Transmitter Design for the Broadcast Channel in the MISO Wireless Communication

TRANSMITTER DESIGN FOR THE BROADCAST CHANNEL IN THE MISO WIRELESS COMMUNICATION

By

HAIBO WANG, B.SC., M.SC.

SEPTEMBER 2006

A Thesis

Submitted to the Department of Electrical & Computer Engineering

and the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree of

Master of Applied Science

McMaster University

MASTER OF APPLIED SCIENCE (2006) McMaster University
(Electrical & Computer Engineering) Hamilton, Ontario

TITLE:	Transmitter Design for the Broadcast Channel in the MISO Wireless Communication
AUTHOR:	Haibo Wang B.Sc. (Electronics and Communications Engineering), M.Sc. (Communications Engineering) Harbin Institute of Technology, Harbin, China
SUPERVISOR:	Thia Kirubarajan, Professor
NUMBER OF PAGES:	1, 60

Dedications

To my parents,

my wife.

Abstract

There are two popular approaches in the communication between multiple receivers and a base station with multiple antennas: dirty paper coding and multiuser diversity. Dirty paper coding can be rather difficult to realize, which motivates people to find some practical schemes. When there are a lot of users, multiuser diversity requires a lot of feedback which decrease the uplink spectrum efficiency.

In this paper, we aim to minimize the probability of error subject to the total transmit power constraint and decrease the amount of feedback required by the multiuser diversity instead of trying to achieve the dirty paper coding. There are two main results in this thesis: First, we formulate the minimization of the average probability of error of all the users as a convex optimization problem, subject to the peak or the average power constraints. The proposed transmitter represents a nonlinear one-to-one mapping between the transmitted data vector and the symbol vector. The transmitted data vector going through the base station antennas is obtained as a solution to the proposed convex error probability optimization problem that can be solved using computationally efficient interior point algorithms. Furthermore, we propose a random unitary beamforming technique to reduce the feedback by selecting a threshold for the users. To improve fairness, an equal ratio scheduling algorithm which could serve the users with different rate requirements is developed. We also give an upper and lower bound on the sum rate achievable in our approach. Monte Carlo simulation results is provided to verify the performance of the proposed algorithms.

Acknowledgements

I'm very lucky to have such a nice supervisor, Dr. Thia Kirubarajan. It is absolutely impossible to express my deep thankfulness to him with several words. His patience, enthusiasm, support, encouragement and wise supervision are essential to the completion of this thesis.

I also want to express my sincere gratitude to Ms. Cheryl Gies. It is her careful daily service and help that saves me from a lot of trouble.

I am also very grateful to Dr. Jiankang Zhang, Dr. Aboulnasr Hassanien, Dr. Amr El-Keyi and Dr. Ramy Gohary for their expertise and insightful discussions on my research and preparation of this thesis. At the same time, I would like to thank the staff in the department, in particular, Helen Jachna whose efficient works impresses me a lot. Last but not least, I would like to thank my family and my wife for their love and support which played an instrumental role in the completion of this project.

Contents

Chapter 1

Introduction

In the past twenty years, wireless communication has become a hot field both in the business and in the technology. While the objective of the first and second generation wireless communication focuses on the voice service, the third and fourth generation wireless communication is supposed to provide multimedia and high-rate data service, which require up to a few Mbps. A distinctive characteristic of the high-rate service, such as web-browsing, file transfer protocol, is the bi-direction asymmetric traffic between the uplink and the downlink. Consequently, it is very natural to increase the throughput of the downlink channel in the cellular communication system.

In the conventional single-input-single-output (SISO) wireless channel, to improve the throughput, we need more frequency spectrum or higher SNR. However, frequency spectrum is limited and expensive. On the other hand, SNR is also limited within some range because of the interference among the users and the cells. Therefore, more efficient modulation and demodulation methods is in practical demand.

1.1 Multiple Antennas

One of the exciting breakthrough of wireless communication in the past decade is the introduction of multiple antennas, which contribute both diversity gain and multiplexing gain to the fading channel. Multiple antennas have long been used in radar, sonar, and signal processing to extract the desired signal. It is well known that multiple antennas could improve the signal-to-noise ratio (SNR) of the received signal by carefully combining the received signal copies sampled from each antenna. Not until the sparkling work by Telatar [28] and Foschini [9], did it become apparent that the ergodic capacity of the multi-antenna channel increases linearly with the minimum of the number of transmit and receive antennas even if the transmitter does not know the channel.

The behavior of the multiple antenna channel is different from that of the single antenna channel. Traditionally, fading is distasteful as it reduces the signal amplitude. While in the multiple antenna channel, fading is beneficial since the variation of the channel result in the uncorrelated of the different channel coefficients, which give rise to a matrix channel whose rank is greater than one with high probability. This matrix channel is often called multiple-input-multiple-output (MIMO) channel. The rank of this matrix channel account for the multiplexing gain coming from the multiple antennas [39].

1.2 Transmitter Design in the Downlink Channel

In the traditional multi-access methods, such as frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA), each user is assigned a unique frequency slot, time slot or pseudorandom sequence to be identified by the base station (BS). When there are multiple antennas, space division multiple access (SDMA), can be used where the users are identified by their channel vector.

One of the most intensively studied MIMO channels is the MIMO broadcast channel [4, 38, 31, 30]. It is shown that the capacity of the MIMO broadcast channel is equivalent to that of the reverse multi-access channel, which is a generalization of the Costa's work [6].

There are three established approaches to MIMO broadcast channel transmitter design. The first is to minimize the total transmitted power subject to individual signal-to-interference-plus-noise ratio (SINR) constraints [19, 1, 24, 2]. The second is to optimize the system performance subject to the total transmitted power constraint[1] [23,14,26]. The last approach to MIMO broadcast channel transmitter design is to diagonalize (or block-diagonalize) the channel to enable efficient interference suppression [27,37,5].

From a capacity viewpoint, receiver and/or transmitter cooperation is necessary for multi-access and broadcast channels to achieve it [38]. Furthermore, recent capacity results on MIMO broadcast channel [4] suggest that instead of trying to suppress the interference and encoding each user independently, the BS should utilize the interference and jointly encode all the users. Obviously, linear BS transmitters do not fully utilize transmitter cooperation as they regard the signals destined for other users as interference and try to suppress this interference as much as possible. Since the optimal multiuser detector for CDMA multi-access channels decodes all the users jointly and in a non-linear way [29], we can expect that in order to achieve the optimal probability of error, a nonlinear transmitter (rather than a linear one) and a joint encoder (rather than an independent one) should be used.

[1]Note that the work in [14] originally considers a MIMO multi-access (MAC) channel and that it can be generalized to the MIMO broadcast channel case using the uplink-downlink duality property [31].

1.3 Multiuser Diversity

One of the basic properties of wireless channel is fading, which comes from the summation of multiple channel coefficient. To combat fading, diversity is widely used in coding and modulation. If there are many users in a wireless communication system, the natural variations of the wireless channel of different users introduce another kind of multiuser diversity.

Transmitting to the best user has been shown to greatly improve the capacity of time-division-multiple-access (TDMA) uplink fading channel when there is only one antenna at the BS [15]. This benefit, termed as multiuser diversity in [32], is due to the independent time-varying channels across the different users. When there are multiple antennas at the BS [32], an opportunistic beamforming technique to increase the throughput of the slowly time-varying channel by inducing faster and larger fluctuations was proposed.

Assuming that there are K single-antenna users and M antennas at the BS, for full channel state information at the BS, the users are required to feed back $2KM$ real numbers after each training, which would be a substantially large number if K is large. This large amount of feedback prohibits practical applications of multiuser diversity if there are a lot of users. In this thesis, we follow the work of [25] and propose an algorithm aimed at reducing the necessary amount of feedback by finding M or less than M quasi-orthogonal users. In our approach, the feedback process is realized in two stages. The first stage determines which beams will be used. In the second stage, the total power is uniformly allocated to the selected beams and the quasi-orthogonal users feed back their SINR to the BS. A user feeds back only if the maximum of the normalized cross-correlations of its channel with the beams is greater than a given threshold. The threshold is deliberately devised so that there are on average d users feeding back at a time, where d is a fixed number. It turns out that the channels of those users who feed back approach orthogonality as K increases. Simulation results demonstrate that the proposed random beamforming has roughly the same performance as the previous techniques. However, the average amount of feedback required in the proposed random beamforming is only d real numbers and d integers, which does not vary with K.

1.4 Dissertation Outline

The thesis is organized as follows. In chapter 2, we formulate the problem of minimizing the probability of error as a convex optimization problem that can be efficiently solved using interior point methods. In chapter 3, we propose a random unitary beamforming to decrease the amount of feedback. The conclusions and future work are reported in chapter 4.

Chapter 2

Convex optimization of error probability

In this chapter, we will introduce the popular approaches to the BS transmitter design subject to total power constraint. We also propose an nonlinear transmitter design method which minimizes the probability of error using convex optimization.

2.1 BS Transmitter design

Previous work on the BS transmitter design under the total transmitted power constraint includes [23], where two iterative algorithms are presented to solve SINR balancing and power minimization using downlink beamforming, and [14], where a minimum mean-square-error (MMSE) based transmitter design approach with different power constraints has been studied. In [26], a duality between the normalized MSE region of the uplink and that of the downlink is given. Furthermore, a closed-form solution to maximize a lower bound of the product of SINRs under the total power constraint is provided in [36], and a non-linear vector perturbation-based technique that approaches the capacity of MIMO broadcast channel is proposed in [11].

Another popular non-linear technique is the Tomlinson-Harashima precoding [35]. The last two nonlinear techniques require that the average power of the transmitted signal vector satisfy the total transmit power constraint. Unfortunately, this condition is hard to satisfy because it is difficult to compute this average power. This is especially true if there is a time-varying near-far effect or if the channel distribution is unknown.

From a capacity viewpoint, receiver and/or transmitter cooperation is necessary for multi-access and broadcast channels to achieve it [38]. Furthermore, recent capacity results on MIMO broadcast channel [4] suggest that instead of trying to suppress the interference and encoding each user independently, the BS should utilize the interference and jointly encode all the users. Obviously, linear[1] BS transmitters do not fully utilize transmitter cooperation as they regard the signals destined for other users as interference and try to suppress this interference as much as possible. Since the optimal multiuser detector for CDMA multi-access channels decodes all the users jointly and in a non-linear way [29], we can expect that in order to achieve the optimal probability of error, a nonlinear transmitter (rather than a linear one) and a joint encoder (rather than an independent one) should be used.

In this chapter, we consider the problem of transmitter design in the multi-input-multi-output broadcast channel where the average probability of error is minimized subject to the peak or the average power constraints. We assume a simple receiver structure in each mobile user that does not need any channel state information (CSI). For each symbol vector in the BS, we select a particular data vector transmitted through the BS antennas to minimize the probability of error. We formulate the problem of minimizing the probability of error as a convex optimization problem that can be efficiently solved using interior point methods. As a by-product, the

[1]Here "linear" means those transmitters that assign a precoding (beamforming) vector to each user.

exact probability of error is obtained after solving this optimization problem. It is proved that the proposed transmitter is optimal unless the signal-to-noise ratio (SNR) is very low or the channel is nearly singular. Simulation results demonstrate that the proposed transmission scheme significantly improves the probability of error as compared to several earlier approaches.

2.2 Convex Optimization

In this section, we will briefly introduce the convex sets and convex optimization problems.

Suppose that C is a set in R^n. $\mathbf{x}_1, \mathbf{x}_2$ are any two points in C and $\mathbf{x}_1 \neq \mathbf{x}_2$. $\forall\, 0 \leq \theta \leq 1$, let

$$\mathbf{y} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$$

if $\mathbf{y}$ is also in C, then C is a convex set.

A function $f : R^n \rightarrow R$ is convex if $\mathbf{dom} f$ is a convex set and if for any $\mathbf{x}, \mathbf{y} \in \mathbf{dom} f$, and θ with $0 \leq \theta \leq 1$, we have

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

Let

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ & f_i(\mathbf{x}) \leq 0, \; i = 1, 2, \cdots, m \\ & h_i(\mathbf{x}) = 0, \; i = 1, 2, \cdots, n \end{aligned} \tag{2.1}$$

describe the problem of find $\mathbf{x}$ that minimizes the function $f_0(\mathbf{x})$ under the conditions that $f_i(\mathbf{x}) \leq 0, \; i = 1, 2, \cdots, m$ and $h_i(\mathbf{x}) = 0, \; i = 1, 2, \cdots, n$. We call $\mathbf{x}$ the *optimization variable* and the function $f_0 : R^n \rightarrow R$ the *objective function* or *cost* *function.* The equations $h_i(\mathbf{x}) = 0$ are called *equality constraints.* If there is no constraint, we say that (2.1) is unconstrained.

The set of all points for which the objective and all constraint function are defined, i.e.

$$D = \bigcap_{i=0}^{m} \mathbf{dom} f_i \quad \cap \quad \bigcap_{i=1}^{n} \mathbf{dom} h_i$$

is called the *domain* of the optimization problem (2.1). A point $\mathbf{x}$ is *feasible* if it satisfies all the constraint $f_i(\mathbf{x}) \leq 0$ and $h_i(\mathbf{x}) = 0$. The problem in (2.1) is said to be *feasible* if there exists at least one feasible point, and *infeasible* otherwise. The set of all feasible points is called the *feasible set* or the *constraint set.*

The *optimal value* f^* of the problem (2.1) is defined as

$$f^* = \inf\{f_0(\mathbf{x}) : \; f_i(\mathbf{x}) \leq 0, \; i = 1, 2, \cdots, m, \; h_i(\mathbf{x}) = 0, \; i = 1, 2, \cdots, n\}$$

We say that $\mathbf{x}^*$ is the *optimal point*, or solves the problem (2.1), if $\mathbf{x}^*$ is feasible and $f_0(\mathbf{x}^*) = f^*$.

A *convex optimization problem* is one of the form

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ & f_i(\mathbf{x}) \leq 0, \; i = 1, 2, \cdots, m \\ & a_i^T \mathbf{x} = b_i, \; i = 1, 2, \cdots, n \end{aligned} \tag{2.2}$$

where $f_0, f_1, \vdots, f_m$ are convex functions. An efficient method to solve a convex optimization problem is interior point algorithm. For details about convex optimization and interior point algorithm, please refer to [3].

2.3 System Model

We consider a single-cell MIMO broadcast channel with M antennas at the BS and K mobile users having one antenna per user. Assuming flat block-fading, the channel is given by a $K \times M$ matrix $\mathbf{H} = [h_{ij}] \in \mathcal{C}^{K \times M}$, where h_{ij} is the path gain from the jth antenna at the BS to the ith user. We assume that $\mathbf{H}$ a is full row-rank matrix and K is less than or equal to M. Perfect channel state information (CSI) is assumed at the transmitter.

