blur the boundaries of what constitutes a "cell", since it may be possible to direct one's strongest beam toward a base that is not necessarily the closest?

- If it is possible for a user with multiple data streams to benefit from allocating to separate time-domain channels, what about allocating them to different base stations? This may be a very complex problem to solve, but in addition to better optimizing the use of network resources, it could have additional benefits such as improving cell "hand off".

Taking all of these ideas into consideration would likely result in a system whose complexity is unacceptable, but nevertheless it will be useful in the future to design network planning algorithms that exploit the spatial properties of MIMO systems even to a limited degree because of the potential they have to improve the capacity of the network.

Chapter 7

Conclusion

This chapter contains some concluding remarks, beginning with a more detailed chapter-by-chapter summary of the new contributions of this thesis, followed by a discussion of areas for future research.

7.1 New Contributions

Each of Chapters 3-6 contains significant new contributions, though some of them are not entirely unique, as they have been recently published concurrently by other authors. These cases will be noted below. In addition, Chapter 2, while mostly being a review and introduction, contains one new contribution: the derivation of single-user MIMO capacity for channels with partial channel information (*e.g.* $\mathbf{H} = \mathbf{AB}$, with only $\mathbf{B}$ known to the transmitter).

Chapter 3 introduced the Block-Diagonalization algorithm, Successive Optimization algorithm, and non-iterative coordinated Tx-Rx processing algorithm. The basic structure of the BD algorithm was proposed concurrently by three authors [37–39]. However, the concept is developed here in much more depth, including the adaptability of the BD algorithm for solving both throughput maximization and power control problems, as well as the BD solution for channels with partial information at the transmitter, all of which constitute unique contributions of this thesis. The SO algorithm and coordinated beamforming algorithm are also new contributions. The coordinated beamforming algorithm presented in Chapter 3 differs only slightly from the Generalized Iterative Zero-Forcing algorithm in Chapter 4. Its main point is to illustrate how it is possible to accomodate any channel dimensions without resorting to iterative algorithms.

In Chapter 4, the Generalized Iterative Zero-Forcing algorithm has also been introduced recently elsewhere [134]. The other new contributions of this chapter, however, are all new. This includes the generalization of the optimal downlink beamforming algorithm to a coordinated transmitter-receiver processing structure with imperfect channel estimates. A slightly different approach to this problem can be found in [90], but otherwise, there is still much room for more research in this area.

The use of measured channels is always important to verify the usefulness of proposed transmission schemes and system designs under realistic conditions. With the high current interest in MIMO systems in the research community, there are several very recent publications which contain experimental results for single-user MIMO channels, but the multi-user aspects have not yet been measured experimentally, so all of the results contained in Chapter 5, both from measurement and modeling, are new contributions. In addition, this chapter contains some new analysis of the SVA model in [46], which shows the connection between the SVA model parameters and the total power of each cluster. This is useful in understanding the role of the model parameters (specifically the product $\Lambda\Gamma$) in determining the eigenvalue distribution of the channel matrix, and thus the channel capacity.

The channel allocation problem discussed in Chapter 6 is an emerging area of research which has not previously addressed for either the downlink beamforming or dirty-paper coding approaches to SDMA. The idea of a pairwise compatibility metric exists in the literature, but previously the channels were only vectors, so a more general compatibility metric (the Frobenius norm) was proposed. The Compatibility Optimization Algorithm is unique in that it optimizes globally over a set of compatibility metrics. The other unique idea in this chapter is the separation of sub-channels belonging to a single user to be allocated to different time-domain channels. The results demonstrate that this is a promising idea.

7.2 Discussion and Future Research

Taken collectively, the new algorithms and results presented in this thesis give some important conclusions and suggest many areas for future research. Some important

points addressed below include the importance of channel information at the transmitter, the comparison of orthogonal and non-orthogonoal beamforming, the question of the number of data streams to be used, and more extensive channel allocation.

7.2.1 Channel Information

One important conclusion to be drawn from the simulations in Chapter 3 is the usefulness of channel information at the transmitter. It has been shown in the literature that for single-user channels there is not much difference between the performance of informed and uninformed transmitters as SNR increases. However, in the multi-user channel, the information is much more important, as spatial multiplexing methods such as block-diagonalization generally require without it, and spatial multiplexing can provide a significant performance gain over TDMA or FDMA.

