MULTIPLE ANTENNA BROADCAST CHANNELS WITH RANDOM CHANNEL SIDE INFORMATION

by

Alon Shalev Housfater

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2011 by Alon Shalev Housfater
Abstract

Multiple Antenna Broadcast Channels with Random Channel Side Information

Alon Shalev Housfater
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2011

The performance of multiple input single output (MISO) broadcast channels is strongly dependent on the availability of channel side information (CSI) at the transmitter. In many practical systems, CSI may be available to the transmitter only in a corrupted and incomplete form. It is natural to assume that the flaws in the CSI are random and can be represented by a probability distribution over the channel. This work is concerned with two key issues concerning MISO broadcast systems with random CSI: performance analysis and system design. First, the impact of noisy channel information on system performance is investigated. A simple model is formulated where the channel is Rayleigh fading, the CSI is corrupted by additive white Gaussian noise and a zero forcing precoder is formed by the noisy CSI. Detailed analysis of the ergodic rate and outage probability of the system is given. Particular attention is given to system behavior at asymptotically high SNR. Next, a method to construct precoders in a manner that accounts for the uncertainty in the channel information is developed. A framework is introduced that allows one to quantify the tradeoff between the risk (due to the CSI randomness) that is associated with a precoder and the resulting transmission rate. Using ideas from modern portfolio theory, the risk-rate problem is modified to a tractable mean-variance optimization problem. Thus, we give a method that allows one to efficiently find a good precoder in the risk-rate sense. The technique is quite general and applies to a wide range of CSI probability distributions.
Acknowledgements

This thesis is the culmination of four years of research work at the University of Toronto. Now that this work is done, I feel incredibly fortunate for having had the opportunity to write this thesis.

I am deeply grateful to my advisor, Professor Teng Joon Lim, for guiding me in my research work and in the writing of this thesis. I thank him for his support and guidance. I have benefited tremendously from his experience, knowledge and insight. I would also like to acknowledge the members of my thesis committee: Prof. Frank Kschischang, Prof. Wei Yu and Prof. Xianbin Wang, for their insightful comments and encouragement.

I would like to thank my family for their patience, affection and help. Last but not least, I thank Viky Gu Fang, my girlfriend and soulmate, for her love, support and faith in me.

Alon Shalev Housfater

Toronto, Canada

July, 2011
Contents

1 Introduction .................................................. 1
  1.1 Motivation ................................................ 2
    1.1.1 Channel Estimation .................................. 3
    1.1.2 Quantization ......................................... 3
    1.1.3 Delay ................................................ 4
  1.2 Outline of the Thesis .................................... 4
    1.2.1 Zero Forcing with Gaussian CSI Noise ............... 4
    1.2.2 Precoder Design for Random CSI ...................... 5
    1.2.3 Additional Results .................................. 6
  1.3 Notation .................................................. 6

2 Broadcast Channels and Transmitter Side Processing .......... 8
  2.1 The MISO Broadcast Channel and its Capacity ............. 8
  2.2 Precoders and Transmitter Side Processing ................ 10
    2.2.1 Linear Precoding ................................... 11
    2.2.2 Cross Rate Precoding ................................ 14
  2.3 Random Channel Side Information .......................... 22
    2.3.1 Random CSI and Fading ............................... 22
    2.3.2 Robust Precoders .................................... 26
    2.3.3 Bounds on Interference Power ......................... 27
3 Zero Forcing with Noisy Channel Side Information

3.1 Transmitter Precoding with Noisy CSI
- 3.1.1 Power Control Policy
- 3.1.2 The Effective Channel and its Distribution
- 3.1.3 Semicorrelated Channels

3.2 Summary

3.3 Appendices
- 3.3.1 Appendix: Theorem 3
- 3.3.2 Appendix: Theorem 4

4 Performance of Zero-Forcing Precoder with Noisy CSI

4.1 The Ergodic Rate
- 4.1.1 Limiting Cases: Feedback Variance and the Ergodic Sum Rate
- 4.1.2 Simulations

4.2 The Ergodic Rate at High SNR
- 4.2.1 The Rate Penalty
- 4.2.2 Simulations

4.3 Outage Probability
- 4.3.1 Outage Probability at High SNR
- 4.3.2 Simulations

4.4 Summary

4.5 Appendices
- 4.5.1 Appendix: Theorem 8
- 4.5.2 Appendix: Theorems 10 & 9
- 4.5.3 Appendix: Theorem 12
List of Figures

2.1 % of Maximal linear rate achieved by both regularized zero forcing and cross rate precoders .................................................. 23

2.2 % of Maximal linear rate achieved by both regularized zero forcing and single iteration optimization of the cross rate .................. 23

3.1 Block diagram for ZF precoding with noisy CSI ............................ 40

4.1 Ergodic rate vs. feedback variance ............................................. 58

4.2 Rate Penalty vs. number of transmit antennas for $p = 2, 4, 6$ ........ 64

4.3 $\log(P_{out})$ vs. number of transmit antennas ............................... 68

4.4 Outage probability vs. feedback variance .................................... 68

5.1 Precoder performance for a two component GMM channel ............. 89
Chapter 1

Introduction

The discovery of the capacity of the Gaussian multiple input multiple output (MIMO) broadcast channel was a triumph of multi-user information theory. The Gaussian MIMO broadcast channel is a fundamental multi-user model where a single transmitter communicates with multiple receivers. The key phenomenon in the broadcast channel is the presence of interference between users. A crucial step in the calculation of the capacity of the MIMO broadcast channel was the realization that this interference can be mitigated by using dirty paper coding (DPC). DPC is a successive precancellation scheme that removes the interference seen by the different receivers in a systematic and optimal manner. However, in order to perform the interference precancellation, the transmitter requires exact knowledge of the channel. From a practical standpoint, the availability of exact channel side information (CSI) at the transmitter is not always possible or desirable. Acquiring exact CSI can be quite taxing on system resources, degrading overall system performance. Also, errors in the CSI may be caused by inherent limitations of the system (such as noise). The aim of this thesis is to investigate the impact of imperfect CSI on system performance. We will do so by addressing two specific questions: What is the impact of CSI error on a system designed with the assumption of perfect CSI? And how should a precoder be designed so as to maximize system performance while

1
accounting for random CSI?

Understanding the impact of CSI errors poses several distinct challenges. First, the notion of imperfect CSI must be rigorously defined by selecting a mathematical model that describes the CSI and its flaws. Selecting an appropriate CSI model for study is delicate work: one must select a model powerful enough to reflect reality while sufficiently simple to admit rigorous analysis. In many natural models, the CSI error propagates through the system in a manner that is difficult to analyze. On the other hand, too simple a model will not give one insights into the performance of real systems. An additional difficulty is the lack of a rigorous information theoretic treatment of incomplete channel side information. Therefore, unlike the more classic subjects of communication and information theory, it is not always clear what the appropriate performance measure should be.

The thesis examines two fundamental topics: the effect of CSI imperfection on system performance and the design of precoders adapted to take account of CSI flaws. The effect of CSI errors on system performance is studied by analyzing a simple model where the CSI imperfection is modeled as additive white Gaussian noise. This noisy CSI is used by the transmitter to form a zero forcing precoder. The outage probability and ergodic sum rate of this system are studied in extensive detail. The design problem is considered in the case where the CSI errors are modeled by some probability distribution. A flexible design methodology of a risk-rate tradeoff is introduced and used to construct an algorithm that can find a good linear precoder for a wide range of CSI error models.

1.1 Motivation

The study of uncertain CSI is especially relevant to the design of future wireless cellular networks. In this setting, the receivers are mobile users and the transmitter is a base station. Several examples will be presented, these illustrate some mechanisms that can
generate imperfect CSI: channel estimation in time division duplex (TDD) systems and quantization and delay in a frequency division duplex (FDD) systems. Furthermore, in practical communication systems some of these mechanisms will be present concurrently and interact in a complex manner to generate imperfect CSI.

1.1.1 Channel Estimation

In TDD mobile cellular systems, a single channel is used for a two way communication between the mobiles and the base station. The mobiles transmit pilot symbols over the uplink channel to the base station which in turn uses these to estimate the channel. However, since the pilot symbols are transmitted over a noisy wireless channel, the base station’s estimation will also be corrupted by noise. This problem can be mitigated to some extent by increasing the number of pilot symbols and allowing the base station to average away the errors. However, increasing the number of pilot symbols will reduce the overall system performance. Thus, the base station must use CSI that will be noisy due to the noise in the pilot symbols reverse channel and a limited number of pilot symbols.

1.1.2 Quantization

In contrast to TDD systems, FDD systems use different frequency bands for uplink and downlink. Thus, the base station cannot estimate the channel from pilot signals sent by the mobiles. Instead, the mobile users must each estimate their own channel from pilot symbols sent by the base station. The estimated channel is quantized and sent to the base station. This quantization is necessary to limit the impact on system throughput due to the feedback overhead [49]. The quantization can introduce significant error into the CSI, with the error determined by the number of bits used to quantize the channel. Also, additional errors are possible due to imprecise estimation by the mobile. Thus, the CSI will be corrupted by several sources of error; these should be accounted for when the CSI is used by the base station.
1.1.3 Delay

In wireless channels with a fast changing channel, delay in the acquisition of CSI by the base station is often a significant source of error. This is an issue in systems where the receivers estimate the channel and transmit it back to the base station [50]. Delay in CSI availability is incurred due to the length of time required for accurate transmission on the reverse channel. Therefore, for a long delay and a fast changing channel, the CSI available at the base station will be outdated. However, depending on the specific conditions of the system, the outdated CSI will still contain some useful information on the channel.

1.2 Outline of the Thesis

The thesis is organized as follows: Chapter 2 introduces some concepts and notation and surveys the relevant literature. Also, it contains some new contributions on the design of a cross rate precoder and bounds on interference power. Chapter 3 introduces the model used for the analysis of the performance impact of CSI noise and proves some fundamental consequences. Chapter 4 proceeds with in-depth performance analysis. Chapter 5 formulates the precoder design problem for random CSI and derives an algorithm to solve it. Detailed summary of the major contributions is given below.

1.2.1 Zero Forcing with Gaussian CSI Noise

In Chapters 3 and 4, the effect of noisy CSI on system performance is analyzed by considering a simple model: a zero forcing precoder is formed from CSI corrupted by AWGN while the channel is taken to be Rayleigh fading. Two performance measures are used, the ergodic sum rate and the outage probability. A computationally efficient formula is given for the ergodic sum rate, which allows one to efficiently evaluate the sum rate for various system configurations. Also, the effect of the CSI noise power on the sum
rate is studied in detail. It is shown that, asymptotically, one must reduce the CSI noise power as the SNR increases in order to prevent the system from becoming interference limited (see also [40]). Under such CSI noise power scaling, the multiplexing gain and rate penalty are calculated in closed form.

Similarly, it is shown that the outage probability can be computed efficiently for all system configurations. It is proven that for a fixed CSI noise power, the system will become interference limited with increasing SNR. However, so long as the CSI noise power is scaled downward with increasing SNR, the outage probability will converge to zero. The decay rate of the outage probability, a quantity analogous to the diversity order of single user fading systems, is given in closed form.

1.2.2 Precoder Design for Random CSI

In Chapter 5, the risk-rate framework for the design of precoders with random CSI is introduced. The uncertainty of the channel (from the transmitter’s perspective) due to imperfect CSI is modeled by a probability distribution over the channel. The risk is defined as the probability of the sum rate being below a fixed rate threshold. This is a concept analogous to the outage probability of fading channels. It is proposed to design a linear precoder by finding the precoding matrix that will minimize risk for a fixed rate or maximize rate for a fixed risk. Since the risk-rate design problem is numerically intractable due to the high dimensional integrations, a related mean-variance optimization problem inspired by ideas from modern portfolio theory is formulated. Next, the CSI probability distribution is assumed to be of a Gaussian mixture form, a general and flexible parametric distribution family. This assumption, coupled with mean-variance optimization, allows one to make some analytic simplifications and derive an algorithm for finding good precoders. Due to the generality and flexibility of the algorithm, it can be used by a system designer to tackle complex CSI error behavior.
1.2.3 Additional Results

Some other contributions are given in Chapter 2. The linear precoder design problem is considered for the two user, two transmit antenna broadcast channel configuration. A new design metric, the cross rate, is introduced as a nonlinear zero-forcing criterion. An efficient optimization algorithm is given for the calculation of the cross-rate optimal precoder. Simulations indicate that this precoder performs better than the regularized zero forcing precoder and comes close to the performance of the optimal, rate maximizing, linear precoder. Also in Chapter 2, a Cramér-Rao type bound on the performance of any linear precoder for an arbitrarily fading broadcast system is derived. The CSI is assumed to be stochastic but the dependency on the channel is arbitrary. It is shown that this bound can be easily evaluated in closed form for a variety of fading distributions.