Let the vector $\mathbf{y} = [y_1 \; y_2 \; \cdots \; y_K]^T$ combine the signals received by all K users where $(\cdot)^T$ denotes the transpose. Then,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.3}$$

where $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_M]^T$ is the vector of signals transmitted from the BS, $\mathbf{n} = [n_1 \; n_2 \; \cdots \; n_K]^T \sim \mathcal{C}N(\mathbf{0}, \mathbf{I})$ is the vector of zero-mean i.i.d. unit-variance complex Gaussian noise, $\mathbf{I}$ is the identity matrix, and $\mathcal{C}N(\cdot, \cdot)$ denotes the complex Gaussian distribution. It should be stressed here that $\mathbf{x}$ is not the vector of original information symbols that have to be sent to the users, but the vector obtained from the above-mentioned vector of information symbols by means of a certain linear or non-linear mapping. We assume the total transmit power at the BS to be P. Hence, the peak and average power constraints can be written as

$$\mathbf{x}^H \mathbf{x} \leq P \tag{2.4}$$

$$\mathrm{E}\{\mathbf{x}^H \mathbf{x}\} \leq P \tag{2.5}$$

respectively, where $\mathrm{E}\{\cdot\}$ is the statistical expectation operator. Let $\mathbf{s} = [s_1 \; s_2 \; \cdots \; s_K]^T$, where s_k denotes the i.i.d information symbol of the kth user. These symbols are assumed to be uncorrelated with the noise. In this chapter, we assume that each user exploits 4-QAM signals $\frac{1}{\sqrt{2}}(\pm 1 \pm j)$ and that the conventional $\mathrm{sgn}(\cdot)$ detector

$$\hat{s}_k = \frac{1}{\sqrt{2}} \left(\mathrm{sgn}\Big(\Re(y_k) \Big) + j\, \mathrm{sgn}\Big(\Im(y_k) \Big) \right) \tag{2.6}$$

is used at the receiver, where $\hat{s}_k$ is the estimate of the symbols of the kth user and $j = \sqrt{-1}$. $\Re(\cdot)$ and $\Im(\cdot)$ stand for the real and imaginary parts, respectively.

2.4 Transmitter Design under Total MMSE Criterion

One popular approach to optimize system performance in a multi-access channel subject to the average power constraint is to minimize the total MMSE of all users [14]. In this section, we extend this approach to the MIMO broadcast channel case. Let p_k denote the power allocated to the kth user and $\mathbf{A} = \mathrm{diag}\{\sqrt{p_1}\ \sqrt{p_2}\ \cdots\ \sqrt{p_K}\}$. Assuming that linear beamforming (precoding) is used at the BS, $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{UAs}$$

where $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_K]$ is the beamforming matrix and $\mathbf{u}_k$ is the normalized beamforming vector for the kth user, i.e., $\|\mathbf{u}_k\| = 1$. $\|\cdot\|$ denotes ℓ^2-norm. The received signal of the kth user can be written as

$$y_k = \mathbf{h}_k^T\mathbf{x} + n_k = \mathbf{h}_k^T\mathbf{UAs} + n_k$$

where $\mathbf{h}_k^T$ is the kth row of $\mathbf{H}$. Then, the received SINR of the kth user is given by

$$\mathrm{SINR}_k = \frac{p_k\mathbf{h}_k^T\mathbf{u}_k\mathbf{u}_k^H\mathbf{h}_k^*}{\sum_{i\neq k} p_i\mathbf{h}_k^T\mathbf{u}_i\mathbf{u}_i^H\mathbf{h}_k^* + 1} \tag{2.7}$$

where $(\cdot)^*$ denotes complex conjugate. The mean square error (MSE) of the kth user can be expressed as

$$\begin{aligned}
\mathrm{MSE}_k &= \mathrm{E}\{|c_k y_k - s_k|^2\} \\
&= |c_k|^2\left(\sum_i p_i\mathbf{h}_k^T\mathbf{u}_i\mathbf{u}_i^H\mathbf{h}_k^* + 1\right) - 2\Re(c_k\sqrt{p_k}\mathbf{h}_k^T\mathbf{u}_k) + 1 \\
&= \left|c_k\sqrt{\sum_i p_i\mathbf{h}_k^T\mathbf{u}_i\mathbf{u}_i^H\mathbf{h}_k^* + 1} - \frac{\sqrt{p_k}\mathbf{h}_k^H\mathbf{u}_k^*}{\sqrt{\sum_i p_i\mathbf{h}_k^T\mathbf{u}_i\mathbf{u}_i^H\mathbf{h}_k^* + 1}}\right|^2 + 1 - \\
&\quad \frac{p_k\mathbf{h}_k^T\mathbf{u}_k\mathbf{u}_k^H\mathbf{h}_k^*}{\sum_i p_i\mathbf{h}_k^T\mathbf{u}_i\mathbf{u}_i^H\mathbf{h}_k^* + 1}
\end{aligned} \tag{2.8}$$

where c_k is the scaling factor at the receiver. Note that the previous MMSE-based transmitter designs in [33, 13, 20, 21, 12] regard this scaling factor as a fixed number, which inevitably results in performance loss. The MSE in (2.8) achieves its minimum

$$\text{MMSE}_k = 1 - \frac{p_k \mathbf{h}_k^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{h}_k^*}{\sum_i p_i \mathbf{h}_k^T \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_k^* + 1} \tag{2.9}$$

when

$$c_k = \frac{\sqrt{p_k} \mathbf{h}_k^H \mathbf{u}_k^*}{\sum_i p_i \mathbf{h}_k^T \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_k^* + 1}$$

Using (2.9), the total MMSE can be expressed as

$$\text{MMSE}_{\text{total}} = \sum_{k=1}^{K} \text{MMSE}_k = K - \sum_{k=1}^{K} \frac{p_k \mathbf{h}_k^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{h}_k^*}{\sum_i p_i \mathbf{h}_k^T \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_k^* + 1}$$

Now, our objective is to minimize the total MMSE over $\mathbf{u}_k$ and p_k subject to the average power constraint. This problem can be written as

$$\min_{\mathbf{u_k}, p_k} -\sum_{k=1}^{K} \frac{p_k \mathbf{h}_k^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{h}_k^*}{\sum_i p_i \mathbf{h}_k^T \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_k^* + 1} \quad \text{s.t.}\ \|\mathbf{u}_k\| = 1, k = 1, 2, \cdots, K, \sum_{k=1}^{K} p_k \leq P \tag{2.10}$$

or, equivalently, as

$$\max_{\mathbf{u_k}, p_k} \sum_{k=1}^{K} \frac{p_k \mathbf{h}_k^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{h}_k^*}{\sum_i p_i \mathbf{h}_k^T \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_k^* + 1} \quad \text{s.t.}\ \|\mathbf{u}_k\| = 1, k = 1, 2, \cdots, K, \sum_{k=1}^{K} p_k \leq P \tag{2.11}$$

Note that the objective functions in (2.10) and (2.11) are not convex. Although iterative algorithms may be designed to solve these problems, their convergence to the global optimum is not guaranteed. Fortunately, according to the uplink-downlink duality property [31], the achievable SINR region in the MIMO broadcast channel is the same as the one in the reciprocal MIMO multi-access channel case, provided that the same average power constraint is used in both cases. Furthermore, the normalized receive beamforming vector of each user in the MIMO multi-access channel is equal to the corresponding normalized transmit beamforming vector in the MIMO broadcast channel. From (2.7) and (2.9), we see that the MMSE of the kth user is related to its

SINR in the broadcast channel (for the multi-access channel case, see [29]). Therefore, if we solve the similar minimization of the total MMSE in the multi-access channel, we will get the normalized transmit beamforming vector $\mathbf{u}_k$ and SINR_k in the MIMO broadcast channel. Then we only need to find the power of each user p_k, which is straightforward using the following linear equations

$$\text{SINR}_k = \frac{p_k \mathbf{h}_k^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{h}_k^*}{\sum_{i \neq k} p_i \mathbf{h}_k^T \mathbf{u}_i \mathbf{u}_i^H \mathbf{h}_k^* + 1}, \ k = 1, 2, \cdots, K$$

because SINR_k, $\mathbf{h}_k$ and $\mathbf{u}_k$ are known. Consider the reciprocal multi-access channel with K users each with one antenna and M antennas at the BS

$$\mathbf{y}_{\text{mac}} = \mathbf{H}^H \mathbf{x}_{\text{mac}} + \mathbf{w}$$

where $\mathbf{y}_{\text{mac}}$, $\mathbf{x}_{\text{mac}}$ and $\mathbf{w}$ are the received signal vector, transmit signal vector and the noise vector, respectively. Let q_k denote the power of the kth user. The total MMSE of all the users is given by [14]

$$\text{MMSE}_{\text{total-mac}} = K - M + \text{tr}((\mathbf{I} + \mathbf{H}^H \mathbf{Q} \mathbf{H})^{-1})$$

where $\mathbf{Q} = \text{diag}\{q_1 \ q_2 \ \cdots \ q_K\}$. In order to minimize the total MMSE, we need to solve

$$\min_{\mathbf{Q}} \ \text{tr}((\mathbf{I} + \mathbf{H}^H \mathbf{Q} \mathbf{H})^{-1}) \quad \text{s. t. } \text{tr}(\mathbf{Q}) \leq P, \mathbf{Q} \succeq 0 \tag{2.12}$$

This problem is proposed in [14] and can be readily shown to be convex. Therefore, (2.10) can be transformed into a convex optimization problem in the reciprocal multi-access channel and solved efficiently using interior point methods. A similar result is given in [26, 16].

From (2.7) and (2.9), it follows that the MMSE of the kth user is related to its SINR as (also see [29])

$$\text{SINR}_k = \frac{1}{\text{MMSE}_k} - 1 \tag{2.13}$$

2.5 Nonlinear Transmitter for MIMO Broadcast Channel

The optimum multiuser detector for MIMO multi-access channel is a nonlinear detector which selects the symbol vector with the minimum Euclidean distance. Also there exists an uplink-downlink duality between the linear transmitter and the receiver. Hence, we conjecture that in order to achieve the optimum performance a nonlinear transmitter should be used instead of a linear one for MIMO broadcast channel. In this section we propose a nonlinear transmitter and formulate the probability of error minimization as a convex optimization problem.

For the sake of convenience of our subsequent derivations, let us convert the complex channel model (2.3) into an equivalent real-valued model. Combining the real and imaginary parts of $\mathbf{y}$ in one vector, we obtain

$$\underline{\mathbf{y}} = \underline{\underline{\mathbf{H}}}\,\underline{\mathbf{x}} + \underline{\mathbf{n}} \tag{2.14}$$

where

$$\underline{\mathbf{y}} = \left[\Re(\mathbf{y})^T \; \Im(\mathbf{y})^T\right]^T, \qquad \underline{\mathbf{x}} = \left[\Re(\mathbf{x})^T \; \Im(\mathbf{x})^T\right]^T$$
$$\underline{\underline{\mathbf{H}}} = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}, \qquad \underline{\mathbf{n}} = \left[\Re(\mathbf{n})^T \; \Im(\mathbf{n})^T\right]^T$$

Note that $\underline{\mathbf{n}}$ is zero-mean i.i.d. real Gaussian noise with the variance 0.5 per entry. From (2.6), it follows that the model in (2.14) is equivalent to that in (2.3) provided that we demodulate $\Re(y_k)$ and $\Im(y_k)$ independently using the same sign detector and combine the estimates so obtained to compute our estimate of the 4-QAM symbol.

Let us define the $2K \times 1$ bit vector as $\mathbf{b} = \sqrt{2}\left[\Re(\mathbf{s})^T \; \Im(\mathbf{s})^T\right]^T$ and introduce the matrix

$$\mathbf{B} = [\mathbf{b}_1 \; \mathbf{b}_2 \; \cdots \; \mathbf{b}_{2^{2K}}] \tag{2.15}$$

which captures all possible 2^{2K} realizations $\mathbf{b}_i$ $(i = 1, 2, \ldots, 2^{2K})$ of vector $\mathbf{b}$. We assume that the vectors $\mathbf{b}_i$ in (2.15) are ordered so that the left half of $\mathbf{B}$ is symmetric to the right half of this matrix multiplied by -1.[2] This specific structure of $\mathbf{B}$ will be used in what follows to simplify the proposed transmitter designs.

For any $\mathbf{b} \in \mathcal{S} \triangleq \{\mathbf{b}_1, \ldots, \mathbf{b}_{2^{2K}}\}$, we obtain the vector $\underline{\mathbf{x}}$ to be transmitted from the BS as

$$\underline{\mathbf{x}} = f(\mathbf{b}) \tag{2.16}$$

where $f(\cdot)$ is the encoding function that provides a one-to-one (generally nonlinear) mapping from $\mathcal{S}$ to 2^{2K} discrete points in the $2M$-dimensional Euclidean space. Hence, the transmitter design problem amounts to obtaining this function, i.e., finding $\underline{\mathbf{x}}$ for any given $\mathbf{b}$.