It has been standard practice here and in other recent publications to assume perfect information at the transmitter. This very optimistic assumption is a good starting point for studying downlink transmission schemes, but of course this is rarely true in practice. Important questions that remain to be answered to their full extent are how can the algorithms be adapted for imperfect or incomplete information (these have begun to be addressed here in two cases). One recent proposal is long-term averaging of the channel information to remove fast fading effects [135].

Multi-user MIMO systems are being considered for two-way communication systems, so the most natural way of obtaining this information is via feedback from the mobiles in the uplink channel. In practical systems, there will be limits on how much of the uplink bandwidth can be used for transmitting data, so other important questions are what exactly should be transmitted and what is the trade-off between feedback bandwidth and gain in system downlink throughput.

7.2.2 Orthogonal vs. Non-Orthogonal Solutions

The algorithms in this thesis can be grouped into two categories: those which decompose the channel into orthogonal sub-channels, and those which balance interference and noise, and are thus non-orthogonal by nature. The orthogonalization algorithms (block

diagonalization and generalized zero-forcing) are able to decompose the channel independent of the particular requirements. The power distribution can then be determined in a separate step depending on the optimization constraints of interest (total throughput, power control). The particular advantage of the block-diagonalization algorithm in particular is the low computational complexity.

The interference balancing algorithms on the other hand are iterative in nature, and thus require significantly more computation, the cost of which can't be known in advance. In addition, they are not as adaptable because each solution is designed around a specific set of requirements. On the other hand, perhaps the most compelling argument for for non-orthogonal beamforming is the increased robustness to channel estimation error seen in the results of Chapter 5 and the capability of incorporating channel information error statistics into the beamformer design, as shown in Chapter 4. The zero-forcing approach may perhaps be ultimately most useful for its role in the hybrid algorithm of Chapter 4, where it helps reduce the computation time of the interference-balancing. The hybrid algorithm may prove even more practical in the future, since one can safely assume that the capabilities of digital signal processors (DSPs) will continue to increase. Nonetheless, achieving further reductions in the required computations to achieve robust multi-user beamforming remains an important research problem. For the case where all users have one antenna, one recent non-orthogonal solution is "regularized" channel inversion: $\mathbf{M}_S = \mathbf{H}_S^*(\mathbf{H}_S\mathbf{H}_S^* + \alpha\mathbf{I})^{-1}$ [68]. This solution has near-optimal performance and can be computed in closed form, and thus is substantially faster than optimal beamforming. An intriguing question is whether this scheme can also be generalized to the block case (as channel inversion has been generalized using the BD algorithm). Such a solution could also be near-optimal while having a computational cost similar to that of the BD algorithm.

7.2.3 The Number of Data Streams

The one thing that made the original BLAST and related algorithms so intriguing is their high data rates, which was achieved by transmitting multiple data streams in parallel. This thesis has focused mainly on methods that allow such multi-stream transmission simultaneously to multiple users, but the usefulness of doing this in multi-user

systems is dependent on various factors. As a system becomes close to fully loaded, where the number of base station antennas is close to the number of users, it is not possible to use multiple data streams for a single user, and so this is not an issue. When the load is light enough to allow multiple data streams, the results in Chapter 5 show that sometimes the "less is more" principle applies here. In a full-rank $\{5, 5\} \times 10$ channel, the system throughput was maximized sending only three data streams to each user, rather than the maximimum possible number of five. For any given set of users, the optimal number of streams per user depends on the interaction between all users channels, so finding the globally optimum solution may require an exhaustive search. One recent work has proposed a method of estimating the best value [136], but this question warrants further study.

7.2.4 Allocation

The channel allocation problem for multi-user MIMO channels is very complex. The optimal solution can only be found by global search, so all practical solutions will trade some performance for cost. The simulation results of Chapter 6 revealed surprisingly good performance for random grouping. The allocation algorithm presented here is only a starting point. In the future, it will be important to consider the channel allocation problem from a network point of view. This raises the questions mentioned in Chapter 6 of which base station a user should belong to and whether more efficient resource usage could be achieved communicating with multiple base stations simultaneously. Future communication networks will rquire channel allocation algorithms can can consistently choose a good distribution of resources at a reasonable cost.

Another emerging area of research that is somewhat related to the channel allocation problem is ad-hoc networking. Typically, this refers to a self-organizing network of nodes that is independent of any base stations. However, in its broadest interpretation, this could also refer to networks where a base station communicates with users via relaying messages among its users. This could be used to extend the geographic reach of a base station, or to reach users whose best propagation path to the base may be via another user. The use of this type of networking when all users have arrays presents many new opportunities and challenges. The same considerations of computational complexity and availability of

channel knowledge would now apply to every member of the network. In addition, when multiple base stations or ad-hoc networks are considered, it is no longer possible to ignore external interference sources. Interference statistics at the receiver can only be known by feedback to the transmitter, which represents another burden on the reverse channel.