1.3 Notation

We denote matrices by upper case letters in boldface type (e.g. \( \mathbf{A} \)) and the \((i,j)\)th entry of \( \mathbf{A} \) is written as \( A_{ij} \). The \( i \)th column vector of a matrix \( \mathbf{A} \) is written as the lower case letter in boldface type \( \mathbf{a}_i \), and the \( i \)th row vector of the matrix \( \mathbf{A} \) is written as \( \mathbf{a}_i^\top \). The trace and determinant of \( \mathbf{A} \) are denoted by \( \text{tr} (\mathbf{A}) \) and \( \text{det} (\mathbf{A}) \), respectively. Also, define \( \text{etr} (\mathbf{A}) = \exp (\text{tr} (\mathbf{A})) \) to be the exponential trace of the matrix \( \mathbf{A} \). The matrix \( \mathbf{I} \) is the identity matrix, and sometimes we will emphasize its dimension by writing \( \mathbf{I}_m \) to denote the \( m \times m \) identity matrix. Also, let \( \mathbb{C}^d \) denote the set of \( d \)-dimensional complex vectors and let \( \mathbb{C}^{m \times n} \) denote the set of \( m \times n \) matrices with complex entries. Writing \( \mathbf{A} > 0 \) implies \( \mathbf{A} \) is a positive definite Hermitian matrix. The magnitude of a complex number \( c \) is written as \(|c|\). The Gamma function, a generalization of the factorial to non-integer values, is given in its integral form as

\[
\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt,
\]
for any $z \in \mathbb{C}$ except the non-positive integers. The complex multivariate Gamma function is given as,

$$
\Gamma_m(\alpha) = \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(\alpha - (i - 1))
$$

for $\text{Re}\{\alpha\} > m - 1$. We will also sometimes use the integral formula [34],

$$
\Gamma_m(\alpha) = \int_{S > 0} \det(S)^{\alpha - m} \text{etr}(-S) \, dS
$$

where the integration is over the cone of $m \times m$ positive definite Hermitian matrices. One form of the Beta function is,

$$
B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x + y)}.
$$

The distribution of a matrix is defined as the joint distribution of its elements. The expectation operator, denoted as $\mathbb{E}\{\cdot\}$, is taken with respect to all random variables in the expression unless stated otherwise. We denote that $z$ is distributed as a complex circularly symmetric standard Gaussian random variable by $z \sim \mathcal{N}_c(0, 1)$. Similarly, the Wishart and inverse Wishart distributions [36] are denoted by the symbols $\mathcal{W}_c$ and $\mathcal{W}_c^{-1}$, respectively. We will also use standard asymptotic notation: Given $g(x) \neq 0$, we write $f(x) = \Theta(g(x))$ as $x \to x_0$ if $f(x)/g(x)$ is bounded as $x \to x_0$ and $f(x) = o(g(x))$ as $x \to x_0$ if $f(x)/g(x) \to 0$ as $x \to x_0$. We will sometimes use the asymptotic notation $f(x) \asymp g(x)$ to indicate that $f(x)/g(x)$ tends to unity as $x$ increases.
Chapter 2

Broadcast Channels and Transmitter Side Processing

2.1 The MISO Broadcast Channel and its Capacity

In this work, we will be concerned with a multi-input single-output broadcast (MISO) channel. The MISO channel consists of $n$ transmit antennas and $p$ users, where each user is equipped with a single antenna. The relationship between the inputs and outputs is modeled as a linear channel with additive white Gaussian noise (AWGN). A single channel use (by all users concurrently) is modeled by the equation,

$$ y = Hx + n_e, $$

(2.1)

where $x \in \mathbb{C}^n$ is an $n$-dimensional vector to be transmitted, $H \in \mathbb{C}^{p \times n}$ is the channel matrix and $y \in \mathbb{C}^p$ is the $p$-dimensional vector of received information. Each entry in $y$ is received by a different user and the users cannot cooperate while decoding their messages. The noise vector $n_e$ is distributed as a circularly symmetric $p$-dimensional complex Gaussian random vector, $n_e \sim \mathcal{N}_c(0, I)$. Equivalently, the channel of a single
user is given by,

\[ y_i = \mathbf{h}_i \mathbf{x} + n_e, \quad i = 1, \ldots, p. \]

The \(1 \times n\) row vector \(\mathbf{h}_i\) corresponds to the \(i\)th row of the channel matrix \(\mathbf{H}\) and the noise \(n_e\) is a scalar Gaussian random variable with mean zero and variance 1. Thus, we observe that all \(p\) channels of the individual users are coupled by the transmission vector \(\mathbf{x}\). Also, the transmitted vector is power constrained, \(\mathbb{E}\{\|\mathbf{x}\|^2\} \leq P\).

Assuming that the channel \(\mathbf{H}\) is known to the transmitter and all receivers, it is known that dirty paper coding is capacity achieving. The dirty paper coding achievable region is given in [4] and the converse is proved in [1]. DPC was first developed in [2] for the single user Gaussian channel with known interference and subsequently extended in [4]. Consider the single user channel with interference,

\[ y = x + s + n, \]

where \(y\) and \(x\) are the received and transmitted vectors, respectively. It can be shown that if \(n\) and \(s\) are both i.i.d. vector Gaussian processes where non-casual knowledge of \(s\) is available at the transmitter, then the capacity of the channel is the same as if \(s\) is not present and is achieved by dirty paper coding. In the setting of the MIMO broadcast channel, this result can be used iteratively to construct a vector that achieves capacity. The transmitter first picks a codeword for receiver 1 from a codebook with covariance \(\Sigma_1\). Next, the transmitter picks a codeword for receiver 2 from a codebook with covariance \(\Sigma_2\). Since the transmitter knows the channel and the symbol intended for receiver 1, it has knowledge of the interference caused by receiver 1 and can choose the codeword for receiver 2 so as to remove the interference caused by it. This can be done in an iterative manner so that the transmitter picks a codeword from a codebook with covariance \(\Sigma_k\) for the \(k\)th receiver such that the interference from the previous \(k-1\) receivers is presubtracted [3]. The transmitted signal is then \(x = x_1 + \ldots + x_p\) where
\( \mathbf{x}_k \) is the codeword for the \( k \)th receiver. Note that the order at which this interference presubtraction is done is arbitrary and we represent this order by a permutation of the integers \( 1, \ldots, p \), denoted by \( \pi \). We write \( R_{\pi(i)} \) to denote the rate achieved by the \( i \)th user for a given encoding order \( \pi \), where \( \pi(i) \) is the user index of the \( i \)th encoded user,

\[
R_{\pi(i)} = \log \left( \frac{1 + \mathbf{h}_{\pi(i)} \left( \sum_{j \geq i} \Sigma_{\pi(j)} \right) \mathbf{h}_{\pi(i)}^*}{1 + \mathbf{h}_{\pi(i)} \left( \sum_{j > i} \Sigma_{\pi(j)} \right) \mathbf{h}_{\pi(i)}^*} \right).
\]

We collect the rates that are simultaneously achieved into a vector, \( \mathbf{R}(\Sigma_1, \ldots, \Sigma_p, \pi) \) where \( \Sigma_1, \ldots, \Sigma_p \) are the input covariance matrices as described and \( \pi \) is the encoding order. The capacity region is then given as the convex hull of the union of all rate vectors over all input covariance matrices \( \Sigma_1, \ldots, \Sigma_p \) such that \( \text{tr} (\Sigma_1 + \ldots \Sigma_p) > P \) and all encoding orders \( \pi \),

\[
\text{co} \left( \bigcup_{\pi} \bigcup_{\text{tr}(\Sigma_1 + \ldots \Sigma_p) > P} \mathbf{R}(\Sigma_1, \ldots, \Sigma_p, \pi) \right).
\]

### 2.2 Precoders and Transmitter Side Processing

The information theoretic derivation of the broadcast channel capacity suggests that *transmitter side processing*, specifically interference pre-subtraction, is crucial in achieving good performance. However, DPC is complicated to implement and as a result many precoders of reduced complexity have been proposed. Precoders are usually distinguished by two design points: the precoder structure and the performance criteria. The two main types of precoder structures in the literature are Tomlinson-Harashima-type precoders (THP) and linear precoders. First introduced in [6] and [7], THP was first used in the single-user intersymbol interference (ISI) channel for subtracting the interference caused by past symbols (see [8] and also [9] for a comprehensive introduction). In [5] and [4] it was observed that THP can be interpreted as an approximation for DPC. Linear precoders are simpler to implement and are performed by a matrix multiplication of the
symbol vector. The possible performance criteria for both THP and linear precoders include: signal to interference and noise ratio (SINR) constraint [14], interference power suppression [16][17], MMSE [26], bit error rate constraints [25], maximal sum-rate [23] and others.

2.2.1 Linear Precoding

Linear precoding is one of the standard approaches to the design of communication systems for MIMO BC. Its power lies in its simplicity of implementation and good performance [15]. In linear precoding, one considers a vector of independently selected symbols \( s \in \mathbb{C}^p \), where the \( i \)th entry is a symbol intended for the \( i \)th user, which is multiplied by a precoding matrix \( D \in \mathbb{C}^{n \times p} \). Mathematically we write

\[
y = HDs + n_e, \tag{2.2}
\]

where we now assume \( n_e \sim \mathcal{N}(0, \sigma^2_eI) \). We observe that the vector \( Ds \) is the transmitted vector \( x \) in equation (2.1). The constraint of constructing the transmission vector by matrix multiplication allows one to implement the transmitter efficiently. We can determine the rates achievable by a fixed linear precoder in the following manner. First, consider a non-precoded system (i.e. \( D = I \)), the weighted sum-rate achievable by such a system can be written,

\[
R_\omega (H) = \sum_{i=1}^{p} \omega_i \log_2 \left( 1 + \frac{h_{ii}^2}{\sigma^2_e + \sum_{j \neq i} h_{ij}^2} \right), \tag{2.3}
\]

where \( \omega \in \mathbb{R}^p \) is a \( p \)-dimensional weighting vector \( \omega = [\omega_1, \ldots, \omega_p]^T \) whose entries are non-negative and sum to one. The weighted sum-rate can be achieved by transmitting to each user using an independent Gaussian codebook. Moreover, for any fixed \( D \), the sum-rate is now \( R_\omega (HD) \). Thus, assuming the symbols \( s \) are independently selected from
a zero-mean unit power Gaussian codebook, the transmitted power can be calculated,

$$\mathbb{E}\{\text{tr}(s^*D^*Ds)\} = \text{tr}(D^*D).$$

Hence, the average power constraint restricts the set of permissible precoders, $\mathcal{F}_P$, to be those whose power is less than or equal to a fixed power budget $P$,

$$\mathcal{F}_P = \left\{ D \in \mathbb{C}^{n \times p} : \text{tr}(D^*D) \leq P \right\}. \tag{2.4}$$

It is natural to use the maximum sum-rate precoder (MRP), given as the solution to the optimization problem,

$$D^{MR} = \arg \max_{D \in \mathcal{F}_P} R_\omega(\mathbf{H}D).$$

Unfortunately, solving this optimization is hard as the problem is non-convex. A simpler and well known alternative is is the zero forcing precoder (ZFP),

$$D^{ZF} = \sqrt{\frac{P}{\text{tr}\left((\mathbf{H}\mathbf{H}^*)^{-1}\right)}}\mathbf{H}^* (\mathbf{H}\mathbf{H}^*)^{-1}.$$

This precoder will remove interference from all receivers, i.e. zero-force the interference. The normalization factor of $\frac{1}{\text{tr}(\mathbf{H}\mathbf{H}^*)^{-1}}$ is required in order to satisfy the transmitter power constraint. However, due to the power normalization factor, the ZFP may suffer from noise enhancement, especially when the channel is close to singular. An alternative to the ZFP is the regularized ZFP (R-ZFP) [17] which aims to reduce the noise enhancement by introducing a regularizing factor into the precoding matrix,

$$D^{RZF} = \sqrt{\frac{1}{z}}\mathbf{H}^* (\mathbf{H}\mathbf{H}^* + \lambda I)^{-1}$$

where $z$ is a normalization constant chosen so that the transmitter power constraint is maintained. This precoding scheme is effective for near singular and square (i.e. same
number of transmit antennas as users) channels. The intuition behind the introduction of the regularizing factor $\lambda I$ is to enable the transmitter to control the singularity of the precoding matrix and the power normalization by this factor, thus reducing the noise enhancement problem. In [17], the value of $\lambda$ is chosen by letting $H$ be Rayleigh fading and finding $\lambda$ that will maximize an approximation to the expected SINR. The maximizing value is shown to be $\lambda = p\sigma_e^2$, under an assumption that the total transmitted power is one. A similar precoder structure is the minimum mean squared error (MMSE) precoder [53]. Here one introduces the mean squared error,

$$e(D) = \mathbb{E}\{\|s - HDs + n_e\|^2\},$$

where the expectation is over the symbols and noise. The MMSE precoder is the precoding matrix $D$ that will minimize the mean squared error subject to the power constraint $\mathbb{E}\{\|Ds\|^2\} = P$. It can be shown that the MMSE precoder is,

$$D^{MMSE} = \sqrt{\frac{P}{z}} H^* \left( HH^* + \frac{p\sigma_e^2}{P} I \right)^{-1},$$

where $z$ is a power normalization factor. Finally, one can note that for large SNR, the MMSE precoder will become the zero forcing precoder. Both the MMSE and regularized ZF precoders introduce a scalar factor $\lambda$ into the precoder, but a related MMSE precoder design given in [52] uses vector scaling. There exist other types of linear precoding designs, for example in [16], where a ZF precoder that maximizes the sum-rate is designed using a convex relaxation approach and [14] were a linear precoder is designed to satisfy SINR constraints. In the following section, a linear precoder is designed that uses a modification of the sum-rate, a problem which is shown to be efficiently solvable.
2.2.2 Cross Rate Precoding