2.5.1 Peak Power Constraint

It can be readily obtained from (2.4) that the peak power is limited in each symbol period as $\underline{\mathbf{x}}^T\underline{\mathbf{x}} \leq P$. According to (2.14), the real and imaginary parts of the received signal of the kth user are given by

$$\underline{y}_k = \underline{\mathbf{h}}_k^T\underline{\mathbf{x}} + \underline{n}_k, \quad \underline{y}_{k+K} = \underline{\mathbf{h}}_{k+K}^T\underline{\mathbf{x}} + \underline{n}_{k+K}$$

respectively. Here, $\underline{y}_k$, $\underline{\mathbf{h}}_k^T$ and $\underline{n}_k$ are the kth element of $\underline{\mathbf{y}}$, the kth row of $\underline{\mathbf{H}}$ and the kth element of $\underline{\mathbf{n}}$, respectively. Each user decodes its bits by means of the conventional sign detector as

$$\hat{b}_k = \operatorname{sgn}(\underline{y}_k) \tag{2.17}$$

[2]It can be easily proved that it is always possible to satisfy this property provided that the columns of $\mathbf{B}$ are properly ordered.

where $\hat{b}_k$ is the estimate of the kth element of $\mathbf{b}$. The probability of erroneously detecting b_k conditioned on that $\mathbf{b}$ being the true bit vector can be computed as

$$\begin{aligned} P_{e_k}^{\mathbf{b}} &= P(\hat{b}_k \neq b_k) = P(b_k \underline{y}_k \leq 0) \\ &= P(b_k(\underline{\mathbf{h}}_k^T \underline{\mathbf{x}} + \underline{n}_k) \leq 0) = \begin{cases} Q(\sqrt{2} b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}}), & b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} \geq 0 \\ 1 - Q(-\sqrt{2} b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}}), & b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} < 0 \end{cases} \end{aligned} \tag{2.18}$$

where $Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-\frac{\eta^2}{2}} \, d\eta$, $t \geq 0$. The second-order derivative of $Q(t)$ is $Q''(t) = \frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}}$. Therefore, $Q(t)$ is convex provided that $t \geq 0$. The probability of error conditioned on $\mathbf{b}$ being the true bit vector and averaged over all the users can be computed as

$$P_e^{\mathbf{b}} = \frac{1}{2K} \sum_{k=1}^{2K} P_{e_k}^{\mathbf{b}} \tag{2.19}$$

where $P_{e_k}^{\mathbf{b}}$ is defined in (2.18). Using (2.19) as the objective function, the transmitter design problem can be formulated as

$$\min_{\underline{\mathbf{x}}} \frac{1}{2K} \sum_{k=1}^{2K} P_{e_k}^{\mathbf{b}} \qquad \text{s.t.} \qquad \underline{\mathbf{x}}^T \underline{\mathbf{x}} \leq P \tag{2.20}$$

Obviously, it is very difficult to solve (2.20) directly. Therefore, let us try to simplify this problem using some properties of $P_{e_k}^{\mathbf{b}}$. As $Q(t)$ is convex for $t \geq 0$, from (2.18) we see that $P_{e_k}^{\mathbf{b}}$ is convex if $b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} \geq 0$, and concave otherwise. It is important to stress that if $b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} < 0$, then $P_{e_k}^{\mathbf{b}} > \frac{1}{2}$ and $P_e^{\mathbf{b}} > \frac{1}{4K}$, which is obviously undesirable. Therefore, we add to the problem (2.20) the additional constraint $b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} \geq 0$ and obtain the following convex optimization problem:

$$\begin{aligned} &\min_{\underline{\mathbf{x}}} \frac{1}{2K} \sum_{k=1}^{2K} Q(\sqrt{2} b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}}) \\ &\text{s.t.} \ \ \underline{\mathbf{x}}^T \underline{\mathbf{x}} \leq P, \quad b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} \geq 0, k = 1, \ldots, 2K \end{aligned} \tag{2.21}$$

Now, we can interpret $f(\cdot)$ in (2.16) as a mapping of $\mathbf{b}$ to the optimal point of (2.21). As the left half of $\mathbf{B}$ is symmetric to its right half, we only need to find the optimal

$\underline{\mathbf{x}}$ for the vectors $\mathbf{b}_i$ contained in the left half of $\mathbf{B}$. A convex approach to multiuser detection in DS-CDMA using the convexity of $Q(t)$ is also proposed in [34].

The optimal point of (2.21) is not necessarily globally optimal as we impose the constraint $b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}} \geq 0$. However, the following theorem which shows that the optimal solution of (2.21) also solves (2.20) in most practically important cases (unless the optimal probability of error is very high, which means that the SNR/transmit power is very low or the channel $\underline{\mathbf{H}}$ is nearly singular) can be used.

Theorem 2.1 *The optimal point of (2.21) solves (2.20) if $(P_e^b)^{\text{opt}} \leq \frac{1}{4K}$, where $(P_e^b)^{\text{opt}}$ is the optimal value of (2.20). Furthermore, if the optimal value of (2.21) is less than or equal to $\frac{1}{4K}$, it is equal to $(P_e^b)^{\text{opt}}$.*

Proof. If $(P_e^{\mathbf{b}})^{\text{opt}} \leq \frac{1}{4K}$, then $(P_{e_k}^{\mathbf{b}})^{\text{opt}}$ must be less than or equal to $\frac{1}{2}$ and $b_k \underline{\mathbf{h}}_k^T \underline{\mathbf{x}}^{\text{opt}}$ must be greater than or equal to zero for all k, where $\underline{\mathbf{x}}^{\text{opt}}$ is the optimal point of (2.20). $(P_{e_k}^{\mathbf{b}})^{\text{opt}}$ is computed from $\underline{\mathbf{x}}^{\text{opt}}$ using (2.18). Therefore, $\underline{\mathbf{x}}^{\text{opt}}$ lies in the feasible set of (2.21). As the feasible set of (2.20) includes that of (2.21), $\underline{\mathbf{x}}^{\text{opt}}$ also solves (2.21), i.e., the optimal point of (2.21) solves (2.20).

Next, the optimal value of (2.21) is an upper bound of $(P_e^b)^{\text{opt}}$ since it has a smaller feasible set. Therefore if the optimal value of (2.21) is less than or equal to $\frac{1}{4K}$, $(P_e^b)^{\text{opt}}$ is also less than or equal to $\frac{1}{4K}$. Combining this with the first statement of Theorem 1, we conclude that the optimal value of (2.21) is equal to that of (2.20) if it is less than or equal to $\frac{1}{4K}$.

The probability of error $P_e^{\mathbf{b}}$ is conditioned on that $\mathbf{b}$ being the true bit vector. The unconditioned probability of error can be expressed as

$$P_e = \frac{1}{2^{2K}} \sum_{\mathbf{b} \in \mathcal{S}} (P_e^{\mathbf{b}})^{\text{opt}} \tag{2.22}$$

where $(P_e^{\mathbf{b}})^{\text{opt}}$ is averaged over all possible realizations of the vector $\mathbf{b}$.

2.5.2 Average Power Constraint

Now, let us consider the average power constraint that can be expressed as

$$\mathrm{E}\{\underline{\mathbf{x}}^T\underline{\mathbf{x}}\} = \frac{1}{2^{2K}} \sum_{i=1}^{2^{2K}} \underline{\mathbf{x}}_i^T \underline{\mathbf{x}}_i \leq P \tag{2.23}$$

where $\underline{\mathbf{x}}_i$ is the selected BS transmit signal vector corresponding to bit vector $\mathbf{b}_i$. Similar to the peak power case, we use the nonlinear transmitter given by (2.16).

As mentioned in the previous section, we only need to find the optimal $\underline{\mathbf{x}}_i$ for the vectors $\mathbf{b}_i$ contained in the left half of matrix $\mathbf{B}$. It should be noted here that for any $\mathbf{b}$, the error probability only depends on $b_k\underline{\mathbf{h}}_k^T\underline{\mathbf{x}}$. Let us use the following notations:

$$\begin{aligned}
\overline{\mathbf{b}} &= \left[\begin{array}{cccc} \mathbf{b}_1^T & \mathbf{b}_2^T & \cdots & \mathbf{b}_{2^{2K-1}}^T \end{array}\right]^T \\
\overline{\mathbf{x}} &= \left[\begin{array}{cccc} \underline{\mathbf{x}}_1^T & \underline{\mathbf{x}}_2^T & \cdots & \underline{\mathbf{x}}_{2^{2K-1}}^T \end{array}\right]^T \\
\overline{\mathbf{n}} &= \left[\begin{array}{cccc} \underline{\mathbf{n}}_1^T & \underline{\mathbf{n}}_2^T & \cdots & \underline{\mathbf{n}}_{2^{2K-1}}^T \end{array}\right]^T \\
\overline{\mathbf{H}} &= \mathbf{I}_{2^{2K-1}} \otimes \underline{\mathbf{H}} \\
\overline{\mathbf{y}} &= \left[\begin{array}{cccc} \underline{\mathbf{y}}_1^T & \underline{\mathbf{y}}_2^T & \cdots & \underline{\mathbf{y}}_{2^{2K-1}}^T \end{array}\right]^T \\
\overline{\mathbf{y}} &= \overline{\mathbf{H}}\overline{\mathbf{x}} + \overline{\mathbf{n}} \\
\underline{\mathbf{y}}_i &= \underline{\mathbf{H}}\,\underline{\mathbf{x}}_i + \underline{\mathbf{n}}_i, \quad i = 1, \ldots, 2^{2K-1}
\end{aligned}$$

where $\otimes$ denotes the Kronecker product, while $\underline{\mathbf{y}}_i$ and $\underline{\mathbf{n}}_i$ are the received signal and the noise vectors corresponding to $\mathbf{b}_i$, respectively. When using the average power constraint, we have to further modify the objective function with respect to (2.20) by averaging it over all possible vectors $\mathbf{b}_i$ that are contained in the left half of matrix $\mathbf{B}$. Hence, the following minimization problem has to be solved:

$$\min_{\overline{\mathbf{x}}} \frac{1}{K2^{2K}} \sum_{k=1}^{K2^{2K}} P_{e_k}^{\overline{\mathbf{b}}} \qquad \text{s.t.} \qquad \overline{\mathbf{x}}^T\overline{\mathbf{x}} \leq P2^{2K-1} \tag{2.24}$$

where $P_{e_k}^{\overline{\mathbf{b}}}$ is defined similar to (2.18). Similar to the peak power constraint case, we can approximate this problem by

$$\begin{aligned} \min_{\overline{\mathbf{x}}} \quad & \frac{1}{K2^{2K}} \sum_{k=1}^{K2^{2K}} Q(\sqrt{2}\bar{b}_k \overline{\mathbf{h}}_k^T \overline{\mathbf{x}}) \\ \text{s.t.} \quad & \overline{\mathbf{x}}^T \overline{\mathbf{x}} \leq P2^{2K-1},\ \bar{b}_k \overline{\mathbf{h}}_k^T \overline{\mathbf{x}} \geq 0, k = 1, \ldots, K2^{2K} \end{aligned} \tag{2.25}$$

where $Q(\sqrt{2}\bar{b}_k \overline{\mathbf{h}}_k^T \overline{\mathbf{x}})$ is defined similar to (2.18), and $\bar{b}_k$ and $\overline{\mathbf{h}}_k^T$ are the kth element of $\overline{\mathbf{b}}$ and the kth row of $\overline{\mathbf{H}}$, respectively.

Note that the power constraint in (2.25) is obtained by averaging the power constraint in (2.23). The optimal $\overline{\mathbf{x}}^{\text{opt}}$ of (2.25) is an $M2^{2K} \times 1$ vector, i.e., it combines optimal transmit signal vectors for all $\mathbf{b}_i$ contained in the left half of $\mathbf{B}$. Let us use the MATLAB notation to define the vector

$$\mathbf{z}_i = \overline{\mathbf{x}}^{\text{opt}}(2M(i-1)+1 : 2Mi) = [\overline{\mathbf{x}}^{\text{opt}}_{2M(i-1)+1}, \cdots, \overline{\mathbf{x}}^{\text{opt}}_{2Mi}]$$

Then the optimal nonlinear mapping function can be expressed as

$$f(\mathbf{b}_i) = \begin{cases} \mathbf{z}_i, & 1 \leq i \leq 2^{2K-1} \\ -\mathbf{z}_{2^{2K}-i+1}, & 2^{2K-1}+1 \leq i \leq 2^{2K} \end{cases} \tag{2.26}$$

From Theorem 1, we have that if the optimal value of (2.25) is less than or equal to $\frac{1}{K2^{2K+1}}$, then it is also the optimal value of (2.24).

One of the attractive properties of our convex formulation-based transmitter design approach is that we can easily add other convex constraints to the corresponding optimization problem. To illustrate this, we will show how to combine both the average and peak power constraints together or introduce individual error probability constraints.

Let P_{peak} denote the maximum peak power that the BS can support. Then, the peak power constraint can be added to (2.25) as follows:

$$
\begin{aligned}
\min_{\bar{\mathbf{x}}} \quad & \frac{1}{K2^{2K}} \sum_{k=1}^{K2^{2K}} Q(\sqrt{2}\bar{b}_k \bar{\mathbf{h}}_k^T \bar{\mathbf{x}}) \\
\text{s.t.} \quad & \bar{\mathbf{x}}^T \bar{\mathbf{x}} \leq P2^{2K-1}, \;\; \bar{b}_k \bar{\mathbf{h}}_k^T \bar{\mathbf{x}} \geq 0, \; k = 1, \ldots, K2^{2K} \\
& \underline{\mathbf{x}}_i^T \underline{\mathbf{x}}_i \leq P_{\text{peak}}, \;\; i = 1, \ldots, 2^{2K-1}
\end{aligned}
\tag{2.27}
$$

Let P_{e_j} denote the maximum acceptable error probability of the jth user. Then we introduce individual error probability constraints to (2.25) given by

$$
\begin{aligned}
\min_{\bar{\mathbf{x}}} \quad & \frac{1}{K2^{2K}} \sum_{k=1}^{K2^{2K}} Q(\sqrt{2}\bar{b}_k \bar{\mathbf{h}}_k^T \bar{\mathbf{x}}) \\
\text{s.t.} \quad & \bar{\mathbf{x}}^T \bar{\mathbf{x}} \leq P2^{2K-1} \\
& \bar{b}_k \bar{\mathbf{h}}_k^T \bar{\mathbf{x}} \geq 0, k = 1, 2, \cdots, K2^{2K} \\
& 2^{-2K} \sum_{m=1}^{2^{2K-1}} \sum_{n=0}^{1} Q(\sqrt{2}\bar{b}_{(m-1)2K+nK+j} \bar{\mathbf{h}}_{(m-1)2K+nK+j}^T \bar{\mathbf{x}}) \leq P_{e_j}, j = 1, 2, \cdots, K
\end{aligned}
\tag{2.28}
$$

where $2^{-2K} \sum_{m=1}^{2^{2K-1}} \sum_{n=0}^{1} Q(\sqrt{2}\bar{b}_{(m-1)2K+nK+j} \bar{\mathbf{h}}_{(m-1)2K+nK+j}^T \bar{\mathbf{x}})$ is the error probability of the jth user using the nonlinear transmitter design.

2.5.3 Optimality of Conventional Detector With Zero Threshold

While the above error probability minimization is formulated based on the assumption that a sign detector with zero threshold is used for each user, one may wonder whether we can obtain any performance gain if we assume that a sign detector with nonzero threshold is used at each user. In order to analyze the performance of this approach, we next assume that a sign detector with nonzero threshold is used at each user.

From (2.17), the sign detector with nonzero threshold is given by

$$
\hat{b}_k = \text{sgn}(\underline{y}_k - \gamma_k) \tag{2.29}
$$

where γ_k is the decision threshold of $\underline{y}_k$, which means that there are two different thresholds for the real and imaginary parts of y_k in (2.6). It should be emphasized that this approach cannot be applied to the peak power constraint case as γ_k is regarded as a variable in the optimization problem. Because there are 2^{2K-1} optimization problems for the peak power constraint case, we will obtain different γ_k from each optimization problem. However, it is impractical for each user to have different γ_k for each bit vector as they do not know the true bit vector sent by the BS. Therefore, in the following, the error probability minimization under the average power constraint is to be studied alone. Before that, we introduce the following notations to develop the error probability minimization problem. Note that in this case the optimal $\underline{\mathbf{x}}$ corresponding to $\mathbf{b}$ may not be equal to the optimal $-\underline{\mathbf{x}}$ corresponding to $-\mathbf{b}$.