7.3 Summary

As was stated in the introduction, the general approach of the algorithms presented here is achieving spatial multiplexing by signal processing at the base station. Some processing is required at the mobile stations, but it is minimal, which is important for most types of mobile devices. The alternative approach is to use the next generation of codes that use the dirty paper principle to avoid inter-user interference. This approach will likely be important in the future, but it will require entirely new codes, and thus new communication protocols and standards. The advantage of achieving interference avoidance by designing appropriate transmit vectors at the base is that it can more easily be applied to existing systems, extending their capacity, allowing them and future systems to meet the ever-increasing demands of new wireless communications services.

Bibliography

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, July, October 1948.

[2] S. Benedetto, G. Montorsi, and D. Divsalar, "Concateneated Convolutaional Codes with Interleavers," *IEEE Communications Magazine*, vol. 41, no. 8, pp. 102–109, August 2003.

[3] E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, November/December 1999.

[4] G. J. Foschini and M. J. Gans, "On Limits of Wireless Personal Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[5] 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 *Proceedings of ISSSE-98*. IEEE, September 1998.

[6] G. G. Raleigh and J. M. Cioffi, "Spatio-Temporal Coding for Wireless Communication," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, March 1998.

[7] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

[8] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[9] D. W. Bliss, K. W. Forsythe, A. O. Hero, and A. L. Swindlehurst, "MIMO Environmental Capacity Sensitivity," in *Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers*. IEEE, October 2000.

[10] S. Catreux, L. J. Greenstein, and V. Erceg, "Some Results and Insights on the Performance Gains of MIMO Systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 839–847, June 2003.

[11] S. M. Alamouti, "A Simple Transmitter Diversity Scheme for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[12] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-Time Block Codes From Orthogonal Designs," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1456–1467, July 1999.

[13] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith, and A. Naguib, "From Theory to Practice: An Overview of MIMO Space-Time Coded Wireless Systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281–302, April 2003.

[14] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-Optimal BLAST Processing With Multiple-Access Interference," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 1, Boston, MA, September 24-28 2000, pp. 87–91.

[15] R. S. Blum, J. H. Winters, and N. Sollenberger, "On The Capacity of Cellular Systems With MIMO," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 2. Atlantic City, NJ: IEEE, Oct. 7-11 2001, pp. 1220–1224.

[16] B. G. Agee, "Exploitation of Internode MIMO Channel Diversity in Spatially Distributed Multipoint Communication Networks," in *Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers*. IEEE, November 2001.

[17] R. W. Heath, M. Airy, and A. J. Paulraj, "Multiuser Diversity for MIMO Wireless Systems With Linear Receivers," in *Conference Record of the 35th Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA: IEEE, November 4-7 2001, pp. 1194–1199.

[18] M. F. Demirkol and M. A. Ingram, "Power-Controlled Capacity for Interfering MIMO Links," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 1. Atlantic City, NJ: IEEE, Oct. 7-11 2001, pp. 187–191.

[19] R. S. Blum, "MIMO Capacity with Interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 793–801, June 2003.

[20] J. Li, K. B. Letaief, M. A. Zhengxin, and Z. Cao, "Spatial Multiuser Access with MIMO Smart Antennas for OFDM Systems," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 3. Atlantic City, NJ: IEEE, Oct. 7-11 2001, pp. 1553–1557.

[21] P. Viswanath, D. N. C. Tse, and V. Anantharam, "Asymptotically Optimal Waterfilling in Vector Multiple Access Channels," *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 241–267, 2001.

[22] D. C. Popescu, O. Popescu, and C. Rose, "Interference Avoidance for Multiaccess Vector Channels," in *Proceedings of the International Symposium on Information Theory*. IEEE, July 2002, p. 499.

[23] J. Wang and K. Yao, "Multiuser Spatio-Temporal Coding for Wireless Communications," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, vol. 1. IEEE, March 2002, pp. 276–279.

[24] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," August 2002, submitted to *IEEE Transactions on Information Theory*.

[25] G. Caire and S. Shamai, "On the Achievable Throughput of a Multi-Antenna Gaussian Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[26] P. Viswanath and D. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Transactions on Information Theory*, vol. 41, no. 8, pp. 1912–1921, August 2003.