The cross rate precoder is a specialized design, published by the author in [18], that is suitable for the two-user broadcast channel \( p = 2 \) with a transmitter equipped with two antennas. Here we give details of this precoder design. Noting the system model (2.2), the effective channel matrix between the symbol vector \( s \) and the received vector \( y \) is \( G = HD \) where,

\[
\begin{pmatrix}
\vec{h}_1 d_1 & \vec{h}_1 d_2 \\
\vec{h}_2 d_1 & \vec{h}_2 d_2
\end{pmatrix}
\]

The vectors \( \vec{h}_1, \vec{h}_2 \) are row vectors of the matrix \( H \) and the vectors \( d_1, d_2 \) are column vectors of the precoding matrix \( D \). We now observe that the achievable rates (in bits per channel use) by both users are,

\[
R_1 = \log_2 \left( 1 + \frac{|\vec{h}_1 d_1|^2}{N_0 + |\vec{h}_1 d_2|^2} \right)
\]

\[
R_2 = \log_2 \left( 1 + \frac{|\vec{h}_2 d_2|^2}{N_0 + |\vec{h}_2 d_1|^2} \right).
\]

These formulae hold when the first entry of \( s \) contains symbols intended for user one and the second entry of \( s \) contains symbols intended for the second user. As mentioned previously, a natural design criterion for the precoding matrix \( D \) is then the sum-rate. In other words, the transmitter should use the precoding matrix that maximizes the sum-rate (subject to power constraints),

\[
\max_{D \in F_p} \log_2 \left( 1 + \frac{|\vec{h}_1 d_1|^2}{N_0 + |\vec{h}_1 d_2|^2} \right) + \log_2 \left( 1 + \frac{|\vec{h}_2 d_2|^2}{N_0 + |\vec{h}_2 d_1|^2} \right).
\]

(2.5)

This problem can be solved using non-convex optimization methods. However, we seek a more efficient algorithm that does not require such techniques. This can be done by considering the rates achievable by the transposed system. By this we mean permuting
the entries of the vector $s$ so that the first entry of $s$ contains symbols intended for user two while the second entry of $s$ contains symbols intended for user one. Thus, we have the transposed rates,

$$R_1^* = \log_2 \left( 1 + \frac{|\tilde{h}_1 d_2|^2}{N_0 + |\tilde{h}_1 d_1|^2} \right)$$

$$R_2^* = \log_2 \left( 1 + \frac{|\tilde{h}_2 d_1|^2}{N_0 + |\tilde{h}_2 d_2|^2} \right).$$

where the rates $R_1^*$, $R_2^*$ are achievable by the first and second users, respectively. It is desirable to maximize the differences $R_1 - R_1^*$ and $R_2 - R_2^*$. Heuristically, if $R_1$ and $R_1^*$ are nearly the same it implies the channel entries $\tilde{h}_1 d_1$ and $\tilde{h}_1 d_2$ have similar magnitude. In other words, the channel coefficient carrying information ($\tilde{h}_1 d_1$) and the one carrying interference ($\tilde{h}_1 d_2$) have similar magnitudes. This implies a low rate. Therefore, to maximize the rate difference of both users jointly, we take as a design criteria the formula,

$$\Upsilon = R_1 - R_1^* + R_2 - R_2^* = R_1 + R_2 - (R_1^* + R_2^*) = -\log_2 \left( \frac{N_0 + |\tilde{h}_1 d_2|^2}{N_0 + |\tilde{h}_1 d_1|^2} \right) - \log_2 \left( \frac{N_0 + |\tilde{h}_2 d_1|^2}{N_0 + |\tilde{h}_2 d_2|^2} \right).$$

Note that by the second equality, one may consider the cross rate, $\Upsilon$, given above as the difference between the sum rate of the system and its corresponding transpose. In order to find a cross-rate optimal precoder $D_{CR}$, we solve the optimization problem,

$$\min_{D \in \mathcal{F}_P} \log_2 \left( \frac{N_0 + |\tilde{h}_1 d_2|^2}{N_0 + |\tilde{h}_1 d_1|^2} \right) + \log_2 \left( \frac{N_0 + |\tilde{h}_2 d_1|^2}{N_0 + |\tilde{h}_2 d_2|^2} \right).$$

(2.6)

Similar to (2.5), the cross rate optimization problem is a non-convex one. However, we are able to obtain the global optimum by a reduction to a sequence of linear programs.
First, note that any precoding matrix, \( D \in \mathcal{F}_P \), can be written as the product,

\[
D = SA,
\]

where the column vectors of \( S \) have unit norm and the power allocation matrix, \( \Lambda \), is,

\[
\Lambda = \begin{pmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{pmatrix}
\]

where \( \lambda_1 \) and \( \lambda_2 \) signify the power allocated to user 1 and user 2, respectively. Our approach is then to optimize the power allocations \( \lambda_1 \), \( \lambda_2 \) and the steering matrix \( S \) in an alternating fashion.\(^1\) Thus, we are given with initial steering and power allocation matrices \( S_0, \Lambda_0 \) and perform the iterations,

\[
S_i \leftarrow \max_{S \in S} \Upsilon (H, S, \Lambda_{i-1}) \tag{2.7}
\]

\[
\Lambda_i \leftarrow \max_{\Lambda \in D_P} \Upsilon (H, S_i, \Lambda) \tag{2.8}
\]

where \( S \) is the set of matrices with unit norm columns and \( D_P \) is the set of diagonal matrices such that the diagonal entries \( \lambda_1^2 + \lambda_2^2 \leq P \). At first glance, we have not gained much as we need to solve problems (2.7) and (2.8) which are again non-convex. In the following, we show that these can be solved efficiently.

### Alternating Optimization Algorithm for the Cross Rate Precoder

We can rewrite problem (2.8) as,

\[
\min_{\lambda_1^2 + \lambda_2^2 \leq P} \left( \frac{N_0 + \lambda_2^2 \eta_{12}}{N_0 + \lambda_2^2 \eta_{22}} \right) \left( \frac{N_0 + \lambda_1^2 \eta_{21}}{N_0 + \lambda_1^2 \eta_{11}} \right) \tag{2.9}
\]

\(^1\)This strategy is known as alternating optimization (AO), see [19] for additional details.
where \( \eta_{ij} = |\hat{h}_i s_j|^2 \geq 0 \). We observe that the minimization is over a compact set (a disk of radius \( P \)) and the objective function is continuous which implies a minimum exists. Since the eigenvalues \( \lambda_1 \) and \( \lambda_2 \) are squared in the objective function of problem (2.9), we restrict the eigenvalues to be non-negative and perform a change of variables \( \tilde{\lambda}_1 = \lambda_1^2 \), \( \tilde{\lambda}_2 = \lambda_2^2 \). We obtain the optimization problem,

\[
\min_{\lambda_1, \lambda_2 \geq 0} \left( \frac{N_0 + \tilde{\lambda}_2 \eta_{12}}{N_0 + \tilde{\lambda}_2 \eta_{22}} \right) \left( \frac{N_0 + \tilde{\lambda}_1 \eta_{21}}{N_0 + \tilde{\lambda}_1 \eta_{11}} \right).
\]

(2.10)

Taking derivatives and equating them to zero, it can be shown that there are no critical points in the interior of the feasible set. Thus, the solution of problem (2.10) must lay on the boundary of the feasible set, i.e. it must satisfy \( \tilde{\lambda}_1 = P - \tilde{\lambda}_2 \). Substituting this condition into the objective function we have the optimization problem,

\[
\min_{0 \leq \tilde{\lambda}_2 \leq P} \left( \frac{N_0 + \tilde{\lambda}_2 \eta_{12}}{N_0 + \tilde{\lambda}_2 \eta_{22}} \right) \left( \frac{N_0 + (P - \tilde{\lambda}_2) \eta_{21}}{N_0 + (P - \tilde{\lambda}_2) \eta_{11}} \right).
\]

Equating the derivative to zero, we obtain two critical points,

\[
\tilde{\lambda}_2^* = \frac{A \pm \sqrt{B}}{C}
\]

(2.11)

where,

\[
A = (\eta_{21} - \eta_{11})(\eta_{12} - \eta_{22})[(\eta_{11} + \eta_{12})N_0 + \eta_{11} \eta_{12} P][(\eta_{21} + \eta_{22})N_0 + \eta_{22} \eta_{21} P]
\]

\[
B = \eta_{12} \eta_{21} (N_0 + \eta_{11} P) - \eta_{11} \eta_{22} (N_0 + \eta_{21} P)
\]

\[
C = \eta_{12} \eta_{21} \eta_{22} + \eta_{11} (\eta_{12} (\eta_{21} + \eta_{22}) - \eta_{21} \eta_{22})
\]

Also, the global minimum \( \tilde{\lambda}_2 \) could be achieved on the boundary of the set \([0, P]\). Thus, in order to optimize the power allocation, we need to find the smallest feasible value among four possibilities: the two solutions of equation (2.11), \( \tilde{\lambda}_2 = 0 \) and \( \tilde{\lambda}_2 = P \).
To solve for the steering vectors (problem (2.7)), we apply results found in [20] on the global optimization of fractional quadratic objective functions with quadratic constraints (see also [21]). The steering vectors $s_1$ and $s_2$ minimize the objective function,

$$\log \left( \frac{N_0 + \lambda_2^2 \eta_{12}}{N_0 + \lambda_2^2 \eta_{22}} \right) + \log \left( \frac{N_0 + \lambda_1^2 \eta_{21}}{N_0 + \lambda_1^2 \eta_{11}} \right),$$

where $\eta_{ij} = |\tilde{h}_i s_j|^2$. Also, the optimization is subject to the constraint that $s_1$ and $s_2$ are unit-norm vectors. It is clear that we may optimize $s_1$ and $s_2$ separately due to the structure of the constraint set and the objective function. The minimization problem is,

$$\min_{|s_j|^2 = 1} \left( \frac{N_0}{\lambda_j} + s_j^* A_i s_j \right) \frac{N_0}{\lambda_j} + s_j^* A_j s_j$$

(2.12)

where $A_k = \tilde{h}_k^* \tilde{h}_k$. We first require upper and lower bounds on the solution to this problem,

$$B_l < \min_{|s_j|^2 = 1} \left( \frac{N_0}{\lambda_j} + s_j^* A_i s_j \right) \frac{N_0}{\lambda_j} + s_j^* A_j s_j < B_u.$$  

(2.13)

These bounds are easy to find, for example, the minimization problem is lower bounded by zero since the objective function is always positive while the upper bound can be found by evaluating the objective function for any feasible $s_j$. The key observation that enables one to solve this problem is that the following two statements are equivalent,

$$\min_{|s_j|^2 = 1} \left( \frac{N_0}{\lambda_j} + s_j^* A_i s_j \right) \frac{N_0}{\lambda_j} + s_j^* A_j s_j < \alpha_k$$

and

$$\min_{|s_j|^2 = 1} s_j^* (A_i - \alpha_k A_j) s_j + \frac{N_0}{\lambda_j^2} (1 - \alpha_k) \leq 0.$$

As shown in [20], one can find the global minimum of problem (2.12) by using a bisection
algorithm where at each step the following optimization subproblem is solved,

$$\min_{|s_j|^2 = 1} s_j^* (A_i - \alpha_k A_j) s_j + \frac{N_0}{\lambda_j^2} (1 - \alpha_k),$$

(2.14)

for a sequence of $\alpha_k$ that will be described next. The bisection procedure is as follows, suppose that at time $k$ one has an upper bound and lower bound, $B_{u,k}$ and $B_{l,k}$, respectively. Set $\alpha_k = \frac{B_{u,k} + B_{l,k}}{2}$ and solve problem (2.14), if the minimum value of the problem is negative, then $\alpha_k$ is a lower bound and set $B_{l,k+1} = \alpha_k$ and $B_{u,k+1} = B_{u,k}$. Otherwise, $\alpha_k$ is an upper bound and set $B_{u,k+1} = \alpha_k$ and $B_{l,k+1} = B_{l,k}$. The bounds in equation (2.13) are used to initialize the bisection. Intuitively, this algorithm operates by generating a sequence of decreasing upper bounds $\alpha_k$ on the minimum value of the problem (2.12) and solutions that will achieve these upper bounds. It can be shown that this algorithm will converge to the global optimum [20]. In order to complete the algorithm, a closed form solution of the subproblem (2.14) will now be given. Note that the eigenvalues of the symmetric matrix $B_{j,\alpha_k} = A_i - \alpha_k A_j$ are,

$$\mu_{j,\alpha_k} = \frac{1}{2} \left( \text{tr} (B_{j,\alpha_k}) + \sqrt{\text{tr} (B_{j,\alpha_k})^2 - 4 \det (B_{j,\alpha_k})} \right)$$