Let

$$\begin{aligned}
\tilde{\mathbf{b}} &= \begin{bmatrix} \mathbf{b}_1^T & \mathbf{b}_2^T & \cdots & \mathbf{b}_{2^{2K}}^T \end{bmatrix}^T \\
\tilde{\mathbf{x}} &= \begin{bmatrix} \underline{\mathbf{x}}_1^T & \underline{\mathbf{x}}_2^T & \cdots & \underline{\mathbf{x}}_{2^{2K}}^T \end{bmatrix}^T \\
\tilde{\mathbf{n}} &= \begin{bmatrix} \underline{\mathbf{n}}_1^T & \underline{\mathbf{n}}_2^T & \cdots & \underline{\mathbf{n}}_{2^{2K}}^T \end{bmatrix}^T \\
\tilde{\mathbf{H}} &= \mathbf{I}_{2^{2K}} \otimes \underline{\mathbf{H}} \\
\tilde{\mathbf{y}} &= \begin{bmatrix} \underline{\mathbf{y}}_1^T & \underline{\mathbf{y}}_2^T & \cdots & \underline{\mathbf{y}}_{2^{2K}}^T \end{bmatrix}^T \\
\tilde{\mathbf{y}} &= \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \qquad\qquad (2.30) \\
\Gamma &= \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{2K} \end{bmatrix}^T
\end{aligned}$$

It can be readily seen that the average power constraint in (2.23) is equivalent to $\tilde{\mathbf{x}}^T\tilde{\mathbf{x}} \leq P2^{2K}$. Similar to (2.25), the simplified probability of error minimization problem under the average power constraint is given by

$$\begin{aligned}
\min_{\tilde{\mathbf{x}},\Gamma} \quad & \frac{1}{K2^{2K+1}} \sum_{k=1}^{K2^{2K+1}} Q(\sqrt{2}\tilde{b}_k(\tilde{\mathbf{h}}_k^T\tilde{\mathbf{x}} - \gamma_{\text{mod}(k,2K)})) \\
\text{s.t.} \quad & \tilde{\mathbf{x}}^T\tilde{\mathbf{x}} \leq P2^{2K} \\
& \tilde{b}_k(\tilde{\mathbf{h}}_k^T\tilde{\mathbf{x}} - \gamma_{\text{mod}(k,2K)}) \geq 0, k = 1, 2, \cdots, K2^{2K+1}
\end{aligned} \tag{2.31}$$

where $\text{mod}(k, 2K)$ denotes the remainder of k divided by $2K$. Since $\tilde{b}_k(\tilde{\mathbf{h}}_k^T\tilde{\mathbf{x}} - \gamma_{\text{mod}(k,2K)})$ is a linear mapping, (2.31) is also a convex optimization problem. The Lagrangian *Karush-Kuhn-Tucker* (KKT) conditions of (2.31) are

$$\tilde{b}_k(\tilde{\mathbf{h}}_k^T\tilde{\mathbf{x}} - \gamma_{\text{mod}(k,2K)}) \geq 0, k = 1, 2, \cdots, K2^{2K+1} \tag{2.32}$$

$$\tilde{\mathbf{x}}^T\tilde{\mathbf{x}} \leq P2^{2K} \tag{2.33}$$

$$\lambda_k \geq 0, k = 1, 2, \cdots, K2^{2K+1}, \nu \geq 0 \tag{2.34}$$

$$\nu(\tilde{\mathbf{x}}^T\tilde{\mathbf{x}} - P2^{2K}) = 0 \tag{2.35}$$

$$\lambda_k\tilde{b}_k(\tilde{\mathbf{h}}_k^T\tilde{\mathbf{x}} - \gamma_{\text{mod}(k,2K)}) = 0, k = 1, 2, \cdots, K2^{2K+1} \tag{2.36}$$

$$-\sum_{k=1}^{K2^{2K+1}} \left(\frac{1}{K2^{2K+1}\sqrt{\pi}} e^{-(\tilde{\mathbf{h}}_k^T\tilde{\mathbf{x}} - \gamma_{\text{mod}(k,2K)})^2} + \lambda_k \right) \tilde{b}_k \begin{bmatrix} \tilde{\mathbf{h}}_k \\ -\mathbf{e}_{\text{mod}(k,2K)} \end{bmatrix} + 2\nu \begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{0} \end{bmatrix} = 0 \tag{2.37}$$

where $\mathbf{e}_k$ denotes the $2K \times 1$ vector whose kth element is 1 and all other elements are zero, $\mathbf{0}$ denotes $2K \times 1$ zero vector. ν and λ_k are dual variables. Because (2.31) is convex, any $\tilde{\mathbf{x}}, \Gamma, \nu$ and λ_k that satisfy the above KKT conditions are primal and dual optimal [3].

Now we assume that a zero-threshold sign detector is used for (2.30). Then (2.31) becomes

$$\begin{aligned} \min_{\tilde{\mathbf{x}}} \quad & \frac{1}{K2^{2K+1}} \sum_{k=1}^{K2^{2K+1}} Q(\sqrt{2}\tilde{b}_k \tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}}) \\ \text{s.t.} \quad & \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} \leq P2^{2K} \\ & \tilde{b}_k \tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}} \geq 0, k = 1, 2, \cdots, K2^{2K+1} \end{aligned} \tag{2.38}$$

It has been shown in the previous section that (2.38) is equivalent to (2.25) because the left half of $\mathbf{B}$ is equivalent to its right half. The Lagrangian KKT conditions of (2.38) are

$$\tilde{b}_k \tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}} \geq 0, k = 1, 2, \cdots, K2^{2K+1} \tag{2.39}$$

$$\tilde{\mathbf{x}}^T \tilde{\mathbf{x}} \leq P2^{2K} \tag{2.40}$$

$$\lambda_k \geq 0, k = 1, 2, \cdots, K2^{2K+1}, \nu \geq 0 \tag{2.41}$$

$$\nu(\tilde{\mathbf{x}}^T \tilde{\mathbf{x}} - P2^{2K}) = 0 \tag{2.42}$$

$$\lambda_k \tilde{b}_k \tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}} = 0, k = 1, 2, \cdots, K2^{2K+1} \tag{2.43}$$

$$-\sum_{k=1}^{K2^{2K+1}} \left(\frac{1}{K2^{2K+1}\sqrt{\pi}} e^{-(\tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}})^2} + \lambda_k \right) \tilde{b}_k \tilde{\mathbf{h}}_k + 2\nu \tilde{\mathbf{x}} = 0 \tag{2.44}$$

Assuming that $\tilde{\mathbf{x}}^{\text{opt}}, \nu^{\text{opt}}$ and $\{\lambda_1^{\text{opt}}, \lambda_2^{\text{opt}}, \cdots, \lambda_{K2^{2K+1}}^{\text{opt}}\}$ are a set of primal and dual optimal points of (2.38), from (2.37), we can see that $\tilde{\mathbf{x}}^{\text{opt}}, \Gamma = \mathbf{0}$ and $\nu^{\text{opt}}, \{\lambda_1^{\text{opt}}, \lambda_2^{\text{opt}}, \cdots, \lambda_{K2^{2K+1}}^{\text{opt}}\}$ are also the primal and dual optimal points of (2.32)–(2.37). Therefore $\Gamma = \mathbf{0}$ is optimal even if we regard Γ as an optimization variable. Based on the above discussion, the following proposition is given as a summary.

Proposition 2.1 *The sign detector with zero threshold is optimal even if we jointly design the transmit vector* $\mathbf{x}$ *and the decision threshold* γ_k.

2.6 Simulation Results

In our simulations, we assume that the channel coefficients are independently zero-mean complex circular Gaussian random variables. Both theoretical and numerical results are compared for the proposed methods. The theoretical plots are obtained from the exact probability of error which comes as a by-product of solving optimization problems. The numerical results correspond to experimental computation of the error probability. For comparison reasons, we also plot the numerical probability of error of the linear transmitters based on the SINR balancing [23] and the total MMSE [14]. In each example, 300 independent channel realizations are used.

In practical scenarios, some users are far from the BS and other are near BS. As a result, different users are subject to different channel gains. To gain an insight into this phenomenon, both symmetrical and asymmetrical channel statistics are considered for performance analysis.

2.6.1 Symmetrical Channel Statistics

In this case, we assume the Gaussian distributed channels of all users have unit variance, i.e., $h_{ij} \sim \mathcal{CN}(0,1)$. Figures 2.1 and 2.2 show the probability of bit error of the methods tested versus the total transmit power. In Figure 2.1, $K = 2$ users and $M = 2$ BS antennas are assumed, while in Figure 2.2, $K = 2$ users and $M = 3$ BS antennas are considered. From Figure 2.1, we can see that at high transmit powers, the performance gain of the proposed technique under the average power constraint is about 2 dB over the linear methods of [23] and [14], which also use the average power constraint. The performance of the proposed technique under the peak power constraint is nearly identical to that of the linear methods of [14] and [23]. From Figure 2.2, we observe that when the number of BS antennas is increased to $M = 3$, the proposed technique with the average power constraint has about 0.7 dB of improvement over the linear methods. Note that in this case the proposed technique with the peak power constraint performs worse than the linear techniques tested. The explanation for this is that the linear techniques use the average rather than the peak power constraint. Therefore more power can be allocated to the transmit vectors corresponding to those bit vectors with high error probability while less power can be allocated to the transmit vectors corresponding to those bit vectors with low error probability.

Recall that the objective function we are aiming to minimize in our optimization problems is the average probability of error of all the users. To illustrate the individual performance, we plot the probability of error of each user in Figure 2.3 when there are two BS antennas and two users. It can be seen that each user in the proposed methods has almost the same error probability under both the peak and average power constraints.

2.6.2 Asymmetrical Channel Statistics

In this example, we assume the Gaussian distributed channels of different users have different variances, i.e., $h_{ij} \sim \mathcal{CN}(0, \beta_i)$. In Figure 2.4, two BS antennas and two users are assumed. The variances of the channel of the users are $\beta_1 = 0.1, \beta_2 = 1$, which means that the channel of the first user is 10dB weaker than that of the second one. Surprisingly, even the proposed method under the peak power constraint beats the linear methods. One reasonable explanation is that SINR and MMSE are not directly associated with the probability of error. Therefore balancing the SINR or minimizing the total MMSE may not minimize the probability of error. In Figure 2.5, the average probability of error of each user is plotted. Although the proposed method is designed to minimize the average probability of error of all the users, significant fairness between the users is achieved.

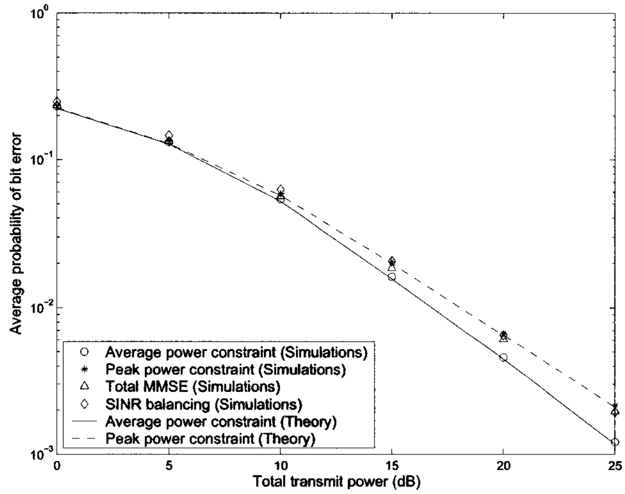

Figure 2.1: Comparison of the bit error probabilities of different methods for MIMO broadcast channel with two BS transmit antennas and two users under symmetrical channel statistics.

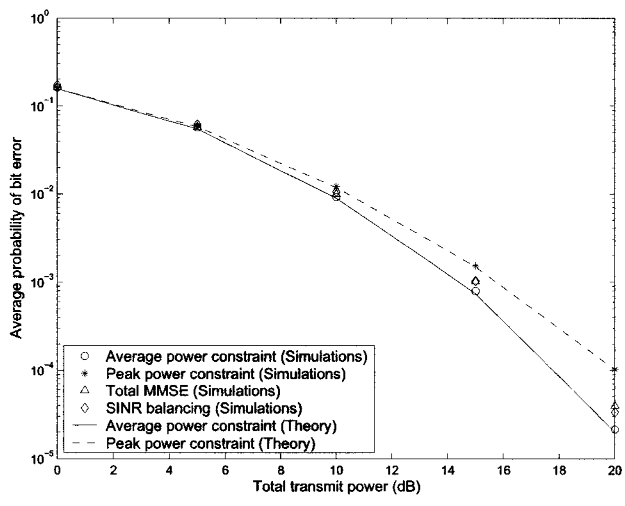

Figure 2.2: Comparison of the bit error probabilities of different methods for MIMO broadcast channel with three BS transmit antennas and two users under symmetrical channel statistics.

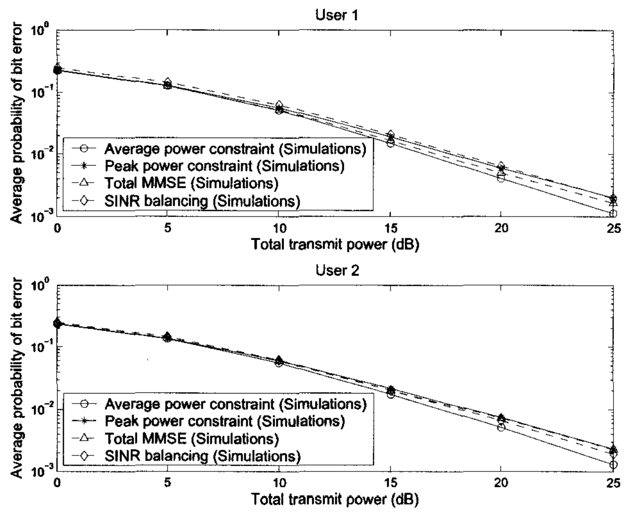

Figure 2.3: Comparison of the bit error probabilities of different methods of each user for MIMO broadcast channel with two BS transmit antennas and two users under symmetrical channel statistics.