[27] W. Yu and J. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," November 2001, submitted to *IEEE Transactions on Information Theory*.

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

[29] C. B. Peel, "On "Dirty-Paper Coding"," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 112–113, May 2003.

[30] H. Boche and M. Schubert, "Multi-Antenna Downlink Transmission with Individual SINR Receiver Constraints for Cellular Wireless Systems," in *Proceedings of the 4th International ITG Conference on Source and Channel Coding*, Informationstechnische Gesellschaft im VDE (ITG). VDE Verlag GmbH, January 2002, pp. 159–166.

[31] M. Schubert and H. Boche, "An Efficient Algorithm for Optimum Joint Downlink Beamforming and Power Control," in *Proceedings of the 55th IEEE Vehicular Technology Conference*, vol. 4. Birmingham, Alabama: IEEE, May 6-9 2002, pp. 1911–1915.

[32] M. Bengtsson and B. Ottersten, "Optimal and Suboptimal Beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed. CRC Press, August 2001.

[33] M. Bengtsson, "A Pragmatic Approach to Multi-User Spatial Multiplexing," in *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop*. IEEE, August 2002.

[34] 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 *Proceedings of the Wireless Communications and Networking Conference*, vol. 1. IEEE, 2000, pp. 426–430.

[35] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint Transmitter Receiver Diversity for Efficient Space Division Multiaccess," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 16–27, January 2002.

[36] Q. H. Spencer and A. L. Swindlehurst, "Some Results on Channel Capacity When Using Multiple Antennas," in *Proceedings of VTC Fall 2000*. IEEE, 2000.

[37] R. L.-U. Choi and R. D. Murch, "A Downlink Decomposition Transmit Preprocessing Technique for Multi-user MIMO Systems," in *Proceedings of IST Mobile & Wireless Telecommunications Summit*, June 2002.

[38] M. Rim, "Multi-user downlink beamforming with multiple transmit and receive antennas," *Electronics Letters*, vol. 38, no. 25, pp. 1725–1726, 5th December 2002.

[39] Q. H. Spencer and M. Haardt, "Capacity and Downlink Transmission Algorithms for a Multi-user MIMO Channel," in *Conference Record of the 36th Asilomar Conference on Signals, Systems and Computers*. IEEE, November 2002.

[40] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multi-User MIMO Channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, February 2004.

[41] ——, "Fast Power Minimization with QoS Constraints in Multi-User MIMO Downlinks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, April 2003.

[42] Q. H. Spencer and A. L. Swindlehurst, "A Hybrid Approach to Spatial Multiplexing in Multi-User MIMO Downlinks," September 2003, submitted to *EURASIP Journal on Wireless Communications and Networking*.

[43] Q. H. Spencer, T. Svantesson, and A. L. Swindlehurst, "MIMO Downlink Spatial Multiplexing Algorithms Applied to Channel Measurements and Models," January 2004, submitted to *Wireless Communications and Mobile Computing*.

[44] Q. H. Spencer and A. L. Swindlehurst, "Channel Allocation in Multi-user MIMO Wireless Communications Systems," in *Proceedings of the IEEE International Conference on Communictions*. Paris: IEEE, June 2004.

[45] Q. H. Spencer, B. Farhang-Boroujeny, and A. L. Swindlehurst, "Untitled," January 2004, to be submitted to *Unknown Journal*.

[46] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, "Modeling the Statistical Time and Angle of Arrival Characteristics of an Indoor Multipath Channel," *IEEE Journal on Selected Areas In Communications*, vol. 18, no. 3, pp. 347–360, March 2000.

[47] T. Svantesson, "A Double-Bounce Channel Model for Multi-Polarized MIMO Systems," in *Proceedings of VTC 2002-Fall*, vol. 2. Vancouver, BC: IEEE, September 2002, pp. 691–695.

[48] M.-S. Alouini and A. J. Goldsmith, "Comparison of Fading Channel Capacity Under Different CSI Assumptions," in *Proceedings of the IEEE 52nd Vehicular Technology Conference*, 4, Ed. Boston, MA, USA: IEEE, September 24-28 2000, pp. 1844–1849.

[49] A. Das and P. Narayan, "Capacities of Time-Varying Multiple-Access Channels with Side Information," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 4–25, January 2002.

[50] G. Caire and S. Shamai, "On the Capacity of Some Channels with Channel State Information," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2007–2019, September 1999.

[51] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, Correlations, and Capacities of Multielement Transmit and Receive Antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 361–368, April 2002.