(2.15)

$$\nu_{j,\alpha_k} = \frac{1}{2} \left( \text{tr} (B_{j,\alpha_k}) - \sqrt{\text{tr} (B_{j,\alpha_k})^2 - 4 \det (B_{j,\alpha_k})} \right).$$

(2.16)

Since $B_{j,\alpha_k}$ is an Hermitian matrix, one may use the spectral theorem, which states that the factorization $B_{j,\alpha_k} = Q_{j,\alpha_k}^* Y_{j,\alpha_k} Q_{j,\alpha_k}$ exists and $Y_{j,\alpha_k}$ is a diagonal matrix whose diagonal entries consists of positive eigenvalues of $B_{j,\alpha_k}$ while $Q_{j,\alpha_k}$ is a unitary matrix which consists of orthonormal eigenvectors of $B_{j,\alpha_k}$. Therefore,

$$s_j^* B_{j,\alpha_k} s_j = \mu_{j,\alpha_k} |\hat{s}_{j,1}|^2 + \nu_{j,\alpha_k} |\hat{s}_{j,2}|^2,$$

where $\hat{s}_j = Q_{j,\alpha_k} s_j$ and since $Q_{j,\alpha_k}$ is unitary, $|\hat{s}_j|^2 = 1$. These considerations allow one
to write an optimization problem equivalent to problem (2.14),

$$\min_{|s_j|^2=1} \mu_{j,\alpha_k} \left| \tilde{s}_{j,1} \right|^2 + \nu_{j,\alpha_k} \left| \tilde{s}_{j,2} \right|^2.$$ 

Since $\tilde{s}_{j,1}$ and $\tilde{s}_{j,2}$ are squared in the objective function above, we perform another change of variables $u_{j,1} = |\tilde{s}_{j,1}|^2$ and $u_{j,2} = |\tilde{s}_{j,2}|^2$, and obtain the problem,

$$\min_{u_{j,1}+u_{j,2}=1, u_{j,1} \geq 0, u_{j,2} \geq 0} \mu_{j,\alpha_k} u_{j,1} + \nu_{j,\alpha_k} u_{j,2},$$

or equivalently,

$$\min_{0 \leq u_{j,1} \leq 1} (\mu_{j,\alpha_k} - \nu_{j,\alpha_k}) u_{j,1}.$$ 

This problem can be solved by noting that the maximizing solution is $u_{j,1} = 0$ if $\mu_{j,\alpha_k} > \nu_{j,\alpha_k}$ and $u_{j,1} = 1$ otherwise. Thus, we conclude that if $\mu_{j,\alpha_k} > \nu_{j,\alpha_k}$ the minimizing vector $d_{j,k}^*$ of problem (2.14) is $Q_{j,\alpha_k}^* e_1$ where $e_1 = [1,0]^T$, otherwise the minimizing vector is $Q_{j,\alpha_k}^* e_2$ where $e_2 = [0,1]^T$. The minimum value of the subproblem is $\min (\mu_{i,\alpha_k}, \nu_{i,\alpha_k}) + \frac{N_0}{N_f} (1 - \alpha_k)$.

**Algorithm**

An efficient algorithm for solving the cross rate problem (2.6) can now be formulated by combining the optimization of the steering matrix and power allocation in an alternating fashion, see Algorithm 1. The alternation between the steering and power allocation technique is guaranteed to converge to a local maximum of the cross rate [19].

**Simulations**

Here, we compare the regularized zero forcing and cross rate precoders. This is done by considering the average fraction of the linear maximal sum rate $R_L$, achieved by each precoder. The linear maximal sum rate is defined as the sum rate that can be achieved
Optimizing Steering Matrix

for \( i = 1 \) and \( i = 2 \) do

Initial Step: Set \( B_{l,0} \leftarrow B_l \) and \( B_{u,0} \leftarrow B_u \)

for \( k \leq 1 \) until \( B_{a,k} - B_{l,k} \leq \epsilon \) do

\[
\alpha_k \leftarrow \frac{B_{u,k-1} + B_{l,k+1}}{2}.
\]

Compute \( \mu_{i,\alpha_k} \) and \( \nu_{i,\alpha_k} \) using eqs. (2.15)-(2.16).

\[
\beta_k \leftarrow \min (\mu_{i,\alpha_k}, \nu_{i,\alpha_k}) + \frac{N_0}{\lambda_j} (1 - \alpha_k).
\]

if \( \beta_k \leq 0 \) then

\[
B_{l,k} \leftarrow B_{l,k-1} \text{ and } B_{u,k} \leftarrow \alpha_k
\]

end

if \( \beta_k > 0 \) then

\[
B_{l,k} \leftarrow \alpha_k \text{ and } B_{u,k} \leftarrow B_{u,k-1}
\]

end

\( d^k_j \leftarrow \) minimizing vector of \( k \)th subproblem (2.14).

end

\( S^2 \leftarrow [d^1_1 d^1_2] \)

Optimizing Power Allocation

Compute \( \lambda_+ \) and \( \lambda_- \) using eq. (2.11).

\[
\lambda_1^2 \leftarrow \max \left( \gamma (H, S^2, \lambda_+), \gamma (H, S^2, \lambda_-), \gamma (H, S^2, P), \gamma (H, S^2, 0) \right)
\]

\[
\lambda_2^2 \leftarrow P - (\lambda_1^2)^2
\]

Algorithm 1: Cross Rate Optimization
Chapter 2. Broadcast Channels and Transmitter Side Processing

by the best linear precoder. We plot the two quantities,

\[ W_{RZF} = \mathbb{E}\left\{ 100 \times \frac{R(H, D^{RZF})}{R_L(H)} \right\} \]

and

\[ W_{CR} = \mathbb{E}\left\{ 100 \times \frac{R(H, D^{CR})}{R_L(H)} \right\} \]

where the expectation is over the channel \( H \), which is drawn from a normal complex circularly symmetric distribution. In practice, 10,000 channel realizations were used to evaluate the expectations. The best linear precoder is computed using MATLAB’s built-in interior point optimization routine. The optimization algorithm finds the precoder that maximizes the sum rate subject to the power constraint. The cross-rate precoder is found using Algorithm 1. Figure 2.1 plots \( W_{RZF} \) and \( W_{CR} \) as the SNR is varied from 0 to 20 dB. It is clear that the cross rate precoder gains an additional 10% of the maximal rate achievable by linear precoders. The cross rate optimization was performed by a two step iteration of algorithm 1. Figure 2.2 plots \( W_{RZF} \) and \( W_{CR} \) where the cross rate is optimized by a single iteration of algorithm 1. This is done by assuming equal power allocation to both users, finding the optimal steering matrix given this power allocation and then optimizing the power allocation. We observe that at low and high SNR, this gives one the same performance as the regularized zero forcing precoder. However, at certain SNR values, the performance gain achieved is up to 12% of the maximal rate achieved by linear precoding.

2.3 Random Channel Side Information

2.3.1 Random CSI and Fading

An important physical phenomenon that one often encounters is that of fading. Fading is caused by changing channel conditions (for example a mobile wireless channel) and is
Figure 2.1: % of Maximal linear rate achieved by both regularized zero forcing and cross rate precoders

Figure 2.2: % of Maximal linear rate achieved by both regularized zero forcing and single iteration optimization of the cross rate
characterized by a fading probability distribution, $H \sim f(H)$. One usually distinguishes between fast and slow fading; fast fading is such that the channel changes quickly over a period of use while in slow fading the channel changes slowly and can be regarded as constant over a period of use. Either way the system will generate a sequence of channels $H_1, H_2, \ldots$ that is distributed according to $f(H)$. We note that despite the fading, the CSI may still be completely accurate. In contrast, random CSI is the scenario where the transmitter does not know the precise value of the channel (sequence of channel realizations when the channel is fading) and this imperfection is represented by a probability distribution. The notion of random CSI can be formalized as follows. The transmitter is assumed to have some information on the channel, $U \in Y$ where $Y$ is the set of possible CSI elements that the transmitter has access to. The CSI is dependent on the channel via a map, $\vartheta : \mathbb{C}^{p \times n} \to Y$, whose domain is the set of all possible channels, $\mathbb{C}^{p \times n}$ and the range is the set of CSI $Y$. For example, by letting $Y = \mathbb{C}^{p \times n}$ and $\vartheta$ be the identity map we obtain the scenario of the perfectly known channel considered before. The CSI map $\vartheta$ may be stochastic or deterministic, depending on the specific CSI mechanism. The structure of the information on the channel $H$ that is available by the CSI $U$ can be represented by a probability distribution $f(H; U)$. This probability distribution is parameterized by the available CSI $U$ and is strongly dependent on the CSI map, $\vartheta$. Finding $f(H; U)$ is not always straightforward, for example when the CSI is generated in a deterministic manner such as quantization. We will give some examples how one could construct such probability distributions in Chapter 5. Finally, in this work we assume the CSI is generated from a single channel realization rather than a sequence of channel realizations. This assumption excludes some important situations (notably that of CSI delay) but is still general enough to encompass two important scenarios: channel estimation error and quantization.
Gaussian CSI

In this situation, each receiver feeds back pilot symbols to the transmitter over an AWGN channel. Assuming that the transmitter uses a linear estimator (e.g., MMSE estimator) to construct a channel estimate, the CSI will be corrupted by Gaussian noise. Mathematically, we set \( Y = C^{p \times n} \) and let the CSI map \( \vartheta : C^{p \times n} \to C^{p \times n} \) be the stochastic function,

\[
\vartheta (H) = H + n,
\]

where \( n \) is complex Gaussian noise. The mean and covariance of the additive noise \( n \) will depend on both the pilot symbols, the channel estimator and the covariance structure of the reverse channels (see Section 5.1.2 for a detailed example).

Quantized CSI

Here, it is assumed that each receiver has access to its own channel which it quantizes and feeds back to the transmitter. The quantization operates as follows: each receiver partitions the space \( C^n \) into \( 2^m \) disjoint labeled partitions, \( Q_1, \ldots, Q_{2^m} \). Note that since \( C^n \) is a noncompact set, at least one of these partitions must be unbounded. Each receiver then feeds back to the transmitter the label of the partition of which its channel is a member. Thus the transmitter receives \( m \) bits from each receiver for a total of \( pm \) bits to represent the full channel matrix. We can easily include this process in the random CSI framework. Denote \( [m] = (1, 2, \ldots, m) \), then the set of CSI is the \( p \)-fold cartesian product of \( [m] \): \( Y = [m] \times [m] \ldots \times [m] = [m]^p \). The CSI map, \( \vartheta : C^{p \times n} \to [m]^p \) is defined by the aggregation of the channel quantization at each receiver as described. This CSI map is deterministic, and in Chapter 5 we use the maximum entropy principle to derive a suitable probability distribution.
2.3.2 Robust Precoders

Several different precoder designs have been published that account for imperfect CSI. These precoders vary in several ways: precoder structure (THP or linear), performance metric (sum rate, weighted sum rate, MSE, SINR constraints) and CSI error model (Gaussian noise, quantization, bounded perturbations). Thus, the design of a robust precoder can be conceptually reduced to a choice of precoder structure, performance metric and CSI imperfection model. The challenge in designing a robust precoder is to find an efficient algorithm that incorporates all these elements in a consistent manner. In [13] and [11], THP precoder designs are given that account for channel estimation error due to a finite amount of training by minimizing the mean square error (MSE) induced by the imperfect CSI. Also, in [13] the authors give a different THP design that accounts for channel quantization.

A popular approach to finding linear precoders under a condition of imperfect CSI is to first assume the channel lies in some bounded region that is known to the transmitter. This turns out to be a natural model when the source of imperfection is quantization. The strategy is then to locate the worst-case channel and find a linear precoder that will maximize some performance measure (e.g. sum rate) under the assumption that the worst-case channel has occurred. Thus, by considering the worst-case scenario, this approach is able to guarantee some minimal system performance. Note that this is a conservative design methodology since it assumes the worst-case channel while in reality, the channel may be more favorable. Such a robust linear precoder was designed in [10] by optimizing a worst case weighted sum-rate, under an assumption of imperfect CSI whose perturbation is confined to a bounded region. In [12], a robust linear precoder was introduced where the imperfections in the CSI are again assumed to be deterministically bounded. The design problem is formulated as a semidefinite programming problem (SDP).
### 2.3.3 Bounds on Interference Power

The following section was published by the author in [24]. It is concerned with bounding the performance penalty due to random CSI. It is clear that an exact analysis of the performance penalty due to imperfect feedback would require analyzing the mismatched precoder and channel, which is generally difficult but can be done in some special cases [23]. We propose to examine noisy feedback linear precoders from an estimation theoretic point of view. This is done by deriving a Cramér-Rao type bound for precoders. Recall that the classical Cramér-Rao bound, a key result in estimation theory, gives a lower bound on the MSE achievable by any estimator of a parameter, given a model relating the parameter and measurement. Similarly, by assuming some regularity conditions, a bound is derived which gives a lower bound on the sum mean squared error (SMSE)\(^2\) achievable at the receiver.