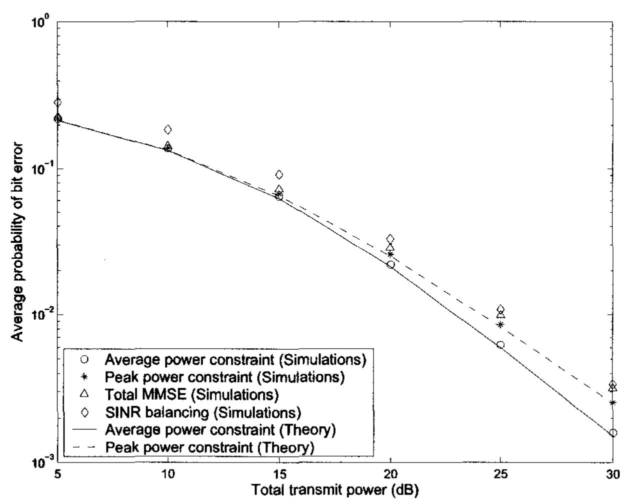

Figure 2.4: Comparison of the bit error probabilities of different methods for MIMO broadcast channel with two BS transmit antennas and two users under asymmetrical channel statistics.

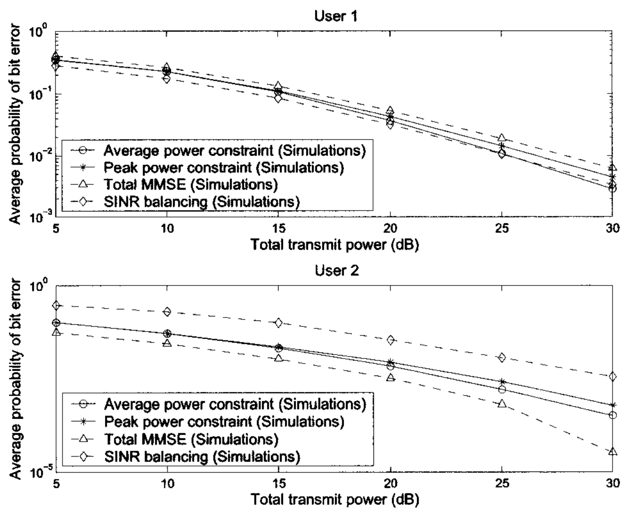

Figure 2.5: Comparison of the bit error probabilities of different methods of each user for MIMO broadcast channel with two BS transmit antennas and two users under asymmetrical channel statistics.

Chapter 3

Random Unitary Beamforming with Partial Feedback Using Multiuser Diversity

In the previous chapter, we assumed that K was less than M and minimized the probability of error. If there are a lot of users in the system, we could use the concept of multiuser diversity to improve the capacity. A fundamental characteristic of the multiuser diversity schemes is that the necessary amount of feedback linearly increases with the number of users. This large amount of feedback prohibits practical applications of multiuser diversity if there are a lot of users. In this chapter, we follow the work of [25] and propose an random unitary beamforming algorithm aimed at reducing the necessary amount of feedback by finding M or less than M quasi-orthogonal users.

3.1 System Model

We consider a single-cell down-link MIMO channel with M antennas at the BS and K mobile users each with one antenna. We assume a frequency flat Rayleigh fading channel where h_{ij} is the path gain from the jth BS antenna to the ith user and it is independent complex circular symmetric Gaussian with unit variance. The input-output relationship of this MIMO system model is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.1}$$

where

$$\begin{aligned} \mathbf{y} &= [y_1\, y_2\, \cdots\, y_K]^T \\ \mathbf{x} &= [x_1\, x_2\, \cdots\, x_M]^T \\ \mathbf{n} &= [n_1\, n_2\, \cdots\, n_K]^T \end{aligned}$$

are the received signal vector, transmitted signal vector and the zero-mean unit-variance complex white Gaussian noise vector, respectively, $(\cdot)^T$ denotes the transpose, and $\mathbf{H} = [h_{ij}] \in \mathbb{C}^{K\times M}$ is the channel matrix. We assume that each user knows its channel $\mathbf{h}_k^T$, which is the kth row of $\mathbf{H}$. We also assume that the BS only knows partial information about $\mathbf{H}$, which depends on how much channel information the users will feed back to the BS.

The input is constrained to satisfy

$$\mathrm{E}\{\mathbf{x}^H\mathbf{x}\} \leq P$$

where P, $\mathrm{E}\{\cdot\}$ and $(\cdot)^H$ denote the maximum allowed transmit power at the BS, statistical expectation and Hermitian transpose operator, respectively.

We define the normalized cross-correlation between $\mathbf{h}_i$ and $\mathbf{h}_j$ as

$$c_{ij} = \begin{cases} 1 & i = j \\ \frac{\mathbf{h}_j^H\mathbf{h}_i}{\|\mathbf{h}_i\|\|\mathbf{h}_j\|} & i \neq j \end{cases} \tag{3.2}$$

where $\|\cdot\|$ denotes the ℓ^2-norm.

We cite the following lemma from multivariate statistical theory, [18] and [10]

Lemma 3.1 *Suppose* $\mathbf{x}$ *is an* $M \times 1$ *vector with i.i.d* $\mathcal{CN}(0,1)$[1] *entries. Let*

$$\overline{\mathbf{x}} = [\Re\{\mathbf{x}\}^T \; \Im\{\mathbf{x}\}^T]^T \tag{3.3}$$

where $\Re(\cdot)$ *and* $\Im(\cdot)$ *stand for the real and imaginary parts, respectively, then* $\overline{\mathbf{x}}/\|\overline{\mathbf{x}}\|$ *is uniformly distributed on the surface of a* $2M$*-dimensional unit ball. This distribution is said to be isotropic.*

Lemma 3.1 is a straightforward extension of the result in [18]. Because the ℓ^2-norm of $\overline{\mathbf{x}}/\|\overline{\mathbf{x}}\|$ is one, it always lies on the surface of the $2M$-dimensional unit ball.

3.2 Proposed random beamforming

To improve the uplink spectrum efficiency, our question is: can we propose an algorithm in which the total amount of feedback does not change with K?

In [25], BS generates a unitary beamforming matrix and transmits it to all the users, assuming there is no channel state information available at the BS. Each user computes its SINRs corresponding to each beam and feed back its maximum SINR and its corresponding beam index to the BS. Then BS selects a user with the maximum SINR for each beam. It is proved in [25] that the asymptotic capacity of this scheme is the same as that with perfect channel information at the BS.

In this chapter, following the work of [25], we consider a linear transmission and reception strategy using random unitary beamforming to decrease the amount of feedback required. We assume that TDMA combined with the space-division-multiple-access (SDMA) is used by the BS to perform its communication. We also assume

[1] $\mathcal{CN}(0,1)$ *means zero-mean complex Gaussian random variable with independent, equal-variance real and imaginary parts.*

$K \geq M$ to gain multiuser diversity. In each time slot, a random unitary beamforming matrix is generated at the BS and M or less than M nearly orthogonal users are found. Therefore, in each time slot at most M users are served simultaneously.

3.2.1 Formulation

We first introduce the following definition.

Definition 3.1 *Given a small real number ε, $0 < \varepsilon < 1$, the ith and jth users are called quasi-orthogonal if $|c_{ij}| \leq \varepsilon$, where c_{ij} is defined in (3.2). For a subset $\mathcal{A} \subset \{1, 2, \cdots, K\}$, if $\forall\ i, j \in \mathcal{A}$, $|c_{ij}| \leq \varepsilon$, we say that users in $\mathcal{A}$ are quasi-orthogonal.*

Let $\mathbf{V}$ denote an $M \times M$ pseudo-random complex unitary beamforming matrix (we assume $\mathbf{V}$ is known at each user and varies in each time slot). If we allocate the power uniformly among the beams, then the received signal of the kth user is given by

$$y_k = \sqrt{\frac{P}{M}}\mathbf{h}_k^T \mathbf{V}\mathbf{s} + n_k \tag{3.4}$$

where

$$\mathbf{s} = [s_1\ s_2\ \cdots\ s_M]^T$$

is the $M \times 1$ transmit symbol vector. n_k is the additive noise at the kth user. Without loss of generality, we assume that $\mathrm{E}\{|s_i|^2\} = 1$.

Assume that the ith beam is allocated to the kth user, then the SINR of the kth user is given by

$$\mathrm{SINR}_{ki} = \frac{\frac{P}{M}|\mathbf{h}_k^T \mathbf{v}_i|^2}{\sum_{j=1, j\neq i}^{M} \frac{P}{M}|\mathbf{h}_k^T \mathbf{v}_j|^2 + 1} \tag{3.5}$$

where $\mathbf{v}_i$ is the ith column of $\mathbf{V}$.

As pointed out in [25], it is reasonable to set a threshold for the SINR because not all the users need to feed back. Only if the maximum (Here 'maximum' means the maximum SINR over the M SINRs of a user.) SINR of a user is larger than this threshold, it feeds back. To select this threshold, we need to know the maximum SINR distribution of a user. However, the M SINRs $\{\text{SINR}_{k1}, \text{SINR}_{k2}, \cdots, \text{SINR}_{kM}\}$ (corresponding to each beam) of a user is not independent each other. Therefore, it's pretty hard (or impossible) to find the probability distribution function of $\max_{i=1,2,\cdots,M} \text{SINR}_{ki}$. Moreover, the work of [25] assumes that all the user have the same SNR. In practice, some user may be near the BS, and some user may be from from the BS. As a result, different user will have different SNR and different SINR distribution. Therefore, it's impossible to set a threshold for the SINR and analyze the performance of the system.

Let

$$\mathbf{z}_k = \frac{\mathbf{h}_k^T \mathbf{V}}{||\mathbf{h}_k||} \tag{3.6}$$

$$\mathbf{z}_k = \begin{bmatrix} z_{k1} & z_{k2} & \cdots & z_{kM} \end{bmatrix} \tag{3.7}$$

With the definition of $\mathbf{z}_k$, we can also compute the SINR as

$$\text{SINR}_{ki} = \frac{\frac{P}{M}||\mathbf{h}_k||^2||z_{ki}||^2}{\sum_{j=1,j\neq i}^{M} \frac{P}{M}||\mathbf{h}_k||^2||z_{kj}||^2 + 1} = \frac{\frac{P}{M}||\mathbf{h}_k||^2||z_{ki}||^2}{\frac{P}{M}||\mathbf{h}_k||^2(1 - ||z_{ki}||^2) + 1} \tag{3.8}$$

From (3.8), we can see that SINR_{ki} monotonically increases with z_{ki}. In the following analysis, we will consider $\max_i |z_{ki}|$ as the decision variable to feed back rather than $\max_i ||\text{SINR}_{ki}||$.

From Lemma 3.1, the combined real and imaginary parts of $\mathbf{h}_i/||\mathbf{h}_i||$ are uniformly distributed on the surface of the $2M$-dimensional unit ball $\mathcal{B}_{2M} = \{\mathbf{x} : \mathbf{x}^T\mathbf{x} \leq 1,\ \mathbf{x} \in \mathcal{R}^{2M}\}$. In this case, the K users are uniformly placed on the surface of $\mathcal{B}_{2M}$. Given $\mathbf{V}$, each beam $\mathbf{v}_i$ can also be regarded as a point on the surface of $\mathcal{B}_{2M}$ (combine the real and imaginary parts of $\mathbf{v}_i$ to a $2M \times 1$ real vector). If K goes to infinity, we will find M users whose normalized channels coincide with the M beams, respectively. In that case, we obtain M orthogonal users. If K is limited, we will find M or less than M quasi-orthogonal users whose normalized cross-correlation is bounded.

To reduce the amount of feedback, in the proposed random beamforming, not all the users feed back. Only if the $\|\mathbf{z}_k\|_\infty$ is larger than a given threshold, i.e., $\{\max_i |z_{ki}| \geq \sqrt{\eta}\}$, does the kth user feed back. Here $\|\cdot\|_\infty$ and η denote the ℓ_∞-norm and the threshold, respectively.

In the following we give a geometric interpretation to this feedback process. Construct M cones $\mathcal{S}_i = \{\mathbf{x} : \frac{|\mathbf{v}_i^T\mathbf{x}|}{\|\mathbf{x}\|} \geq \sqrt{\eta},\ \mathbf{x} \in \mathcal{C}^M\},\ i = 1, 2, \cdots, M$. Note that $0 < \eta < 1$. If K is sufficiently large, $\mathcal{S}_i$ can be made small enough under the constraint that each cone has, on average, one user. Those users who fall within these cones will have low normalized cross-correlation. If $\mathbf{V}$ is isotropically distributed, it scans the surface of $\mathcal{B}_{2M}$ uniformly. Therefore, fairness is ensured in the sense that each user has the same probability to be served by the BS (i.e., fall into $\mathcal{S}_i$).

Fig. 3.1 gives an example when the k_1th and k_2th user fall into $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. We assume that there are two BS antennas. To simplify the figure, we also assume $\mathbf{H}$ and $\mathbf{V}$ are real matrices. The small triangles denote the normalized channels of the users. From Fig. 3.1 we can see that the k_1th and k_2th users have low normalized cross-correlation.

Suppose we want on average d users[2] to feed back in each time slot. Because $\mathbf{h}_k$ is independent among different users, to satisfy this condition we need

$$\text{Prob}(\max_i |z_{ki}|^2 \geq \eta) = \frac{d}{K}, \tag{3.9}$$

where $\text{Prob}(\max_i |z_{ki}|^2 \geq \eta)$ is the probability that the kth user feeds back. Note that (3.9) only ensures that the average number of users who feed back in each time slot is d. Sometime there maybe less or more than d users feeding back.

[2]In section 3.2.2 we will discuss in detail how to select d.

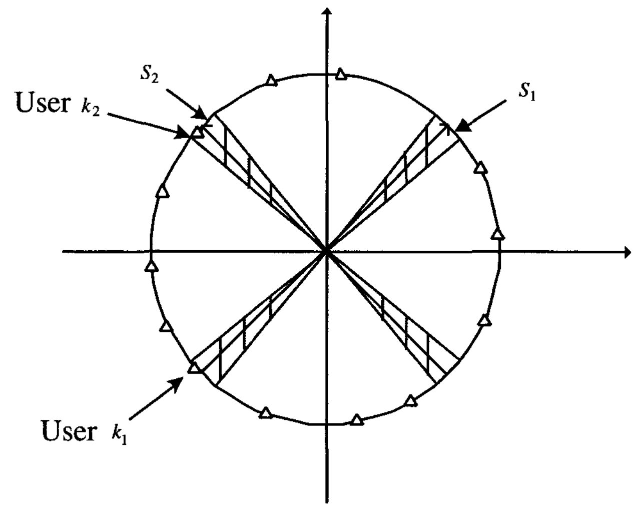

Figure 3.1: The k_1th and k_2th users fall into $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, when there are two BS antennas.

Since $\{|z_{ki}|^2,\ i = 1, 2, \cdots, M\}$ are not independent to one another, it is hard to find the distribution of $\max_i |z_{ki}|^2$.