[52] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading Correlation and its Effect on the Capacity of Multielement Antenna Systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, March 2000.

[53] M. T. Ivrlač, W. Utschick, and J. A. Nossek, "Fading correlations in wireless mimo communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 819–828, June 2003.

[54] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO Channels: Asymptotic Evaluation Under Correlated Fading," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 829–838, June 2003.

[55] Z. Hong, K. Liu, R. W. Heath, Jr., and A. M. Sayeed, "Spatial Multiplexing in Correlated Fading via the Virtual Channel Representation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 856–866, June 2003.

[56] G. German, "Ray-based Analysis of Indoor MIMO Channels for Statistical Model and Channel State Feedback," Master's thesis, Brigham Young University, December 2003.

[57] M. T. Ivrlač and J. A. Nossek, "Correlated Fading in MIMO-Systems–Blessing or Curse?" in *39th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, USA, October 2001.

[58] D. A. Gore, R. U. Nabar, and A. Paulraj, "Selecting an Optimal Set of Transmit Antennas for a Low Rank Matrix Channel," in *Proceedings of ICASSP 2000*. IEEE, 2000.

[59] A. M. Tehrani, R. Negi, and J. Cioffi, "Space-Time Coding And Transmission Optimization for Wireless Channels," in *Proceedings of Asilomar 1998*. IEEE, 1998.

[60] ——, "Space-Time Coding Over A Code Division Multiple Access System," in *Proceedings of WCNC 1999*. IEEE, 1999.

[61] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and Performance of Some Basic Space-Time Architectures," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 303–320, April 2003.

[62] M. O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice Code Decoder for Space-Time Codes," *IEEE Communications Letters*, vol. 4, no. 5, pp. 161–163, May 2000.

[63] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communications: Performance Criterion and Code Construction," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 744–765, March 1998.

[64] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[65] R. W. Heath, Jr. and A. J. Paulraj, "Diversity Versus Multiplexing in Narrowband MIMO Channels: A Tradeoff Based on Euclidean Distance," December 2002, submitted to *IEEE Transactions on Communications*.

[66] T. J. Oechtering and H. Boche, "On the Capacity of a Distributed Multiantenna System Using Cooperative Transmitters," in *Proceedings of the 57th IEEE Vehicular Technology Conference*. Jeju, Korea: IEEE, April 22-25 2003, pp. 75–79.

[67] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink Capacity Evaluation of Cellular Networks with Known-Interference Cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 802–811, June 2003.

[68] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication," June 2003, submitted to *IEEE Transactions on Communications*, Available at `http://mars.bell-labs.com`.

[69] C. B. Peel, “A Hodge-Podge ...” Ph.D. dissertation, Brigham Young University, December 2003.

[70] J. H. Winters, J. Salz, and R. D. Gitlin, “The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems,” *IEEE Transactions on Communications*, vol. 42, no. 2, pp. 1740–1751, Feb/Mar/Apr 1994.

[71] T. Haustein, C. von Helmolt, E. Jorswieck, V. Jungnickel, and V. Pohl, “Performance of MIMO Systems With Channel Inversion,” in *Proceedings of the IEEE 55th Vehicular Technology Conference*, vol. 1. Vancouver, BC: IEEE, September 2002, pp. 35–39.

[72] W. Qiu, H. Tröger, and M. Meurer, “Joint Transmission (JT) in Multi-user MIMO Transmission Systems,” in *COST 273 TD(02)008*. EURO-COST, January 2002.

[73] E. Visotsky and U. Madhow, “Optimum Beamforming Using Transmit Antenna Arrays,” in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 1. Houston, TX: IEEE, May 16-20 1999, pp. 851–856.

[74] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, “Transmit Beamforming and Power Control for Cellular Wireless Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437–1450, October 1998.

[75] H. Boche and M. Schubert, “Optimum SIR Balancing Using Extended 1-Norm Beamforming Optimization,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3. IEEE, 2002, pp. 2945–2948.

[76] H. Boche, M. Schubert, and E. A. Jorswieck, “Throughput Maximization for the Multiuser MIMO Broadcast Channel,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. Hong Kong: IEEE, April 6-10 2003, pp. 808–811.

[77] H. Boche and M. Schubert, “Effective Bandwidth Maximization for Uplink/Downlink Multi-Antenna Systems,” in *Proceedings of the IEEE International*

Conference on Communications. Anchorage, Alaska: IEEE, May 2003, pp. 3215–3219.