The bound is derived for a general model: the feedback is assumed to be stochastically dependent on the channel realization while the channel itself is fading according to some given probability distribution. This relation can be fully described by assuming a joint distribution of feedback and channel. Next, we consider power constrained precoders and show that a certain regularity condition can be significantly weakened. Combining the two results, we obtain a bound on precoding MSE whose validity depends on certain differentiability and continuity properties of the joint distribution of feedback and channel fading only.

#### Linear Precoders and Noisy CSI

It is clear that a precoder is a map that takes the symbols and feedback matrix and outputs a precoded data vector. Such a point of view is quite general and as such, encompasses any conceivable precoding scheme under any power control policy.

\(^2\)The SMSE is defined as the trace of the error covariance matrix of the effective channel, formed by the precoder and channel in series.
Definition. A precoder is linear when the symbol vector $s$ is mapped to the transmitted vector by a multiplication by a precoder matrix $\Phi(U)$. The precoding matrix is dependent on the feedback matrix $U$ and can be also viewed as a function $\Phi : \mathbb{R}^{p \times n} \rightarrow \mathbb{R}^{n \times p}$ which maps the feedback matrix $U$ into a precoding matrix. Also, let $\varphi_{ij} : \mathbb{R}^{p \times n} \rightarrow \mathbb{R}$ denote the scalar valued function in the $i,j$ entry of the precoder $\Phi$.

A simple precoder that can be expressed in terms of the feedback matrix $U$ is a standard zero forcing precoder,

$$\Phi_{ZF}(U) = \frac{1}{\gamma}U^T(UU^T)^{-1},$$

(2.17)

where $\gamma$ is a normalizing factor. Nonetheless, a precoder is rarely an elementary function of the feedback. High performance precoders may involve the optimization of an objective function such that no closed form solution exists [25].

The mean squared error (MSE) is an important and practical figure of merit in pre-coded broadcast systems. Noting that receivers in a broadcast system cannot cooperate and only the variance of the error at each receiver is meaningful, one may define the mean square error as the trace of the error covariance matrix,

$$\mathbb{E}\left\{ \text{tr}\left( H\Phi(U)s + n_e - s \right)\left( H\Phi(U)s + n_e - s \right)^T \right\}.$$ 

The expression can be simplified by extracting the term due to the AWGN. Using the independence of $n_e$ and $s$, and the fact that $n_e$ is zero mean, one may write the MSE as,

$$\mathbb{E}\left\{ \text{tr}\left( H\Phi(U)s - s \right)\left( H\Phi(U)s - s \right)^T \right\} + p\sigma_e^2.$$ 

Henceforth, we will ignore the term due to the AWGN as it produces a constant increase in the MSE, independent of the precoder, while we are interested in the precoding’s impact on the MSE at the receiver. Next, using the assumption $\mathbb{E}\{ss^T\} = I$, we arrive
at the expression that will be of most interest to us,

\[ S(\Phi) = \mathbb{E}\left\{ \operatorname{tr}\left( H\Phi(U) - I \right)\left( H\Phi(U) - I \right)^T \right\}. \tag{2.18} \]

One may consider \( S \) as a measure of the mean square error between the compound channel \( H\Phi \) and the identity matrix (that would result from a perfect inversion of the channel) for a fixed precoder \( \Phi \).

**A Cramér-Rao Bound for Precoders**

The well known Cramér-Rao bound [27] gives a lower bound on the variance of an arbitrary estimator of an unknown deterministic parameter. Similarly, a Bayesian CRLB [28] gives a lower bound on error variance when estimating a random variable. In the following theorem, we obtain a similar lower bound on the MSE (2.18) for an arbitrary precoder.

**Theorem 1.** For any linear precoder \( \Phi \), the following bound holds:

\[ S(\Phi) \geq \sum_{i=1}^{P} \sum_{j=1}^{N} \mathbb{E}\left\{ \Phi_{ii} \right\}^2 \frac{\mathbb{E}\left\{ \left( \frac{\partial \ln(f(H,U))}{\partial H_{ji}} \right)^2 \right\}}{\frac{\partial^2 \ln(f(H,U))}{\partial H_{ji}^2}} \]

\[ = \sum_{i=1}^{P} \sum_{j=1}^{N} \mathbb{E}\left\{ \Phi_{ii} \right\}^2 \mathbb{E}\left\{ \frac{\partial \ln(f(H,U))}{\partial H_{ji}} \right\}^2 \tag{2.19} \]

where \( H_{i,j} \) denotes the \((i,j)\) element of \( H \). Let \( f(H,U) \) be the joint probability density of \( U \) and \( H \), \( f(U|H) \) is the probability density of \( U \) conditional on \( H \) and \( f(H) \) is the marginal probability density of \( H \). The following conditions are assumed:

1. \( \frac{\partial f(H,U)}{\partial H_{i,j}} \) and \( \frac{\partial^2 f(H,U)}{\partial H_{i,j}^2} \) are absolutely integrable with respect to \( U \) and \( H \) for any element \( H_{i,j} \).

2. Define the quantity,

\[ B(H) = \int H\Phi(U) f(U|H) \, dU \tag{2.20} \]
then it is assumed that

\[ \lim_{H \to \infty} f(H) B(H) = \lim_{H \to -\infty} f(H) B(H) = 0. \] (2.21)

where a limit of a matrix means a corresponding limit in each one of its entries.

Proof. See Appendix.

Thus far we considered linear precoders in some generality as Theorem 1 was derived for any fading and feedback noise distributions under certain regularity conditions. In the following we show that condition 2 of Theorem 1 can be considerably weakened for a power constrained precoder as defined in equation (2.4).

**Lemma 1.** Given a linear precoder with a power constraint, the following is true

\[ \lim_{H \to \infty} f(H) B(H) = \lim_{H \to -\infty} f(H) B(H) = 0 \]

so long as \( \lim_{H \to \infty} f(H) H = \lim_{H \to -\infty} f(H) H = 0 \) where a limit of a matrix means a corresponding limit in each one of its entries.

Proof. Taking the \( i, j \) entry of \( B(H) \), we have

\[ f(H) B(H)_{i,j} = f(H) \sum_{r=1}^{n} H_{i,r} \int \varphi_{r,j}(U) f(U|H) \, dU. \]

The power constraint (2.4) implies that \( \varphi_{i,j} \) for any \( i, j \) is bounded,

\[ |\varphi_{i,j}| \leq \sqrt{P}. \]

Thus \( f \varphi_{r,j}(U) f(U|H) \, dU \) is finite. This further implies that

\[ \lim_{H \to \infty} f(H) B(H)_{i,j} \]
Chapter 2. Broadcast Channels and Transmitter Side Processing

\[
= \sum_{r=1}^{n} \lim_{H \to \infty} f(H) H_{i,r} \int \varphi_{r,j}(U) f(U|H) \, dU
\]
\[
= 0,
\]
so long as \( \lim_{H \to \infty} f(H) H_{i,r} = \lim_{H \to -\infty} f(H) H_{i,r} = 0 \), which gives us the theorem above.

This result translates a condition on a power constrained precoder to a condition on the fading distribution only. We can use this result to weaken the conditions of Theorem 1 when it is applied to power constrained precoders. These conditions do not depend on the precoder but only on joint and marginal distributions of the feedback and channel.

**Corollary 2.** For a precoder \( \Phi \) that is linear and power constrained, bound (2.19) holds so long as the following two conditions are satisfied:

1. \( \frac{\partial f(H,U)}{\partial H} \) and \( \frac{\partial^2 f(H,U)}{\partial H^2} \) are absolutely integrable with respect to \( U \) and \( H \).
2. \( \lim_{H \to \infty} f(H) H = \lim_{H \to -\infty} f(H) H = 0 \).

**Applications of the Cramér-Rao Bound**

In this section we give some practical examples of the precoder bound of Theorem 1. In order to simplify the formulae we will assume a fair precoder. A fair precoder entails the condition that \( \mathbb{E}\{\varphi_{ii}\} = \mathbb{E}\{\varphi_{jj}\} \), i.e. the diagonal entries in the precoder matrix are identical on average. Intuitively, this condition states that, on average, each symbol has the same power as it is transmitted to its intended receiver. Moreover, we make the additional assumption that the precoder is made to satisfy the power constraint by a multiplicative constant only. More precisely, the precoder is a linear function of the square root of the power. Denote the expectation of any diagonal element of a fair precoder as \( \tilde{\varphi} \), where the precoder satisfies a unity power constraint \( \text{tr}(\Phi^T \Phi) \leq 1 \). Note that using these assumptions, we can write \( \mathbb{E}\{\varphi_{ii}\} = \sqrt{P}\tilde{\varphi} \). Also, we assume that the transmitter power is uniformly allocated to each data stream.
We discuss a scenario where the channel matrix is fading according to a standard Gaussian distribution and the feedback is corrupted by Gaussian noise. Next, we bound the precoding MSE when the channel is fading according to Nakagami-m and Weibull distributions.

**Gaussian Feedback and Fading**

Assume that the entries of channel matrix $H$ are i.i.d. with distribution $f(H_{ij}) \sim \mathcal{N}(0, \sigma_c^2)$. Next, the transmitter has access to the matrix $U \in \mathbb{R}^{p \times n}$ which has the following relationship to the channel realization $H$,

$$U = H + N_f$$

where $N_f$ is a random $p \times n$ matrix with i.i.d. entries, each distributed as $\mathcal{N}(0, \sigma_f^2)$. Under these distributional assumptions, the joint density of the channel $H$ and its estimate $U$ is also Gaussian [29]. Indeed, it is not hard to see that the joint density has the covariance matrix,

$$
\begin{pmatrix}
\sigma_c^2 I \otimes I & \sigma_c I \otimes I \\
\sigma_c I \otimes I & (\sigma_c^2 + \sigma_f^2) I \otimes I
\end{pmatrix}
$$

with correlation coefficient $\rho = \left(\frac{\sigma_c^2}{\sigma_f^2} + 1\right)^{-1/2}$ between any pair of entries of the matrices $U$ and $H$. It will prove important later on to note that the joint distribution is Gaussian for any choice of $\sigma_c^2$ and $\sigma_f^2$. Substituting into Theorem 1, we calculate the log-derivative,

$$-E\left\{ \frac{\partial^2 \ln (f(H, U))}{\partial H_{ij}^2} \right\} = \frac{1}{\sigma_c^2} + \frac{1}{\sigma_f^2}.$$  

Substituting into (2.19) and normalizing by the number of users, we obtain the bound

$$\frac{1}{p} S(\Phi) \geq nP \left( \frac{1}{\sigma_c^2} + \frac{1}{\sigma_f^2} \right)^{-1} \sigma^2.$$  

(2.23)
This bound gives some important insights into the system behavior in the high power regime. First, note that as the power is increased, the user-normalized MSE will increase linearly with power. However, for small $\sigma_f^2$, the bound trivially states that the MSE will be non-negative. Thus, a necessary condition to prevent the MSE from becoming large for large power is that we must scale the feedback variance $\sigma_f^2$ linearly with transmission power.

### Nakagami Fading

The Nakagami-$m$ distribution is a statistical distribution common when modeling fading envelopes of signals [30]. Let us now assume that the channel coefficients are i.i.d. distributed as Nakagami-$m$ random variables,

$$f(H_{ij}) = \frac{2^m m^m}{\Gamma(m) \Omega^m} H_{ij}^{2m-1} \exp\left( -\frac{m H_{ij}^2}{\Omega} \right) \quad \text{for } H_{ij} \geq 0$$

where $m \geq 0.5$ is the shape parameter and $\Omega > 0$ controls the spread of the distribution. We assume an identical feedback model to the one given in equation (2.22), i.e. the conditional density of the feedback $U$ on channel realization $H$ is $N(H, \sigma_f^2 I)$. The expected second derivative of the log-joint density can be computed to be,

$$\frac{1}{\sigma_f^2} + \frac{2m}{\omega} - \frac{1 - 2m}{H_{ij}^2}.$$  \hspace{1cm} (2.24)

Taking the expected value of this expression, we obtain the bound

$$\frac{1}{p} S(\Phi) \geq n P \left( \frac{1}{\sigma_f^2} + \frac{3 - 4m}{\omega (m - 1)} \right)^{-1} \tilde{\varphi}^2.$$

This bound is only valid for $m > 1$ as the second inverse moment of the Nakagami-$m$ distribution does not exist otherwise.
Weibull Fading

We also mention the Weibull distribution since recent theoretical and experimental work [31] shows that the Weibull distribution accurately models some wireless channels. Assuming that the channel coefficients are i.i.d. distributed as a Weibull random variable, 

$$f(H_{ij}) = \frac{k}{\lambda} \left( \frac{H_{ij}}{\lambda} \right)^{k-1} \exp \left( - \left( \frac{H_{ij}}{\lambda} \right)^k \right) \text{ for } H_{ij} \geq 0$$

where $k > 0$ is the fading parameter and $\lambda > 0$ is the average fading power. Also, the feedback noise is identical to previous scenarios, $U | H \sim N(H, \sigma_f^2 I)$. The second derivative of the log-joint density is,

$$\frac{1}{\sigma_f^2} + 1 - k - \frac{(k - k^2)(H_{ij}^2/\lambda)^k}{H_{ij}^2}.$$

The expected value of the expression above only exists for $k > 2$, yielding the bound

$$\frac{1}{p} S(\Phi) \geq n P \left( \frac{1}{\sigma_f^2} + \frac{(k - 1)^2 \Gamma \left( \frac{k-2}{k} \right)}{\lambda^2} \right)^{-1} \varphi^2. \quad (2.25)$$

Combining the three bounds (2.23), (2.24) and (2.25), we obtain a strong conclusion that holds for three well known fading models and any linear precoder. Namely, in order to prevent the MSE from becoming large for large power, one must scale the feedback variance $\sigma_f^2$ inversely with power. In Chapter 4, a similar statement is shown to also hold for the ergodic sum-rate and outage probability.