Recall

$$\mathbf{z}_k = \frac{\mathbf{h}_k^T \mathbf{V}}{\|\mathbf{h}_k\|} = \frac{\mathbf{h}_k^T \mathbf{V}}{\|\mathbf{h}_k^T \mathbf{V}\|} = \frac{\tilde{\mathbf{h}}_k^T}{\|\tilde{\mathbf{h}}_k^T\|} \tag{3.10}$$

where

$$\tilde{\mathbf{h}}_k = \mathbf{V}^T \mathbf{h}_k$$

has the same distribution as $\mathbf{h}_k$. Fortunately, because $|\tilde{h}_{ki}|^2$ has a common exponential distribution ($\tilde{h}_{ki}$ is the ith element of $\tilde{\mathbf{h}}_k$), according to order statistics [7], $\{|z_{ki}|^2, i = 1, 2, \cdots, M\}$ have the same joint distribution as the spacings of random division of an unit interval. Therefore, from [7], we have

$$\begin{aligned} \text{Prob}(\max_i |z_{ki}|^2 \geq \eta) = M(1-\eta)^{M-1} - \binom{M}{2}(1-2\eta)^{M-1} \\ + \cdots + (-1)^{j-1}\binom{M}{j}(1-j\eta)^{M-1} \end{aligned} \tag{3.11}$$

where the series in (3.11) continues as long as $1-j\eta > 0$ and $\binom{M}{j}$ denotes the binomial coefficient.

Because $\text{Prob}(\max_i |z_{ki}|^2 \geq \eta)$ is a non-increasing function of η, η increases with K according to (3.9). If K is sufficiently large, $\eta \geq \frac{1}{2}$. Then (3.11) reduces to

$$\text{Prob}(\max_i |z_{ki}|^2 \geq \eta) = M(1-\eta)^{M-1} \tag{3.12}$$

Combine (3.9) and (3.12), we obtain

$$KM(1-\eta)^{M-1} = d \tag{3.13}$$

$$\eta = 1 - \left(\frac{d}{KM}\right)^{\frac{1}{M-1}} \tag{3.14}$$

From (3.14), as $K \to \infty$, $\eta \to 1$. The users in different cones approach orthogonality, which coincides with the previous analysis. To satisfy $\eta \geq \frac{1}{2}$, from (3.13), the minimum K required for a given η is

$$K_{\min} = \left\lceil \frac{d}{M(1-\eta)^{M-1}} \right\rceil \tag{3.15}$$

where $\lceil \cdot \rceil$ is the ceiling operator. We can see that $K_{\min}$ increases exponentially with M.

The proposed random beamforming is described in detail in Table 3.1. We stress that in Table 3.1 the feedback process is divided into two steps. If no user feeds back for some beams at a time slot, less than M beams will be used. In the fourth step of the proposed random beamforming, if each user knows the total transmit power, the BS only needs to transmit the number of beams to be used which is an integer between one and M. This is an easier task compared to transmitting the allocated power which is a real number.

In contrast to the opportunistic beamforming [32], the proposed random beamforming provides performance improvement in both slow and fast fading environment because the performance of the proposed random beamforming depends not on the

Table 3.1: Proposed random beamforming

Aim:	Given K and d, find M or less than M quasi-orthogonal users.
Init:	According to (3.9) and (3.11), the BS computes η and transmits η to each user.
Step 1	In the beginning of each time slot, the BS generates an isotropically distributed $\mathbf{V}$ and transmits $\mathbf{V}$ to each user.
Step 2	If $\max_i |z_{ki}|^2 \geq \eta$, the kth user feeds back the beam index $i_k^* = \text{argmax}_i |z_{ki}|^2$.
Step 3	The BS determines how many beams will be used and then uniformly allocates the power to those beams.
Step 4	The BS transmits the allocated power to the users who fed back in step 2. And these users compute their SINRs.
Step 5	If $\max_i |z_{ki}|^2 \geq \eta$, the kth user feeds back its SINR corresponding to the i_k^*th beam.
Step 6	THE BS selects a user with the maximum SINR for each beam to perform communication.

rate and dynamic range of the channel variations, but on the independence of the users' channels. The amount of multiuser diversity depends on the number of users. From Table 3.1 and (3.9), we see that the average amount of feedback of the proposed random beamforming per time slot is d real numbers and d integers, which does not vary with K.

3.2.2 Selection of d

Let the random variable $\mathbf{A}$ denote the number of users who feed back at a particular time slot. To simplify the notation, we set $\mu = \text{Prob}(\max_i |z_{ki}|^2 \geq \eta)$. In the proposed random beamforming, the event that the kth user feeds back at a time slot is equivalent to a Bernoulli trial. Accordingly, $\mathbf{A}$ has a binomial distribution with mean d. Then we have

$$\text{Prob}(\mathbf{A}=a)=\binom{K}{a}\mu^a(1-\mu)^{K-a},\ a=0,1,\cdots,K \tag{3.16}$$

Now we assume that there are a users feeding back at a particular time slot. Because $\mathbf{h}_k$ are independent from one another and $\mathbf{V}$ is unitary, i^* ($i^*=\text{argmax}_i\ |z_{ki}|^2$) is uniformly distributed among $\{1,2,\cdots,M\}$. In the second step of Table 3.1, each of these a users who feeds back has the same probability to be associated with any of the M beams. The process that the a users are associated with the M beams is equivalent to a distribution problem: Distribute a distinguishable balls into M distinguishable boxes. The event that the kth user is associated with the ith beam is equivalent to that the kth ball is placed into the ith box. It is easily shown that there are totally M^a ways to place a distinguishable balls into M distinguishable boxes. If no user feeds back for the ith beam, the ith beam will not be used.

Let M_{av} denote the average number of the used beams per time slot, and E_j denote the event "j beams are used in a time slot". Assuming that $N(m,n)$ is the number of ways that m distinguishable balls are to be placed into n distinguishable boxes and that none of the boxes are empty, from [22], we have

$$N(m,n)=\sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}i^m,\ m\geq n \tag{3.17}$$

When there are a users feeding back at a time slot, the probability that j beams are used is

$$\text{Prob}(E_j|\mathbf{A}=a)=\binom{M}{j}\frac{N(a,j)}{M^a},\ j\leq\min(M,a) \tag{3.18}$$

The average number of used beams per time slot is

$$\begin{aligned} M_{\text{av}} &= \sum_{a=1}^{K} \text{Prob}(\mathbf{A}=a) \sum_{j=1}^{\min(M,a)} \text{Prob}(E_j|\mathbf{A}=a)j \\ &= \sum_{a=1}^{K} \sum_{j=1}^{\min(M,a)} \binom{M}{j} \frac{N(a,j)}{M^a} \binom{K}{a} \mu^a (1-\mu)^{K-a} j \\ &= \sum_{a=1}^{K} \sum_{j=1}^{\min(M,a)} \sum_{i=1}^{j} (-1)^{j-i} \binom{j}{i} i^a \binom{M}{j} j \frac{\binom{K}{a}(d/K)^a (1-d/K)^{K-a}}{M^a} \end{aligned} \tag{3.19}$$

Fig. 3.2 plots the theoretical and simulation-based average number of used beams per time slot with respect to K. We assume that there are four antennas at the BS. The simulation results are obtained by averaging over 10,000 independent channel realizations. We can see that M_{av} does not change with K.

We admit that it is more reasonable to select d based on the sum rate. However, given K, M and P, it's pretty hard to get a closed-form solution of the sum rate with respect to d. A straightforward method is to plot the curve of the sum rate with respect to d and select d which maximizes the sum rate. For simplicity, we use a suboptimal way to choose d according to the average number of used beams per time slot plotted in Fig. 3.2. There is a tradeoff when choosing d. Large d cause large M_{av} and bad cross-correlation among the selected users. Small d cause small M_{av} and good cross-correlation among the selected users. From Fig. 3.2, when $d = 2M$, $M_{av} \simeq 3.5$. We take $d = 2M$ as an acceptable choice between these tradeoffs.

3.2.3 Scheduling

We see that the proposed random beamforming always picks the user inside the cone $\mathcal{S}_i$ with the maximum SINR for the ith beam. As a result, this will introduce unfairness among the users, especially in slow fading environment where the channel is fixed for a long time. Those users with high SINR will have larger probability to

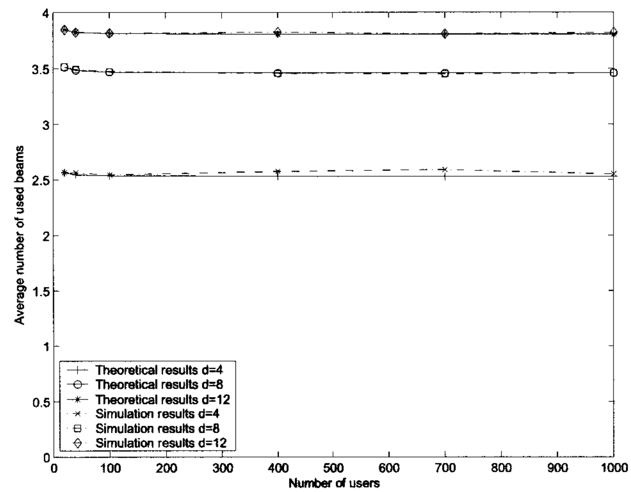

Figure 3.2: Average number of used beams per time slot with four BS antennas.

be served than those users with low SINR. In this subsection, we propose an equal ratio scheduling that is able to serve the users with different rate requirements.

Recall that $\mathbf{V}$ is isotropically distributed. From Lemma 3.1, if $\mathcal{S}_i$ scans the surface of $\mathcal{B}_{2M}$ uniformly, all of the users will have the same probability to lie in $\mathcal{S}_i$. Therefore, all of the users will be fairly treated. The probability that the kth user falls into any of $\mathcal{S}_i$ is

$$\text{Prob}(\max_i |z_{ki}|^2 \geq \eta) = \frac{d}{K} \tag{3.20}$$

We assume that the latency time is t_c time slots. Let r_k^d denote the desired rate of the kth user. We use $r_k^a(t)$ to denote the average rate of the kth user from the $(t - t_c + 1)$th time slot to the tth time slot. We stress that $r_k^a(t)$ means the average rate in the past t_c time slots. The proposed *equal ratio scheduling (ERS)* selects the user with the minimum average-to-desired rate ratio at the tth time slot

$$k^* = \arg \min_{k \in \mathcal{K}_i(t)} \frac{r_k^a(t)}{r_k^d} \tag{3.21}$$

for the ith beam, where $\mathcal{K}_i(t)$ denotes the set of users who feed back for the ith beam at the tth time slot. The proposed random beamforming with ERS is described

Table 3.2: Proposed random beamforming with ERS

Aim:	Given K and d, find M or less than M quasi-orthogonal users.
Init:	According to (3.9) and (3.11), the BS computes η and transmits η to each user.
1	In the beginning of each time slot, the BS generates an isotropically distributed $\mathbf{V}$ and transmits $\mathbf{V}$ to each user.
2	If $\max_i \|z_{ki}\|^2 \geq \eta$, the kth user feeds back the beam index $i_k^* = \mathrm{argmax}_i \|z_{ki}\|^2$.
3	The BS selects a user for each beam with the minimal average-to-desired rate ratio.
4	The BS transmits the allocated power so that the selected users can compute their SINR.
5	The selected users feed back their SINRs corresponding to the preferred beams.

in detail in Table 3.2. The average amount of feedback in the proposed random beamforming with ERS per time slot is M_{av} real numbers and d integers.

A fundamental issue in ERS is how to update $r_k^a(t)$ in each time slot. A straightforward method is to store the rate of each user in the past t_c time slots for updating $r_k^a(t)$. Although this approach provides the exact average rate, it requires a lot of memory if t_c and K are large. In [32], proportional fair scheduling uses an exponentially weighted low-pass filter to update $r_k^a(t)$ given by

$$r_k^a(t+1) = \begin{cases} (1-\frac{1}{t_c})r_k^a(t) + \frac{1}{t_c}\log_2(1+\mathrm{SINR}_k) & k = k^* \\ (1-\frac{1}{t_c})r_k^a(t) & k \neq k^* \end{cases} \tag{3.22}$$

In the following we point out that this low-pass filter may not accurately track the average rate if t_c is small relative to $\frac{K}{M_{\mathrm{av}}}$.

Assume the kth user is not served in the past t_c time slots, $r_k^a(t)$ should be zero. From (3.22),

$$r_k^a(t+t_c) = (1-\frac{1}{t_c})^{t_c} r_k^a(t) \tag{3.23}$$

Let

$$g(x) = (1-\frac{1}{x})^x, \qquad x \geq 2 \tag{3.24}$$

We have $g(2) = 0.25$ and $\lim_{x\to\infty} g(x) = e^{-1}$. Next, we show that $g(x)$ is increasing with x. Let $h(x) = \ln g(x)$, then

$$\begin{aligned} h(x) &= x \ln(1 - \frac{1}{x}) \\ h'(x) &= \ln(x-1) - \ln(x) + \frac{1}{x-1} \\ h''(x) &= \frac{-1}{x(x-1)^2} \end{aligned}$$

Since $h''(x) < 0$ for $x \geq 2$, $h'(x)$ is decreasing with x for this domain. Because $\lim_{x\to\infty} h'(x) = 0$, $h'(x)$ is always greater than zero. Therefore $g(x)$ is increasing with x. From the definition of $r_k^a(t)$, it should be zero at the $(t+t_c)$th time slot if the kth user is not served in the past t_c time slots. However, from (3.23),

$$\frac{r_k^a(t+t_c)}{r_k^a(t)} = (1 - \frac{1}{t_c})^{t_c} \in [0.25, e^{-1}] \tag{3.25}$$

which shows that the exponentially updated average rate maybe a poor approximation of the exact average rate.

In this section we propose a linear method to update $r_k^a(t)$ as follows

- If $k = k^*$

$$\begin{aligned} r_k^a(t+1) &= \Big((t_c - n_k(t))r_k^i(t) + \log_2(1 + \mathrm{SINR}_k)\Big)/t_c \\ n_k(t+1) &= 1 \end{aligned}$$

- If $k \neq k^*$

$$\begin{aligned} r_k^a(t+1) &= (t_c - n_k(t))r_k^i(t)/t_c \\ n_k(t+1) &= \min(t_c, n_k(t)+1) \\ r_k^i(t+1) &= r_k^i(t) \end{aligned} \tag{3.26}$$

where $n_k(t)$ denotes the number of time slots between the tth time slot and the last transmission of the kth user, and $r_k^i(t)$ denotes the average rate of the kth user in

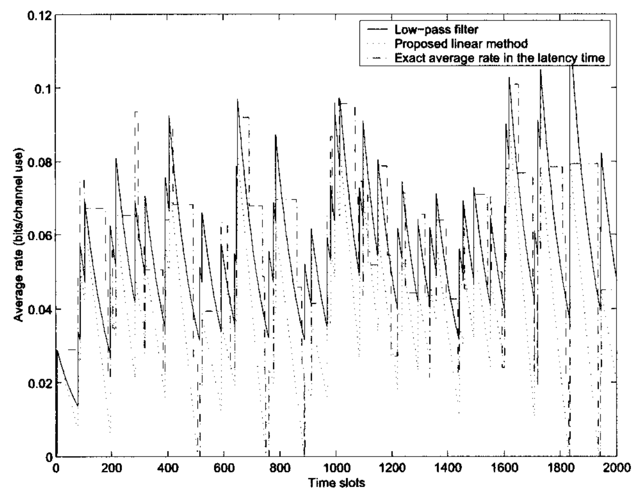

Figure 3.3: Performance of different approaches to approximate the average rate of the 1st user over 2000 time slots.

the last t_c time slot. If the kth user is not served in the past t_c time slots, $r_k^a(t)$ will definitely be zero.