[78] ——, "SIR Balancing for Multiuser Downlink Beamforming – A Convergence Analysis," in *IEEE International Conference on Communications (ICC)*, vol. 2, New York City, April 2002, pp. 841–845.

[79] ——, "A General Duality Theory for Uplink and Downlink Beamforming," in *Proceedings of the IEEE 56th Vehicular Technology Conference*, vol. 1. Vancouver, BC: IEEE, September 2002, pp. 87–91.

[80] E. Jorswieck and H. Boche, "Rate Balancing for the Multi-Antenna Gaussian Broadcast Channel," in *Proceedings of the IEEE 7th International Symposium on Spread Spectrum Techniques and Applications*, vol. 2. Prague, Czech Republic: IEEE, September 2-5 2002, pp. 545–549.

[81] H. Boche, M. Scbubert, and E. A. Jorswieck, "Trace Balancing for Multiuser MIMO Downlink Transmission," in *Conference Record of the 36th Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, CA: IEEE, November 2002, pp. 1379–1383.

[82] H. Boche and M. Schubert, "Analysis of Different Precoding/Decoding Strategies for Multiuser Beamforming," in *Proceedings of the 57th IEEE Vehicular Technology Conference*. Jeju, Korea: IEEE, April 22-25 2003, pp. 39–43.

[83] M. Schubert and H. Boche, "Joint 'Dirty Paper' Pre-coding and Downlink Beamforming," in *Proceedings of the IEEE 7th International Symposium on Spread Spectrum Techniques and Applications*, vol. 2. Prague, Czech Republic: IEEE, September 2-5 2002, pp. 536–540.

[84] M. Schubert, D. Karadoulamas, H. Boche, and G. Lehmann, "Joint Downlink Beamforming and Power Control for 3G WCDMA," in *Proceedings of the 57th IEEE Semiannual Vehicular Technology Conference*. Jeju, Korea: IEEE, April 22-25 2003, pp. 331–335.

[85] T. Haustein, M. Schubert, and H. Boche, "On Power Reduction Strategies for the Multi-User Downlink With Decentralized Receivers," in *Proceedings of the 57th IEEE Vehicular Technology Conference*, vol. 2. IEEE, April 22-25 2003, pp. 1007–1011.

[86] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[87] D. P. Palomar, M. A. Lagunas, and J. Cioffi, "Optimum Linear Joint Transmit/Receive Processing for MIMO Channels with QoS Constraints," 2003, to appear, *IEEE Transactions on Signal Processing*.

[88] D. P. Palomar and M. A. Lagunas, "Joint Transmit-Receive Space-Time Equalization in Spatially Correlated MIMO Channels: A Beamforming Approach," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 730–743, June 2003.

[89] Z. Pan, K.-K. Wong, and T. Ng, "MIMO Antenna System for Multi-User Multi-Stream Orthogonal Space Division Multiplexing," in *Proceedings of the IEEE International Conference on Communications*, vol. 5. Anchorage, Alaska: IEEE, May 2003, pp. 3220–3224.

[90] M. Bengtsson, "Pragmatic Multi-User Spatial Multiplexing with Robustness to Channel Estimation Errors," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, April 2003.

[91] E. Visotsky and U. Madhow, "Space-Time Transmit Precoding With Imperfect Feedback," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2632–2639, September 2001.

[92] H. Boche and E. Jorswieck, "Analysis of Diversity and Multiplexing Tradeoff for Multi-Antenna Systems with Covariance Feedback," in *Proceedings of the IEEE 56th Vehicular Technology Conference*, vol. 2. Vancouver, BC: IEEE, September 2002, pp. 864–868.

[93] D. Hughes-Hartogs, "Ensemble Modem Structure for Imperfect Transmission Media," U.S. Patents Nos. 4,679,227 (July 1987), 4,731,816 (March 1988), and 4,833,706 (May 1989).

[94] J. A. C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.

[95] J. Campello, "Practical Bit Loading for DMT," in *Proceedings of the IEEE International Conference on Communications*. IEEE, 1999, pp. 801–805.

[96] R. V. Sonalkar and R. R. Shively, "An Efficient Bit-Loading Algorithm for DMT Applications," *IEEE Communications Letters*, vol. 4, no. 3, pp. 80–82, March 2000.

[97] T. Haustein and H. Boche, "Optimal Power Allocation for MSE and Bit-Loading in MIMO Systems and the Impact of Correlation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. Hong Kong: IEEE, April 6-10 2003, pp. 405–408.