### 2.4 Summary

In this chapter, we reviewed some of the basic notions of broadcast channels. The capacity of the MISO broadcast channel was stated and DPC was discussed. The various types of practical precoders were surveyed and the important class of linear precoders was
introduced. Also, we presented a new design metric, the cross rate, and showed how it arises by considering the difference of sum rates of a system and its transpose. The cross rate was used to efficiently design a linear precoder. Simulation results indicate the new precoder design achieves a significant rate gain compared to standard linear precoder designs. The notions of incomplete and random CSI were rigorously defined and specific examples of Gaussian and quantized CSI were given. The literature on robust precoders was surveyed. Finally, we derived a Cramér-Rao type bound on the performance of any linear precoder for an arbitrarily fading broadcast system. The feedback provided to the precoder is taken to be stochastic and dependent on the channel realization in an arbitrary way.

2.5 Appendix: Cramér-Rao Bound, Theorem 1

The derivation of Theorem 1 is a modification of a calculation found in [32] where a bound is derived on the MSE when estimating a random variable. We will denote vectorization of an arbitrary matrix $A$ by $\bar{A}$ where the vectorization is done by stacking the column vectors of $A$. Proceeding with the derivation of Theorem 1, one can simplify the sum MSE of the system as

$$\mathbb{E}\{\text{tr} (H\Phi s - s)(H\Phi s - s)^T\} = \mathbb{E}\{\text{tr} (H\Phi - I)(H\Phi - I)^T\}$$

by taking a conditional expectation and using the assumption that $\mathbb{E}\{ss^T\} = I$. Now, by a property of the Kronecker product, we can write $\text{vec}(H\Phi) = (\Phi^T \otimes I) \text{vec}(H) = (\Phi^T \otimes I) \bar{H}$. Note that the conditional expectation of the error (see equation (2.20)) can now be written in vector form as,

$$D(\bar{H}) = f(\bar{H}) \int \left( (\Phi^T \otimes I) \bar{H} - I \right) f_{U|H} dU$$

$$= f(\bar{H}) B(\bar{H}) - f(\bar{H}) I$$
Differentiate the \( i \)-th entry of \( \mathbf{D}(\tilde{\mathbf{H}}) \) with respect to the \( j \)-th entry of the vector \( \tilde{\mathbf{H}} \)

\[
\frac{\partial}{\partial \tilde{H}_j} [\mathbf{D}(\tilde{\mathbf{H}})]_i = \int (\mathbf{\Phi}^T \otimes \mathbf{I})_{ij} f(\tilde{\mathbf{H}}, \mathbf{U}) \ d\mathbf{U} + \int [(\mathbf{\Phi}^T \otimes \mathbf{I}) \tilde{\mathbf{H}} - \mathbf{I}]_i \frac{\partial f(\tilde{\mathbf{H}}, \mathbf{U})}{\partial \tilde{H}_j} \ d\mathbf{U}.
\]

Combining all derivatives in a matrix form,

\[
\frac{\partial}{\partial \tilde{H}} \mathbf{D}(\tilde{\mathbf{H}}) = \int (\mathbf{\Phi}^T \otimes \mathbf{I}) f(\tilde{\mathbf{H}}, \mathbf{U}) \ d\mathbf{U} + \int (\mathbf{\Phi}^T \otimes \mathbf{I}) (\tilde{\mathbf{H}} - \mathbf{I}) \frac{\partial f(\tilde{\mathbf{H}}, \mathbf{U})^T}{\partial \tilde{H}} \ d\mathbf{U}.
\]

and integrating with respect to \( \tilde{\mathbf{H}} \) and using the fundamental theorem of calculus together with assumption 2, we obtain the equation

\[
\mathbb{E}\{\mathbf{\Phi}^T \otimes \mathbf{I}\} + \int \int (\mathbf{\Phi}^T \otimes \mathbf{I}) (\tilde{\mathbf{H}} - \mathbf{I}) \frac{\partial f(\tilde{\mathbf{H}}, \mathbf{U})^T}{\partial \tilde{H}} \ d\mathbf{U} \ d\tilde{\mathbf{H}} = \mathbf{0}
\]

Premultiplying by an arbitrary \( 1 \times p^2 \) vector \( \mathbf{a}^T \) and postmultiplying by an arbitrary \( np \times 1 \) vector \( \mathbf{B} \),

\[
\int \mathbf{a}^T ((\mathbf{\Phi}^T \otimes \mathbf{I}) \tilde{\mathbf{H}} - \mathbf{I}) \frac{\partial \ln (f(\tilde{\mathbf{H}}, \mathbf{U}))^T}{\partial \tilde{H}} \mathbf{B} f(\tilde{\mathbf{H}}, \mathbf{U}) \ d\mathbf{U} \ d\tilde{\mathbf{H}} = -\mathbf{a}^T \mathbb{E}\{\mathbf{\Phi}^T \otimes \mathbf{I}\} \mathbf{B}
\]

Applying the Cauchy-Schwarz inequality, we have

\[
(\mathbf{a}^T \mathbb{E}\{\mathbf{\Phi}^T \otimes \mathbf{I}\} \mathbf{B})^2 \leq \int \mathbf{a}^T ((\mathbf{\Phi}^T \otimes \mathbf{I}) \tilde{\mathbf{H}} - \mathbf{I}) (\mathbf{\Phi}^T \otimes \mathbf{I}) (\tilde{\mathbf{H}} - \mathbf{I})^T \mathbf{a} f(\tilde{\mathbf{H}}, \mathbf{U})
\]

\[
\cdot \int \mathbf{B}^T \frac{\partial \ln (f(\tilde{\mathbf{H}}, \mathbf{U}))}{\partial \tilde{H}} \frac{\partial \ln (f(\tilde{\mathbf{H}}, \mathbf{U}))^T}{\partial \tilde{H}} \mathbf{B} f(\tilde{\mathbf{H}}, \mathbf{U}) \ d\mathbf{U}.
\]

where the integrals are with respect to both \( \mathbf{U} \) and \( \tilde{\mathbf{H}} \). Since \( \mathbf{a} \) and \( \mathbf{B} \) are arbitrary, let them both be vectors containing all zeros except at the \( k \)-th place where there is a one.
for \( k = 1, \ldots, p^2 \). Therefore, for any \( k \), we have the inequality

\[
E\left\{ \left( \left( \Phi^T \otimes I \right) \bar{H} - \bar{I} \right) \left( \left( \Phi^T \otimes I \right) \bar{H} - \bar{I} \right)^T \right\}_{kk} \geq \frac{E\left\{ \left( \left( \Phi^T \otimes I \right) \right)_{kk} \right\}^2}{E\left\{ \left( \frac{\partial \ln(f(H,U))}{\partial H} \right)^T \right\}_{kk}}.
\]

Now, since \( k \) is an integer between 1 and \( p^2 \), we can write it as \( k = p(i - 1) + j \) where \( i = 1, \ldots, p \) and \( j = 1, \ldots, p \). Moreover, it is not hard to see that \( \left( \Phi^T \otimes I \right)_{kk} = \Phi_{ii} \) and \( \left( \left( \Phi^T \otimes I \right) \bar{H} - \bar{I} \right)_{kk} = \left( \left( \Phi^T \otimes I \right) \bar{H} - \bar{I} \right)_{kj}^2 = (H \Phi - I)_{ji}^2 \). Thus, we have the inequalities for all \( i, j \)

\[
E\left\{ (H \Phi - I)_{ji}^2 \right\} \geq \frac{E\left\{ \Phi_{ii} \right\}^2}{E\left\{ \left( \frac{\partial \ln(f(H,U))}{\partial H_{ji}} \right)^2 \right\}}.
\]

Summing over all \( i, j \), we have a bound for the trace of the error covariance matrix,

\[
E\left\{ \text{tr} (H \Phi - I) (H \Phi - I)^T \right\} \geq \sum_{i,j} \frac{E\left\{ \Phi_{ii} \right\}^2}{E\left\{ \left( \frac{\partial \ln(f(H,U))}{\partial H_{ji}} \right)^2 \right\}}.
\]

The steps to show \( E\left\{ \left( \frac{\partial \ln(f(H,U))}{\partial H_{ji}} \right)^2 \right\} = \frac{E\left\{ \frac{\partial^2 \ln(f(H,U))}{\partial H^2_{ji}} \right\}}{E\left\{ \frac{\partial \ln(f(H,U))}{\partial H_{ji}} \right\}^2} \) and thus the final inequality of Theorem 1 are identical to those found in [32].
Chapter 3

Zero Forcing with Noisy Channel Side Information

As discussed previously, the capacity region of a broadcast MIMO channel with exact CSI is achieved by the precoding strategy of dirty paper coding [1]. This strategy hinges on the availability of accurate CSI at the transmitter. In this chapter, we are interested in understanding the impact of imperfect CSI on a system which uses a linear zero-forcing precoder (ZF). Linear ZF precoding is a transmitter side processing strategy that takes the provided channel matrix, computes its pseudo-inverse and multiplies this inverse by the vector of symbols. In essence, this operation pre-distorts the symbols in order to remove the interference caused by the channel so that only the intended symbols are visible to each receiver.

In the following, we consider a linear zero forcing precoder that is formed by a noisy channel estimate, where the estimation error is modeled as additive complex Gaussian noise. Also, the channel matrix will be modeled as random and flat fading with a complex Gaussian distribution, known in the literature as Rayleigh fading. Using this stochastic model, we analyze the overall mismatched system that is formed by the channel and the precoder. It will be proven that the effective channel matrix, formed by the channel
and the precoder, has a complex matrix variate student’s T-distribution. We will also show that the channel of any given user has a marginal distribution of a complex vector T-distribution, this distribution will be used to derive new results on the ergodic rate and outage probability in Chapter 4. Finally, we discuss an extension of these results to semi-correlated channels.

Preliminaries: Finite Random Matrices

In this chapter, we will make heavy use of finite random matrix theory and also some concepts from matrix integration theory (see [33] and [35]). A random matrix is defined to be a matrix valued random variable. The associated probability density function of a random matrix $A$ is the joint probability density of all of its entries. In order to integrate with respect to a matrix (e.g. to compute an expected value), we regard the integration as over $m$-dimensional complex space, where the dimension $m$ is equal to the product of matrix dimensions.

3.1 Transmitter Precoding with Noisy CSI

Consider a multi-input multi-output (MIMO) broadcast system with $n$ transmit antennas and $p$ single-antenna users such that $n > p$. The relationship between the inputs and outputs is modeled as a linear channel with additive white Gaussian noise (AWGN). Mathematically, we write

$$y = Hx + n_e$$

where $x$ is an $n$-dimensional vector to be transmitted, $H \in \mathbb{C}^{p \times n}$ is the channel matrix and $y$ is the $p$-dimensional vector of received information. The noise vector $n_e$ is distributed as a circularly symmetric $p$-dimensional complex Gaussian random vector, $n_e \sim \mathcal{N}_c(0, \sigma_e^2 I)$. We further assume a fading model where the entries of the channel matrix $H$ are independent and identically distributed (i.i.d.) as a complex circularly
symmetric Gaussian random variable with variance $\sigma_c^2$. We require that the average transmitted power is less than or equal to $P$ (i.e. $\mathbb{E}\{x^*x\} \leq P$). Next, we assume that the transmitter has access to the CSI matrix $U \in \mathbb{C}^p \times n$, which is the channel realization $H$, corrupted by additive Gaussian noise. This relationship is written,

$$U = H + N_f.$$  \hspace{1cm} (3.1)

The CSI noise matrix $N_f$ is a random $p \times n$ matrix whose entries are circularly symmetric i.i.d. with a distribution $\mathcal{N}_c(0, \sigma_f^2)$, independent of the channel $H$. A zero forcing precoder multiplies the intended symbols, denoted as $s$, by the pseudo-inverse of the CSI matrix $U^\dagger$. In the absence of CSI noise, this will amount to inverting the channel so that there is no interference at each receiver. The full system with the zero forcing precoder can be written,

$$y = \varrho HU^* (UU^*)^{-1} s + n_e$$

where $x = \varrho U^* (UU^*)^{-1} s$ is the transmitter output. The scalar $\varrho$ is associated with the power control policy. See Figure 3.1 for a block diagram of the system.
3.1.1 Power Control Policy

It is clear that the power of the transmitted vector $\mathbf{x}$ is also a random variable which depends on the channel and CSI realizations. The average power can be calculated as,

$$\mathbb{E}\{\mathbf{x}^*\mathbf{x}\} = \varrho^2 \mathbb{E}\{\mathbf{s}^* (\mathbf{U}\mathbf{U}^*)^{-1} \mathbf{s}\}$$

where the expectation is with respect to both $\mathbf{U}$ and $\mathbf{s}$. Observe that the matrix $\mathbf{U}$ is marginally distributed as $\mathcal{N}_c(\mathbf{0}, (\sigma_f^2 + \sigma_c^2) \mathbf{I})$. Therefore, the $p \times p$ Hermitian positive definite matrix $\mathbf{S}^{-1} = (\mathbf{U}\mathbf{U}^*)^{-1}$ is distributed as a complex inverse Wishart random matrix with $n$ degrees of freedom and an inverse scale matrix $\frac{1}{\sigma_f^2 + \sigma_c^2} \mathbf{I}$, or $\mathbf{S}^{-1} \sim \mathcal{W}^{-1}_c \left( n, \frac{1}{\sigma_f^2 + \sigma_c^2} \mathbf{I} \right)$. Using properties of the complex inverse Wishart distribution [36],

$$\mathbb{E}\{\mathbf{x}^*\mathbf{x}\} = \frac{\varrho^2}{(n - p) (\sigma_f^2 + \sigma_c^2)} \quad (3.2)$$

where we made the assumption that the input symbols $\mathbf{s}$ have equal power and their total power is unity ($\mathbf{s}^* \mathbf{s} = 1$). Therefore, one can satisfy the average power constraint by setting,

$$\varrho^2 \leq (\sigma_f^2 + \sigma_c^2) (n - p) \text{ } P. \quad (3.3)$$

Note that the expectation (3.2) does not exist for square matrices $n = p$, thus we restrict ourself to the case where $n > p$. We will give some remarks on the scenario where $n = p$ (i.e. a square channel matrix) later on.