One can get an intuitive feeling about the performance of the aforementioned methods by inspecting Fig. 3.3. We assume there are four BS antennas, one hundred users and the latency time is one hundred time slots. The total transmitted power is 20dB. We plot the sample average rate of the first user over 2000 consecutive time slots to see if it is accurately tracked. In Fig. 3.3 we can see that in some time slots, the exact average rate obtained by averaging over the past t_c time slots is zero and the low-pass filter method cannot track it as the range of $g(x)$ is $[0.25, e^{-1}]$. Note that t_c is small relative to $\frac{K}{M_{\text{av}}}$, the first user may not be served in consecutive t_c time slots causing the average rate to go to zero.

In order to satisfy the rate requirements of all the users, the sum of the desired rate should be less than what the channel can support. Based on the fact that all users have the same probability to lie in $\mathcal{S}_i$, the ERS can allocate time slots to users proportional to their average vs. desired rate ratio.

Table 3.3: Amount of Feedback

Random beamforming scheduling	K real numbers + K integers
Proposed random beamforming (average)	d real numbers + d integers
Proposed random beamforming with ERS (average)	M_{av} real numbers + d integers

Table 3.3 gives the amount of feedback of the aforementioned algorithms. Generally, K is much larger than M. We can see that the proposed random beamforming with or without ERS requires far less feedback than the random beamforming scheduling.

3.2.4 Performance Analysis

In this section we give a lower and upper bound for the sum rate of the proposed random beamforming.

Assume powerful coding and decoding are used at the BS and the users, respectively, we employ the Shannon-limit rate to evaluate the performance of the proposed random beamforming. The rate of the kth user and the sum rate of all the users are given by

$$r_k = \log_2(1 + \text{SINR}_k) \tag{3.27}$$

$$r_{\text{sum}} = \sum_{k=1}^{K} r_k = \sum_{k=1}^{K} \log_2(1 + \text{SINR}_k) \tag{3.28}$$

respectively.

Assume that there are m selected beams and the ith beam is allocated to the kth user, let $\rho_m = \frac{P}{m}$, from (3.5), the SINR of the kth user is

$$\text{SINR}_k = \frac{\rho_m |\mathbf{h}_k^T \mathbf{v}_i|^2}{\rho_m \sum_{j=1, j\neq i}^{m} |\mathbf{h}_k^T \mathbf{v}_j|^2 + 1} \leq \rho_m \|\mathbf{h}_k\|^2$$

Since $||\mathbf{h}_k^T||^2$ is chi-square distributed with $2M$ degrees of freedom, an upper bound of the sum rate is given by

$$\begin{aligned} r_{\text{sum}} &= \sum_{m=1}^{M} \text{Prob}(E_m) m \text{E}\{\log_2(1+\text{SINR}_k)\} \\ &= \sum_{m=1}^{M}\sum_{a=m}^{K} \text{Prob}(E_m, \mathbf{A}=a) m \text{E}\{\log_2(1+\text{SINR}_k)\} \\ &= \sum_{m=1}^{M}\sum_{a=m}^{K} \text{Prob}(E_m|\mathbf{A}=a)\text{Prob}(\mathbf{A}=a) m \text{E}\{\log_2(1+\text{SINR}_k)\} \\ &\le \sum_{m=1}^{M}\sum_{a=m}^{K} \text{Prob}(E_m|\mathbf{A}=a)\text{Prob}(\mathbf{A}=a) m \int_0^{\infty} \frac{x^{M-1}e^{-x}}{(M-1)!} \log_2(1+\rho_m x)dx \end{aligned} \quad (3.29)$$

where $\text{Prob}(\mathbf{A}=a)$ and $\text{Prob}(E_m|\mathbf{A}=a)$ are defined in (3.16) and (3.18), respectively. If the kth user falls in $\mathcal{S}_i$, then $\frac{|\mathbf{h}_k^T\mathbf{v}_i|^2}{||\mathbf{h}_k||^2} \ge \eta$ and $\sum_{j=1,j\neq i}^{M} \frac{|\mathbf{h}_k^T\mathbf{v}_j|^2}{||\mathbf{h}_k||^2} \le 1-\eta$, and we have

$$\begin{aligned} \text{SINR}_k &= \frac{\rho_m|\mathbf{h}_k^T\mathbf{v}_i|^2}{\rho_m\sum_{j=1,j\neq i}^{m}|\mathbf{h}_k^T\mathbf{v}_j|^2+1} \\ &\ge \frac{\rho_m||\mathbf{h}_k||^2\eta}{\rho_m||\mathbf{h}_k||^2(1-\eta)+1} \end{aligned} \quad (3.30)$$

Let $X_m = \frac{\rho_m||\mathbf{h}_k||^2\eta}{\rho_m||\mathbf{h}_k||^2(1-\eta)+1}$, the cumulative distribution function (cdf) of X_m is

$$\begin{aligned} F_{X_m}(X_m \le x) &= \text{Prob}(\frac{\rho_m||\mathbf{h}_k||^2\eta}{\rho_m||\mathbf{h}_k||^2(1-\eta)+1} \le x) \\ &= \text{Prob}(||\mathbf{h}_k||^2 \le \frac{x}{\rho_m\eta - x\rho_m(1-\eta)}) \end{aligned} \quad (3.31)$$

Note that $X_m \le \frac{\eta}{1-\eta}$. As $||\mathbf{h}_k||^2$ has $\chi^2(2M)$ distribution, the probability distribution function (pdf) of X_m is

$$p_{X_m}(x) = \frac{1}{(M-1)!}(\frac{x}{\rho_m\eta - x\rho_m(1-\eta)})^{M-1} e^{-\frac{x}{\rho_m\eta - x\rho_m(1-\eta)}} \frac{\eta}{\rho_m(\eta - x(1-\eta))^2} \quad (3.32)$$

Combine (3.30) and (3.32), we can get a lower bound of the sum rate as given by

$$\begin{aligned} r_{\text{sum}} &= \sum_{m=1}^{M} \text{Prob}(E_m) m \text{E}\{\log_2(1+\text{SINR}_k)\} \\ &\geq \sum_{m=1}^{M} \sum_{a=m}^{K} \text{Prob}(E_m|\mathbf{A}=a)\text{Prob}(\mathbf{A}=a) m \int_0^{\frac{\eta}{1-\eta}} p_{X_m}(x) \log_2(1+x) dx \end{aligned} \quad (3.33)$$

3.3 Simulation Results

In order to compare the performance of the proposed random beamforming with that of the random beamforming scheduling [25], computer simulations have been carried out. In the random beamforming scheduling, at each time slot, the BS transmits the beamforming matrix to all the users. Each user computes its SINRs corresponding to each beam and feeds back the maximum SINR to the BS. The BS then selects a user with the maximum SINR for each beam.

Both symmetric and asymmetric fading channel statistics are studied because a near-far effect is inevitable in the practical cellular communication system. We adopt the total transmitted power P at the BS instead of the SNR since less than M beams may be used in some time slots and P will be assigned only to those used beams. In that case we will distribute the transmitted power uniformly into the selected beams. We assume $d = 2M$ in the following examples.

3.3.1 Symmetric Fading Channel Statistics

In the symmetric fading channel, we assume the channels of all the users have the same fading statistics, i.e., $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{I}_M)$. This models the scenario in which all the users have nearly the same distance from the BS. Fig.s 3.4 and 3.5 plot the sum rate of the proposed random beamforming with respect to K when there are four and eight BS antennas, respectively. The total transmitted power at the BS is fixed at 20dB.

The sum rate is obtained by averaging over 1000 independent channel realizations. In the proposed random beamforming with ERS, we assume that all of the users have the same desired rate.

In Fig. 3.4, with four BS antennas, the sum rate of the proposed random beamforming is slightly higher than that of the random beamforming scheduling when $K \leq 300$. We also plot the upper and lower bound of the proposed random beamforming. We can see that the upper bound is loose and the lower bound is about 2.5 bits/channel user less than the sum rate of the proposed random beamforming with ERS. In Fig. 3.5, the sum rate of the proposed random beamforming is always larger than that of the random beamforming scheduling with eight BS antennas. A rational explanation to this is that when K is small, selecting less than M quasi-orthogonal users (proposed random beamforming) may yield better performance than always selecting M users (random beamforming scheduling) because the latter may have larger normalized cross-correlation. Notice that, when there are 500 users, the sum rate with four BS antennas is larger than that with eight BS antennas, which contradicts with the full channel state information case where the sum rate linearly increases with the number of BS antennas. This property is also observed in [25] where the sum rate achieves its maximum with four BS antennas when the total transmitted power is 10dB.

The sum rate of the proposed random beamforming with ERS is less than those of the proposed random beamforming and random beamforming scheduling because it is aimed to serve the users whose average rates are much smaller compared to their desired rates. In both figures the maximum gap between the sum rate of the proposed random beamforming with ERS and those of the remaining two algorithms is about 1 bit/channel use, which is the price we pay for fairness. We remark that the sum rates of the aforementioned three algorithms in a slow fading channel is the same as those in fast fading channel because their performances do not depend on the rate of the

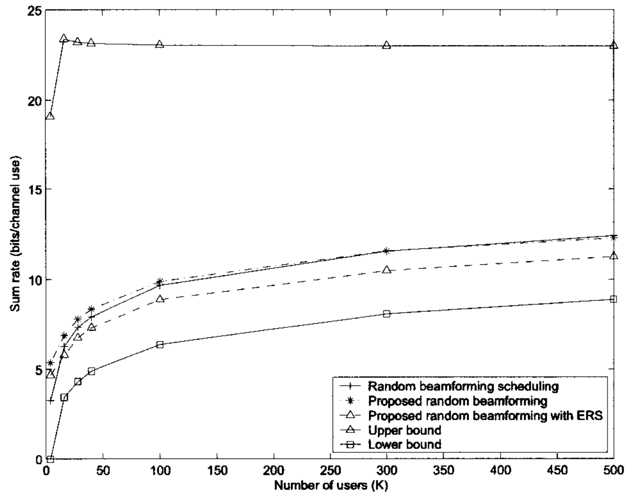

Figure 3.4: Sum rate vs. number of users for the symmetric channel with four BS antennas, $P = 20$dB.

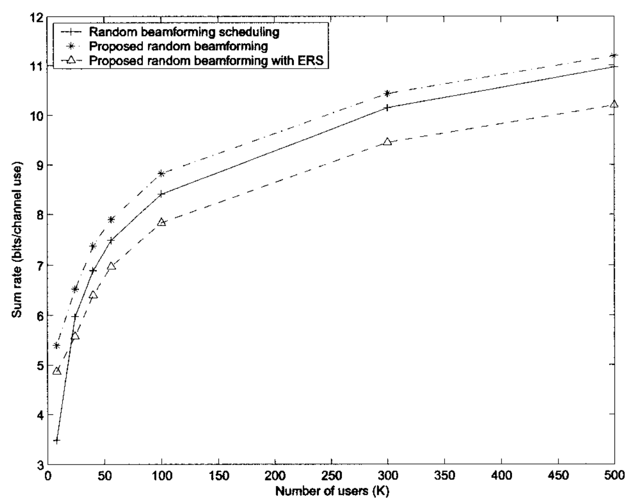

Figure 3.5: Sum rate vs. number of users for the symmetric channel with eight BS antennas, $P = 20$dB.

channel fluctuations. Their performances depend on the normalized cross-correlation between the channel and $\mathbf{V}$.

To illustrate the performance of the ERS, we perform two separate experiments in which each user is assumed to have the same or different rate requirement, respectively. We assume that there are four BS antennas and the latency time is 300 time slots.

Fig. 3.6 compares the average rate of the aforementioned methods at the 600th time slot in slow fading channel with 20 users which have the same desired rate. Clearly, the fluctuation of the rate curve of the proposed random beamforming with ERS is far smaller than that of the remaining two algorithms. The difference between the maximum and the minimum rate is about 0.04 bits/channel use for the proposed random beamforming with ERS while it is about 0.42 bits/channel use for the random beamforming scheduling.

Fig. 3.7 and 3.8 plots the average rate of each user at the 600th time slot with different rate requirements in slow and fast fading channel, respectively, when the proposed random beamforming with ERS is used. We can see that in both figures the average rate roughly approaches or exceeds the desired rate. Compared to the slow fading, fast fading improves the fairness among the users because in fast fading the channel varies rapidly so that each user will experience a good channel condition within limited time slots.

3.3.2 Asymmetric Channel Statistics

In this subsection, we study the performance of the proposed techniques when the channel of different users have different variances. We assume $\mathbf{h}_k \sim \mathcal{CN}(0, \alpha_k \mathbf{I}_M)$ and $10 \log_{10} \alpha_k$ is uniformly distributed in $[-10, 0]$ dB.

Fig. 3.9 and 3.10 show the sum rate of the proposed random beamforming with four and eight BS antennas, respectively. We can see that similar to the symmetric

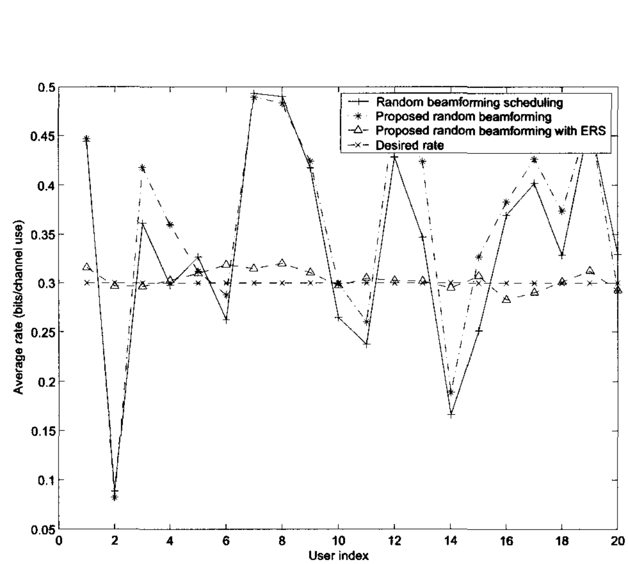

Figure 3.6: Average rate of each user with the same desired rate at the 600th time slot in slow fading and symmetric channel, $P = 20$dB.