[98] J. Gao and M. Faulkner, "On Implementation of Bit-Loading Algorithms for OFDM Systems With Multiple-Input Multiple-Output," in *Proceedings of the IEEE 56th Vehicular Technology Conference*, vol. 1. Vancouver, BC: IEEE, September 2002, pp. 199–203.

[99] J. G. Proakis, *Digital Communications*, 3rd ed. McGraw Hill, 1989.

[100] T. Svantesson and J. Wallace, "Statistical Characterization of the Indoor MIMO Channel Based on LOS/NLOS Measurements," in *Proceedings of the 36th Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, November 2002.

[101] J.-P. Kermoal, L. Schumacher, P. E. Mogensen, and K. I. Pedersen, "Experimental Investigation of Correlation Properties of MIMO Radio Channels for Indoor Picocell Scenarios," in *Proceedings of VTC Fall 2000*. Boston, MA: IEEE, September 2000.

[102] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A Stochastic MIMO Radio Channel Model With Eperimental Validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, August 2002.

[103] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek, and R. S. Thomä, "Capacity of MIMO Systems Based on Measured Wireless Channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 561–569, April 2002.

[104] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Correlation Analysis Based on MIMO Channel Measurements in an Indoor Environment," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 713–720, June 2003.

[105] J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, "Experimental Characterization of the MIMO Wireless Channel: Data Acquisition and Analysis," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 335–343, March 2003.

[106] D. McNamara, M. Beach, P. Fletcher, and P. Karlsson, "Temporal Variation of Multiple-Input Multiple-Output (MIMO) Channels in Indoor Environments," in *Proceedings of the IEE Eleventh International Converence on Antennas and Propagation*, vol. 2. Manchester, UK: IEE, April 17-20 2001, pp. 578–582.

[107] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-Input-Multiple-Output Measurements and Modeling in Manhattan," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 321–331, April 2003.

[108] A. Adjoudani, E. C. Beck, A. P. Burg, G. M. Djuknic, T. G. Gvoth, D. Haessig, S. Manji, M. A. Milbrodt, M. Rupp, D. Samardzija, A. B. Siegel, I. Sizer, Tod, C. Tran, S. Walker, S. A. Wilkus, and P. W. Wolniansky, "Prototype Experience for MIMO BLAST Over Third-Generation Wireless System," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 440–451, April 2003.

[109] S. Howard, H. Inanoglu, J. Ketchum, M. Wallace, and R. Walton, "Results From MIMO Channel Measurements," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 4. IEEE, September 15-18 2002, pp. 1932–1936.

[110] C. C. Martin, J. H. Winters, and N. R. Sollenberger, "Multiple-Input Multiple-Output (MIMO) Radio Channel Measurements," in *Proceedings of the IEEE 52nd Vehicular Technology Conference*, vol. 2. Boston, MA: IEEE, September 24-28 2000, pp. 774–779.

[111] A. A. M. Saleh and R. A. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," *IEEE Journal on Selected Areas of Communications*, vol. SAC-5, pp. 128–13, February 1987.

[112] J. W. Wallace and M. A. Jensen, "Modeling the Indoor MIMO Wireless Channel," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 5, pp. 591–599, May 2002.

[113] R. B. Ertel and J. R. Reed, "Angle and Time of Arrival Statistics for Circular and Elliptical Scattering Models," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1829–1840, November 1999.

[114] C. Oestges, V. Erceg, and A. J. Paulraj, "A Physical Scattering Model for MIMO Macrocellular Broadband Wireless Channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 721–729, June 2003.

[115] P. Zetterberg, M. Bengtsson, D. McNamara, P. Karlsson, and M. A. Beach, "Performance of Multiple-Receive Multiple-Transmit Beamforming in WLAN-Type Systems Under Power or EIRP Constraints with Delayed Channel Estimates," in *Proceedings of the 55th Vehicular Technology Conference*, vol. 4. IEEE, 2002, pp. 1906–1910.

[116] T. Svantesson and L. Swindlehurst, "A Performance Bound for Prediction of a Multipath MIMO Channel," in *Proceedings of the 37th Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, November 2003.

[117] K. Sulonen, P. Suvikunnas, L. Vuokko, J. Kivinen, and P. Vainikainen, "Comparison of MIMO Antenna Configurations in Picocell and Microcell Environments," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 703–712, June 2003.