3.1.2 The Effective Channel and its Distribution

In order to proceed with the analysis, we analyze the overall probabilistic behavior of the system in the following sense. The end-to-end system is written,

$$\mathbf{y} = \varrho \mathbf{H}\mathbf{U}^* (\mathbf{U}\mathbf{U}^*)^{-1} \mathbf{s} + \mathbf{n}_c. \quad (3.4)$$
Let $V = HU^*(UU^*)^{-1}$ denote the effective channel formed by joining the precoder matrix and the physical channel matrix (see also Figure 3.1). We will show that the distribution of the effective channel is a special case of the complex matrix variate T distribution, defined as follows.

**Definition.** A random $m \times k$ matrix, $A$, with a probability density function,

$$
\frac{\varphi^{-mk} \Gamma_k (\tau + k) \Gamma_{k-m} (\tau)}{\pi^{mk} \Gamma_k (\tau) \Gamma_{k-m} (\tau + k)} \det \left( I + \frac{1}{\varphi} (A - \mu I_{m \times k})^* (A - \mu I_{m \times k}) \right)^{-(\tau + m)},
$$

(3.5)

where $0 < \varphi < +\infty$, $\mu \in \mathbb{R}$ and $\tau$ is a positive integer, is denoted as $A \sim T_{m,k}(\tau, \mu, \varphi)$. When $m = 1$, the matrix $A$ is now a random vector and we write $A \sim T_k(\tau, \mu, \varphi)$. This distribution is a special case of the complex matrix variate T distribution.

More general definitions can be found in [35] for complex matrices and [33] for real matrices. Using this definition, we can state the distribution of the effective channel $V$, as follows

**Theorem 3.** Assume the $p \times n$ channel matrix $H \sim N_c(0, \sigma_c^2 I \otimes I)$ and the $(p \times n)$ CSI matrix $U$ has a conditional distribution $U|H \sim N_c(H, \sigma_f^2 I \otimes I)$. The effective channel $V = HU^*(UU^*)^{-1}$ is distributed as complex matrix variate T distribution, $V \sim T_{p,p}(n, \mu, \varphi)$ where the parameters $\mu$ and $\varphi$ are given as,

$$
\mu = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_f^2} \text{ and } \varphi = \frac{\sigma_f^2 \sigma_c^2}{(\sigma_c^2 + \sigma_f^2)^2}.
$$

**Proof.** See Appendix 3.3.1.

As a result, one can write the channel as,

$$
y = \rho V s + n_e
$$

where $V \sim T_{p,p}(n, \mu, \varphi)$. Moreover, it is straightforward to show that $V = \mu I_{p \times p} + \varphi^{1/2} G$.
where $\mathbf{G} \sim \mathcal{T}_{p,p} (n, 0, 1)$. Thus, the channel can be written as,

$$
\mathbf{y} = \varrho \mu \mathbf{s} + \varrho \varphi^{1/2} \mathbf{G s} + \mathbf{n}_e,
$$

where the term $\mu \mathbf{s}$ corresponds to the intended (deterministic) effect of the precoding when no CSI noise is present, but multiplied by the factor $\mu$ due to noisy CSI (note that $0 < \mu \leq 1$). The term $\varphi^{1/2} \mathbf{G s}$ is the stochastic component of the effective channel, induced by the imperfect inversion of the channel matrix, $\mathbf{H}$. Indeed, should the CSI be perfect (i.e. $\sigma_f^2 = 0$), then $\mu = 1$ and $\varphi = 0$, and this corresponds to perfect inversion of the channel matrix, resulting in $\mathbf{y} = \varrho \mathbf{s} + \mathbf{n}_e$.

On the receiver side, each user must decode its own data without any additional information from other users. The signal received by user $k$ can be written as,

$$
\mathbf{y}_k = \varrho \mu \mathbf{s}_k + \varrho \varphi^{1/2} \mathbf{g}_k \mathbf{s} + \mathbf{n}_e.
$$

(3.6)

where $\mathbf{g}_k$ is the $k$th row of the random matrix $\mathbf{G}$. Due to the independent decoding done by each user, the marginal distribution of the channel row, $\mathbf{g}_k$, associated with user $k$ is important in determining system performance in the next chapter.

**Theorem 4.** The vector $\mathbf{g}_k$ has a complex vector $T$ distribution, $\mathbf{g}_k \sim \mathcal{T}_p (n, 0, 1)$. The probability density function is,

$$
\frac{\Gamma (n + 1)}{\pi^p \Gamma (n - p + 1)} (1 + \mathbf{g}_k \mathbf{g}_k^*)^{-(n+1)}.
$$

**Proof.** See Appendix 3.3.2.

**Remark 1:** It is interesting to consider the case of square ($n = p$) channel matrices. As observed in [17], when a ZF precoder is used the system behavior is rather pathological as the mean of the transmitter power is infinite. Moreover, for the noisy CSI scenario, we observe that when $n = p$, the effective channel $\mathbf{V}$ is distributed as a complex
matrix variate Cauchy random matrix [35]. It is not hard to show that, similar to the scalar Cauchy distribution, this distribution does not have any finite moments. Therefore, the fading induces impulsive noise conditions\(^1\) on the overall channel. Also, since the expected power does not exist, we can observe that a zero forcing precoder, used in a square channel, cannot satisfy an expected power constraint. This phenomenon was noted in [17], where the regularized ZF precoder was introduced. Due to the functional form of the precoder, the distribution induced by the random CSI on the regularized zero forcing precoder is such that the expected power exists for square channels. Therefore, the regularized ZF precoder can satisfy an expected power constraint in a square channel. Unfortunately, the distributions and moments are difficult to find and are generally unknown and thus performance analysis of the kind given in this work is difficult.

### 3.1.3 Semicorrelated Channels

The distributional result of Theorem 3 can be generalized in a straightforward way to semi-correlated channels and channel feedback noise. By a semi-correlated random Gaussian matrix we refer to a random Gaussian matrix whose entries are correlated inside any given row but are independent between different rows. This is quite a realistic model since, heuristically, if the users are sufficiently far apart, one would expect their physical channels to be independent from each other.

Consider a channel distributed as a normal complex random matrix with covariance matrices, \( \mathbf{A} \) and \( \mathbf{B} \) and a zero mean, \( \mathbf{H} \sim \mathcal{N}_c (0, \mathbf{A} \otimes \mathbf{B}) \). The covariance between any two entries \( H_{ij}, H_{nm} \) can be computed [33],

\[
\text{Cov}(H_{ij}, H_{nm}) = \mathbf{A}_{in} \mathbf{B}_{jm}.
\]

---

\(^1\)Non-Gaussian noise of high or infinite variance is usually called impulsive in the literature as it may induce high interference levels, and is sometimes modeled as an alpha-stable distribution (non-Gaussian).
We observe that to have independence between different user channels, the covariance matrix \( \mathbf{B} \) should be diagonal. Moreover, without any loss of generality, we can set the covariance matrix \( \mathbf{B} \) to be the identity since if it is not, one can create a new pair of covariance matrices \( \tilde{\mathbf{A}}, \tilde{\mathbf{B}} \) so that \( \tilde{\mathbf{A}} = \mathbf{A} \mathbf{B} \) and \( \tilde{\mathbf{B}} = \mathbf{I} \). The Gaussian random matrix with covariance matrices \( \tilde{\mathbf{A}}, \tilde{\mathbf{B}} \) will have an identical covariance structure as the one with a non-identity diagonal left covariance matrix. Thus, we may assume a semi-correlated normal channel has a distribution \( \mathbf{H} \sim \mathcal{N}_c(\mathbf{0}, \Sigma_c \otimes \mathbf{I}) \). By these considerations, the following theorem on the effective channel can be proven.

**Theorem 5.** Assume the \( p \times n \) channel matrix \( \mathbf{H} \sim \mathcal{N}_c(\mathbf{0}, \Sigma_c \otimes \mathbf{I}) \) and the \( (p \times n) \) CSI matrix \( \mathbf{U} \) has a conditional distribution \( \mathbf{U}|\mathbf{H} \sim \mathcal{N}_c(\mathbf{H}, \Sigma_f \otimes \mathbf{I}) \). The effective channel \( \mathbf{V} = \mathbf{H} \mathbf{U}^*(\mathbf{U} \mathbf{U}^*)^{-1} \) is distributed as generalized complex matrix variate \( T \) distribution,

\[
\frac{\Gamma_p(n+p)}{\pi^p \Gamma_p(n)} \det(\Psi)^{-p} \det(\Omega)^{-p} \det(\mathbf{I} + \Psi^{-1}(\mathbf{V} - \Xi)^* \Omega^{-1}(\mathbf{V} - \Xi))^{n-p}.
\]

where,

\[
\Xi = \left( \Sigma_c^{-1} + \Sigma_f^{-1} \right)^{-1} \Sigma_f^{-1}
\]
\[
\Psi = \left( \Sigma_f^{-1} - \Sigma_f^{-1} \left( \Sigma_f^{-1} + \Sigma_c^{-1} \right)^{-1} \Sigma_c^{-1} \right)
\]
\[
\Omega = \left( \Sigma_c^{-1} + \Sigma_f^{-1} \right)^{-1}.
\]

**Proof.** See Appendix 3.3.1. \( \square \)

Theorem 5 indicates that the fundamental stochastic nature of the effective channel does not change due to correlations inside user channels, but as expected, correlations are formed between elements in the effective channel. Interestingly, while the users' channels are independent, their effective channels are no longer independent due to the zero forcing precoder.
3.2 Summary

We investigated a broadcast system with noisy CSI at the transmitter, modeled as the true channel matrix corrupted by additive Gaussian noise. An exact probabilistic analysis of a zero forcing precoder, mismatched to a complex valued random Gaussian channel was performed. It was shown that the overall channel has a complex matrix variate T-distribution with parameters depending on the system dimensions, fading variance and CSI noise variance. Moreover, we showed an identical conclusion holds for semi-correlated channels where the distribution of the resulting effective channel is a T-distributed random matrix with a more complicated covariance structure.