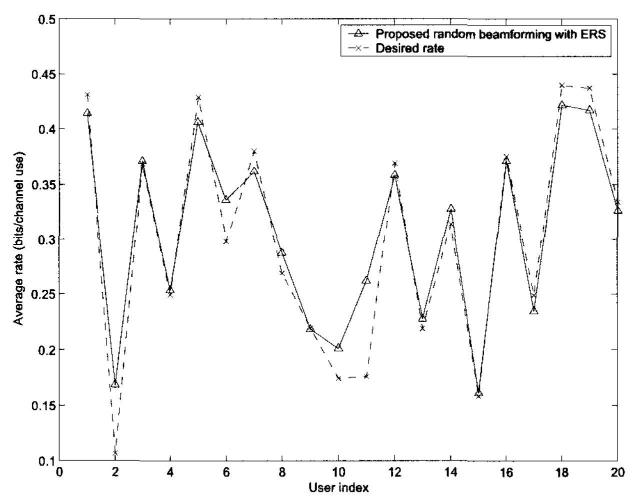

Figure 3.7: Average rate of each user with different desired rate at the 600th time slot in slow fading and symmetric channel, $P = 20$dB.

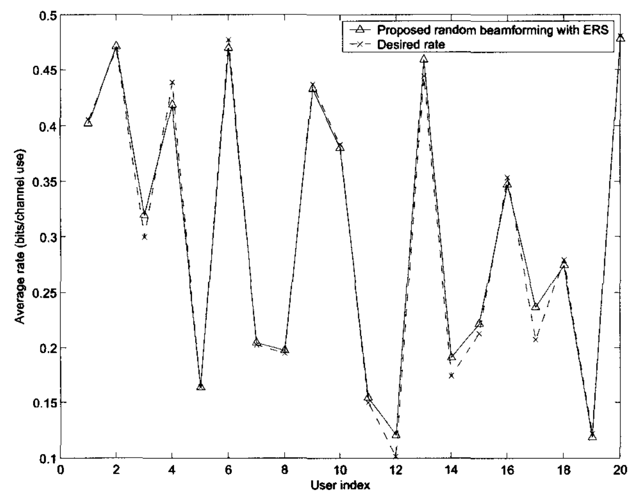

Figure 3.8: Average rate of each user with different desired rate at the 600th time slot in fast fading and symmetric channel, P = 20dB.

channel, in spite of a little rate loss due to smaller channel gain, Fig. 3.9 and 3.10 demonstrate that the sum rate of the proposed random beamforming is comparable to that of the random beamforming scheduling. And the sum rate of the proposed random beamforming with ERS is roughly 1 bit/channel use less than that of the other two algorithms.

Fig. 3.11 and 3.12 plot the average rate of each user in the 600th time slot over slow and fast fading channels, respectively. We assume there are 20 users and the latency time is 300 time slots. We can see that in fast fading ERS exactly satisfies the rate requirement of each user. In slow fading, the exact rate of some users is a little less than the desired rates, however, the rate gap is negligible.

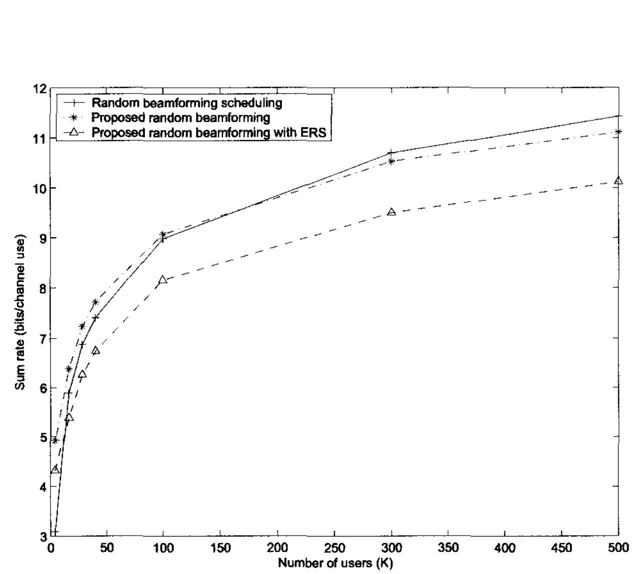

Figure 3.9: Sum rate vs. number of users for the asymmetric channel with four BS antennas, $P = 20$dB.

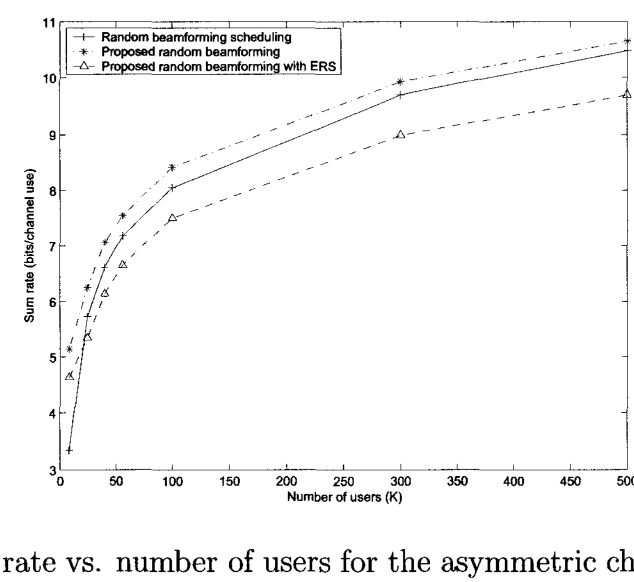

Figure 3.10: Sum rate vs. number of users for the asymmetric channel with eight BS antennas, $P = 20$dB.

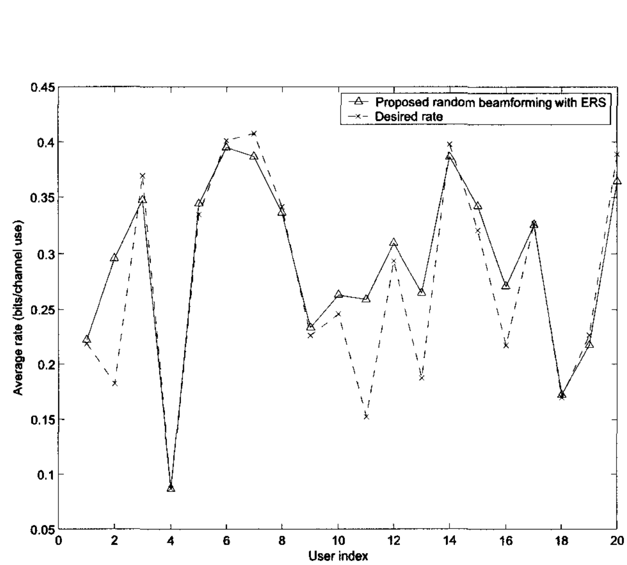

Figure 3.11: Average rate of each user with different desired rate at the 600th time slot in slow fading and asymmetric channel, $P = 20$dB.

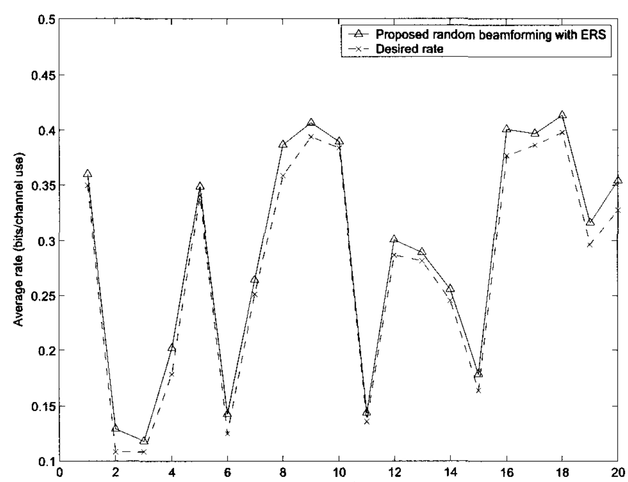

Figure 3.12: Average rate of each user with different desired rate at the 600th time slot in fast fading and asymmetric channel, $P = 20$dB.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

In this thesis, BS transmitter design and downlink transmission schemes are considered. The problems of bit error probability minimization in the MIMO broadcast channel under the peak and average power constraints are formulated as convex optimization problems, which can be efficiently solved using an interior point algorithm. It has been shown that the solutions of these convex optimization problems are globally optimal unless the SNR is very low or the channel is nearly singular. Simulation results show that the proposed approaches significantly improve the performance compared to the existing methods of [23] and [14]. As a by-product, the exact probability of bit error can be obtained after solving the optimization problems.

Another contribution in this thesis is that we propose an algorithm to decrease the amount of feedback required to exploit multiuser diversity in MISO downlink transmission. Compared to the previous approaches, the necessary average feedback rate of the proposed random beamforming is fixed and does not increase with the number of users. A lower and upper bound of the sum rate of the proposed random beamforming are derived to evaluate its performance. Simulation results show that the sum rate of the proposed random beamforming is better than that of random beamforming scheduling when the number of users is small.

An equal rate scheduling is also proposed to serve users with different rate requirements. This scheduling algorithm can be regarded as allocating the time slots proportional to the users' rate requirements. Simulation results demonstrate that despite a little sum rate loss, the proposed random beamforming with equal rate scheduling roughly satisfies the rate requirements of each user both in slow and fast fading channels.

4.2 Future Work

Since the work of [4], there has been a lot of research on the broadcast channel. However, none of the approaches achieve its capacity. A recent work [8] combine turbo coding and vector quantization to present a realization of multidimensional dirty paper coding scheme. The gap to capacity of such systems at low SNR is large. The bottleneck lies in the quantization code. To further improve the performance, better multidimensional quantization code which achieve the ultimate shaping gain (1.53dB) needs to be found.

Another interesting topic is the application of semidominant programming in the approach of maximum-likelihood detection. There has been a lot of work in this area. It is shown in a recent result [17] that the performance of two semi-definite relaxation model is pretty close to that of the maximum-likelihood detection.

Bibliography

[1] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications.* Boca Raton, FL: CRC Press, Aug. 2001.

[2] M. Biguesh, S. Shahbazpanahi, and A. B. Gershman, "Robust power adjustment for transmit beamforming in cellular communication systems," in *Proc. ICASSP'03*, 2003.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge University Press, 2004.

[4] G. Caire and S. S. Shamai, "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Trans. Info. Theory*, vol. 49, pp. 1691–1706, July 2003.

[5] L.-U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," vol. 3, pp. 20–24, Jan. 2004.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] H. A. David, *Order Statistics*, 2nd ed. New York: John Wiley & Sons, 1981.

[8] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Info. Theory*, vol. 51, pp. 3417–3432, Oct. 2005.

[9] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.: Kluwer Academic Press*, vol. 6, pp. 311–335, Mar. 1998.

[10] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Info. Theory*, vol. 48, pp. 1473–1484, June 2002.

[11] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication-Part II: perturbation," *IEEE Trans. Commun.*, vol. 53, pp. 537–544, Mar. 2005.

[12] E. S. Hons, A. K. Khandani, and W. Tong, "An optimized transmitter precoding scheme for synchronous DS-CDMA," in *Proc. ICC*, New York, May 2002, pp. 1818–1822.

[13] W. M. Jang, B. R. Vojcic, and R. L. Pickholtz.

[14] E. A. Jorswieck and H. Boche, "Transmission strategies for the MIMO MAC with MMSE receiver: average MSE optimization and achievable individual MSE region," *IEEE Trans. Signal Processing*, vol. 51, pp. 2872–2881, Nov. 2003.

[15] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. ICC*, June 1995, pp. 331–335.

[16] Z.-Q. Luo, T. N. Davidson, G. B. Giannakis, and K. M. Wong, "Transceiver optimization for block-based multiple access through isi channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 1037–1052, Apr. 2004.

[17] A. Mobasher, M. Taherzadeh, R. Sotirov, and A. K. Khandani, "A near maximum likelihood decoding algorithm for MIMO systems based on semi-definite programming," *IEEE Trans. Info. Theory*, Aug. 2005, (Submitted).

[18] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* New York: John Wiley & Sons, 1982.

[19] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Select. Areas Commun.*, vol. 16, 1998.

[20] D. Reynolds, A. Host-Madsen, and X. Wang, "Adaptive transmitter precoding for time division duplex cdma in fading multipath channels: strategy and analysis," *EURASIP Joural of App. Sig. Proc.*, vol. 2002, pp. 1325–1334, Dec. 2002.

[21] D. Reynolds, X. Wang, and K. N. Modi, "Interference suppression and diversity exploitation for multi-antenna cdma with ultra-low complexity receiver," *IEEE Trans. Signal Processing*, vol. 53, pp. 3226–3237, Aug. 2005.

[22] K. H. Rosen, *Discrete Mathematics and Its Applications*, 5th ed. New York: McGraw-Hill, 2003.

[23] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004.

[24] ——, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Signal Processing*, vol. 53, pp. 2324–2334, July 2005.

[25] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Info. Theory*, vol. 51, pp. 506–522, Feb. 2005.

[26] S. Shi and M. Schubert, "MMSE transmit optimization for multi-user multi-antenna systems," in *Proc. ICASSP'05*, vol. 3, Philadelphia, USA, Mar. 2005, pp. 409–412.

[27] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 461–471, Feb. 2004.

[28] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[29] S. Verdu, *Multiuser Detection.* New York: Cambridge Univ. Press, 1998.

[30] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. Info. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[31] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Info. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.

[32] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE J. Select. Areas Commun.*, vol. 48, pp. 1277–1294, June 2002.

[33] B. R. Vojcic and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, vol. 46, pp. 1346–1355, Oct. 1998.

[34] X. Wang, W.-S. Lu, , and A. Antoniou, "Constrained minimum-BER multiuser detection," *IEEE Trans. Signal Processing*, vol. 48, pp. 2903–2909, Oct. 2000.

[35] C. Windpassinger, R. F. Fischer, T. Vencel, and J. B. Huber, "Precoding in multi-antenna and multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1305–1306, July 2004.

[36] K.-K. Wong, R. D. Murch, and K. B. Letaief, "Performance enhancement of multiuser MIMO wireless communication systems," *IEEE Trans. Commun.*, vol. 50, pp. 1960–1970, Dec. 2002.

[37] ——, "A joint-channel diagonalization for multiuser MIMO antenna systems," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 773–786, July 2003.

[38] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Info. Theory*, vol. 50, pp. 1875–1892, Sept. 2004.

[39] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Info. Theory*, vol. 49, pp. 1073–1096, May 2003.