[118] D. Aktas and H. El Gamal, "Multiuser Scheduling for MIMO Wireless Systems," in *Proceedings of the IEEE 58th Vehicular Technology Conference*. Orlando, FL: IEEE, October 6-9 2003.

[119] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," *IEEE Personal Communications*, vol. 3, no. 3, pp. 10–31, June 1996.

[120] J. Jiang, T.-H. Lai, and N. Sounarajan, "On Distributed Dynamic Channel Allocation in Mobile Cellular Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 10, pp. 1024–1037, October 2002.

[121] J. S. Blogh, P. J. Cherriman, and L. Hanzo, "Adaptive Antenna Array Assited Dynamic Channel Allocation Techniques," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 305–311, February 2001.

[122] ——, "Dynamic Channel Allocation Techniques Using Adaptive Modulation and Adaptive Antennas," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 312–321, February 2001.

[123] L. Chen, H. Murata, S. Yoshida, and S. Hirose, "A dynamic channel assignment algorithm for cellular system with adaptive array antennas," in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 1. Houston, TX, USA: IEEE, May 16-20 1999, pp. 204–208.

[124] A. Alexiou and R.-H. Yan, “Downlink Capacity Enhancement By Employing SDMA In GSM,” in *Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, MA, USA, March 16-17 2000, pp. 413–417.

[125] C. Farsakh and J. A. Nossek, “Channel Allocation and Downlink Beamforming in an SDMA Mobile Radio System,” in *Proceedings of the Sixth IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 2. Toronto, Canada: IEEE, September 27-29 1995, pp. 687–691.

[126] ——, “A Real Time Downlink Channel Allocation Scheme for an SDMA Mobile Radio System,” in *Proceedings of the Seventh IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 3, Taipei, Taiwan, October 15-18 1996, pp. 1216–1220.

[127] S. Harano, Y. Akaiwa, and J. Uchibori, “Performance of Dynamic Channel Assignment Methods In Cellular Systems Using Beam Tilting And Adaptive Array,” in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 4. Amsterdam, Netherlands: IEEE, September 19-22 1999, pp. 2092–2095.

[128] L. C. Godara, “Applications of Antenna Arrays to Mobile Communications. I. Performance Improvement, Feasibility, and System Considerations,” *Proceedings of the IEEE*, vol. 85, no. 7, pp. 1031–1060, July 1997.

[129] V. K. Lau, Y. Liu, and T. A. Chen, “Optimal Multi-User Space Time Scheduling for Wireless Communications,” in *Proceedings of the IEEE 56th Vehicular Technology Conference*, vol. 4. Vancouver, BC: IEEE, September 2002, pp. 1939–1942.

[130] P. Viswanath, D. N. C. Tse, and R. Laroia, “Opportunistic Beamforming Using Dumb Antennas,” *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[131] B. Farhang-Boroujeny and Q. Spencer, “Layering Techniques for Space-Time Communication in Multi-User Networks,” in *Proceedings of the IEEE 58th Vehicular Technology Conference*. Orlando, FL: IEEE, October 6-9 2003.

[132] T. Ohgane, Y. Ogawa, and K. Itoh, “A Study on a Channel Allocation Scheme With an Adaptive Array in SDMA,” in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 2, Phoenix, AZ, USA, May 4-7 1997, pp. 725–729.

[133] I. Koutsopoulos and L. Tassiulas, “Adaptive Resource Allocation in SDMA-Based Wireless Broadband Networks With OFDM Signaling,” in *Proceedings of IEEE INFOCOM 2002*, vol. 3, 2002, pp. 1376–1385.

[134] R. L.-U. Choi, M. T. Ivrlač, R. D. Murch, and J. A. Nossek, “Joint Transmit and Receive Multi-user MIMO Decomposition Approach for the Downlink of Multi-user MIMO Systems,” in *Proceedings of the IEEE 58th Vehicular Technology Conference*. Orlando, FL: IEEE, October 6-9 2003.

[135] M. T. Ivrlač, R. L.-U. Choi, R. D. Murch, and J. A. Nossek, “Effective Use of Long-term Transmit Channel State Information in Multi-user MIMO Communication Systems,” in *Proceedings of the IEEE 58th Vehicular Technology Conference*. Orlando, FL: IEEE, October 6-9 2003.

[136] K.-K. Wong, “Adaptive Space-Division-Multiplexing and Bit-and-Power Allocation in Multiuser MIMO Flat Fading Broadcast Channels,” in *Proceedings of the IEEE 58th Vehicular Technology Conference*. Orlando, FL: IEEE, October 6-9 2003.