3.3 Appendices

3.3.1 Appendix: Theorem 3

The joint density of the complex channel $H$ and its complex estimate $U$ can be written,

$$
\pi^{-2np} \det(\Sigma_c)^{-n} \det(\Sigma_f)^{-n} \etr\left(-\Sigma_c^{-1}HH^* - \Sigma_f^{-1}(U - H)(U - H)^*\right).
$$

Note that,

$$
\etr\left\{-\Sigma_c^{-1}HH^* - \Sigma_f^{-1}(U - H)(U - H)^*\right\}
= \etr\left\{-\left(\Sigma_c^{-1} + \Sigma_f^{-1}\right)\left(H - \left(\Sigma_c^{-1} + \Sigma_f^{-1}\right)^{-1}\Sigma_f^{-1}U\right)\left(H - \left(\Sigma_c^{-1} + \Sigma_f^{-1}\right)^{-1}\Sigma_f^{-1}U\right)^* \right. \\
- \left. \left(\Sigma_f^{-1} - \Sigma_f^{-1}\left(\Sigma_f^{-1} + \Sigma_c^{-1}\right)^{-1}\Sigma_f^{-1}\right)UU^*\right\}.
$$

Therefore, we can rewrite the joint density as,

$$
\pi^{-2np} \det(\Psi)^n \det(\Omega)^{-n} \etr\left(-\Omega^{-1}(H - \Xi U)(H - \Xi U)^* - \PsiUU^*\right).
$$
Thus, we see that $H | U \sim \mathcal{N}_c(\Xi U, \Omega \otimes I)$ and $U \sim \mathcal{N}_c(0, \Psi^{-1} \otimes I)$ jointly. Using properties of the complex matrix variate normal distribution, this implies that $V | U \sim \mathcal{N}_c(\Xi, \Omega \otimes (UU^*)^{-1})$. Therefore, the joint density of $U$ and $V$ is,

$$\pi^{-np-p^2} \det (\Psi)^n \text{etr} (-\Psi UU^*) \det (UU^*)^p \det (\Omega)^{-p} \text{etr} (-\Omega^{-1} (V - \Xi) UU^* (V - \Xi)^*) .$$

To obtain the marginal density of $V$, we integrate over the matrix $U$,

$$\pi^{-p(n+p)} \frac{\det (\Psi)^n}{\det (\Omega)^p} \int_{U \in \mathbb{C}^{p \times n}} \det (UU^*)^p \text{etr} \left(- (\Psi + (V - \Xi)^* \Omega^{-1} (V - \Xi)) UU^* \right) dU.$$ 

Let $S = UU^*$ and $Q$ is a semi-unitary matrix of dimensions $p \times n$. Using a result from [35], we can perform a change of variables to arrive at the integral,

$$\pi^{-p(n+p)} \det (\Psi)^n \det (\Omega)^{-p} 2^{-p} \int_{S \in P(\mathbb{C}^{p \times p})} \int_{Q \in U(p,n)} \det (S)^n \text{etr} \left(- (\Psi + (V - \Xi)^* \Omega^{-1} (V - \Xi)) S \right) dS Q dQ.$$ 

Integrating over the semiunitary matrices, it is known that,

$$\int_{Q \in U(p,n)} Q dQ = \frac{2^{p} \pi^{np}}{\Gamma_p (n)}.$$ 

Continuing the derivation of the density of $V$, we have the expression,

$$\frac{\pi^{-p^2}}{\Gamma_p (n)} \det (\Psi)^n \det (\Omega)^{-p} \int_{S \in P(\mathbb{C}^{p \times p})} \det (S)^n \text{etr} \left(- (\Psi + (V - \Xi)^* \Omega^{-1} (V - \Xi)) S \right) dS.$$ 

We now integrate over $S$ using the gamma integral,

$$\int_{S \in P(\mathbb{C}^{p \times p})} \det (S)^{\alpha-p} \text{etr} (-BS) dS = \Gamma_p (\alpha) \det (B)^{-\alpha},$$
for $B \in P(C^n)$. Thus, we obtain the density of $V$ to be,

$$\frac{\Gamma_p(n+p)}{\pi^{p^2} \Gamma_p(n)} \det(\Psi)^{-p} \det(\Omega)^{-p} \det(I + \Psi^{-1}(V - \Xi)^* \Omega^{-1}(V - \Xi))^{-(n-p)}.$$ 

### 3.3.2 Appendix: Theorem 4

The quadratic form $(G - \mu I)^*(G - \mu I)$ in the density (3.5) can be expanded as,

$$(G - \mu I)^*(G - \mu I) = (G_1 - \mu I_{p_1 \times p})^*(G_1 - \mu I_{p_1 \times p}) + (G_2 - \mu I_{p_2 \times p})^*(G_2 - \mu I_{p_2 \times p}).$$

To obtain the density of $G_1$, we integrate over the matrix $G_2$,

$$\frac{\varphi^{-p^2} \Gamma_p(n+p)}{\pi^{p^2} \Gamma_p(n)} \int_{G_2 \in \mathbb{C}^{p \times p}} \det(I + \frac{1}{\varphi} (G_1 - \mu I_{p_1 \times p})^* (G_1 - \mu I_{p_1 \times p}))$$

$$+ (G_2 - \mu I_{p_2 \times p})^*(G_2 - \mu I_{p_2 \times p}))^{-(n+p)} dG_2.$$

Performing the change of variables $G_2 \rightarrow (G_2 + \mu I_{p_2 \times p}) \left(I + \frac{1}{\varphi} (G_1 - \mu I_{p_1 \times p})^* (G_1 - \mu I_{p_1 \times p})\right)^{1/2}$ with a corresponding Jacobian $\det(I + \frac{1}{\varphi} (G_1 - \mu I_{p_1 \times p})^* (G_1 - \mu I_{p_1 \times p}))^{p_2}$,

$$\frac{\varphi^{-p^2} \Gamma_p(n+p)}{\pi^{p^2} \Gamma_p(n)} \det(I + \frac{1}{\varphi} (G_1 - \mu I_{p_1 \times p})^* (G_1 - \mu I_{p_1 \times p}))^{-(n+p-p_2)}$$

$$\times \int_{G_2 \in \mathbb{C}^{p \times p}} \det(I + \frac{1}{\varphi} G_2^* G_2)^{-n-p} dG_2.$$

Integrating using a result from [35],

$$\int_{G_2 \in \mathbb{C}^{p \times p}} \det(I + \frac{1}{\varphi} G_2^* G_2)^{-n-p} dG_2 = \frac{\varphi^{p_2 p \pi^{p_2} \Gamma_{p_2}(n)}}{\Gamma_{p_2}(n + p)}$$

we have the density of $G_1$,

$$\frac{\varphi^{-p_1 p} \Gamma_p(n+p) \Gamma_{p-p_1}(n)}{\pi^{p_1} \Gamma_{p-p_1}(n + p) \Gamma_p(n)} \det(I + \frac{1}{\varphi} (G_1 - \mu I_{p_1 \times p})^* (G_1 - \mu I_{p_1 \times p}))^{-(n+p_1)}.$$
Chapter 4

Performance of Zero-Forcing Precoder with Noisy CSI

In this chapter, we use the distributional results given in Chapter 3 to explore the performance penalty incurred by noisy feedback. We use two standard performance measures: ergodic sum rate and outage probability. Moreover, we examine these performance measures in both the asymptotic regime of high SNR and in the regime of finite SNR. In the asymptotic (in SNR) regime of the ergodic rate, it turns out one must scale the CSI noise variance as a function of SNR in order to prevent the system from becoming interference limited. This result is similar to those obtained in [40], where the number of bits in the CSI are scaled as a function of SNR. We derive the scaling of the logarithmic growth in the ergodic rate (the multiplexing gain). Also, we give precise information on the asymptotic rate penalty due to the noisy feedback. The asymptotics of the outage probability are determined when the noise variance is scaled with SNR and also for a fixed noise variance. We show that, assuming a fixed CSI noise variance, the outage probability converges to a non-zero value at high SNR. Also, we prove that the outage probability converges to zero if the CSI noise variance is scaled polynomially with respect to SNR. The outage is shown to decay to zero polynomially. This rate of decay is determined in
closed form. In the case of finite SNR, we derive closed form expressions that one may use to efficiently compute the outage probability and ergodic rate and also study their behavior by simulations and by examining some special cases.

**Special Function Theory**

We use special function theory extensively in this chapter. Special function theory is an important field in applied mathematics. Its purpose is to study certain functions and their properties. These special functions are often solutions to families of ordinary differential equations; they cannot in general be represented in an elementary closed-form expression but are nonetheless very useful in many situations. Special functions can have thousands of different identities and properties; they are well studied and can be evaluated numerically in an efficient manner [38].

We will use several of these functions in this chapter, the most basic special function we use is the Gaussian hypergeometric function \( _2F_1(a, b; c; x) \) which is a solution to the hypergeometric differential equation,

\[
x (1 - x) \frac{d^2 w}{dx^2} + (c - (a + b + 1) x) \frac{dw}{dx} - abw = 0.
\]

There are many integral expressions for the Gaussian hypergeometric function, for example,

\[
_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(c - b) \Gamma(b)} \int_0^1 u^{b-1} (1 - u)^{c-b-1} (1 - xu)^{-a} \, du,
\]

when \( c > b > 0 \). This special function is of central importance in mathematical physics and applied mathematics, see [38] for detailed information. The other hypergeometric function that will be of central interest throughout much of this chapter is the function \( _3F_2(a_1, a_2, a_3; b_1, b_2; x) \), it is less well studied than the Gaussian hypergeometric function but it is related to it through several integral identities. The identity of most use to us
is,

\[ 3F_2 \left( a_1, a_2, a_3; b_1, b_2; x \right) = \frac{\Gamma (b_2)}{\Gamma (a_3) \Gamma (b_2 - a_3)} \int_0^1 t^{a_3-1} (1 - t)^{-a_3+b_2-1} 2F_1 (a_1, a_2; b_1; tx) \, dt. \]

(4.1)

Another special function that will be interest to us is the psi function, \( \psi (x) \), defined as the logarithmic derivative of the gamma function,

\[ \psi (x) = \frac{d}{dx} \ln (\Gamma (x)). \]

For integer values \( n \), the psi function \( \psi (n) \) can be evaluated in closed form [43],

\[ \psi (n) = -\gamma + \sum_{k=1}^{n-1} \frac{1}{k}. \]

The number \( \gamma = 0.57721\ldots \) is the Euler-Mascheroni constant, see [43]. Finally, one can recognize the sum \( \sum_{k=1}^{n-1} \frac{1}{k} \) as the \((n-1)\)th harmonic number.

### 4.1 The Ergodic Rate

In this section, we examine the ergodic rate of the system and the effect of the CSI noise on this ergodic rate. First, using equation (3.6), the signal power at user \( k \) is \( \vartheta^2 \left| \varphi^{1/2} g_{kk} + \mu \right|^2 \), the interference power is \( \vartheta^2 \varphi \ell_k \) where

\[ \ell_k = \sum_{l \neq k} |g_{k,l}|^2 \]

is the sum of the squares of the interference terms and the AWGN power is \( \sigma_e^2 \). Therefore, one can write the signal to interference and noise ratio (SINR) of the \( k \)th user as

\[ \text{SINR} = \frac{\left| \varphi^{1/2} g_{kk} + \mu \right|^2}{\sigma_e^2 \vartheta^{-2} + \varphi \ell_k}. \]

(4.2)
Moreover, due to the symmetry of the system all receivers have identical SINR, and we may write the ergodic sum rate of the system as,

$$R(P) = p \mathbb{E}\{\log_2 (1 + \text{SINR})\},$$

where $P$ is the total power available for transmission. The power $P$ enters the SINR expression via formula (3.3) which specifies the value of the constant $\varrho$, which is used to satisfy the average power constraint. Substituting the SINR expression (4.2) into the sum rate expression, we can write the ergodic sum rate as,

$$p \mathbb{E}\left\{\log_2 \left(1 + \frac{\varrho^{1/2} g_{kk} + \mu}{1 + \varrho^2 \varphi_k^{1/2}}\right)^2\right\} = p \mathbb{E}\{\log_2 (1 + \alpha |g_{kk} + \beta|^2 + \alpha \ell_k)\} - p \mathbb{E}\{\log_2 (1 + \alpha \ell_k)\}$$

(4.3)

where we used the notation $\alpha = \frac{\varrho^2 \varphi}{\sigma_e^2}$ and $\beta = \frac{\mu \varphi}{\varphi_e \varphi}$ is used. Several assumptions have been tacitly made when computing the ergodic rate above. First, we assume that each receiver has perfect information on its own effective channel, i.e. the vector $\vec{g}_k$. Moreover, the fading and the feedback noise processes change fast enough so as to average the effective channel gain and thus support the bit rate, $R$. We will investigate the case where both assumptions are violated in Section 4.3. Unfortunately, the ergodic rate above cannot be succinctly expressed in closed form. Nonetheless, we can utilize the structure of the effective channel distribution given in Theorem 3 to express the ergodic rate as a simple double integral, and thus allow for numerical computation of the ergodic rate. First, we give an important lemma that will be used throughout.

**Lemma 2.** Let $u$ a $k$-dimensional random vector, distributed as a standard complex $T$-distribution with $\tau$ degrees of freedom, $u \sim T_k(\tau, 0, 1)$, then

$$\mathbb{E}\{\log_2 (1 + \xi \|u\|^2)\} = \frac{k(\xi - 1)}{\tau + 1} F_2 (1, 1, k + 1; 2, \tau + 2; 1 - \xi) + \psi(\tau + 1) - \psi(\tau - k + 1).$$
Proof.

\[
\mathbb{E}\{\log_2 (1 + \xi \|u\|^2)\} = \frac{\Gamma (\tau + 1)}{\pi^k \Gamma (\tau - k + 1)} \int_{\mathbb{S}^{k-1}} \frac{\ln (1 + \xi \|u\|^2)}{(1 + \|u\|^2)^{\tau + 1}} \, du
\]
\[
\overset{(a)}{=} \frac{2 \Gamma (\tau + 1)}{\Gamma (\tau - k + 1) \Gamma (k)} \int_0^{+\infty} \frac{\ln (1 + \xi r^2)}{(1 + r^2)^{\tau + 1}} r^{2k-1} \, dr
\]
\[
\overset{(b)}{=} \frac{\Gamma (\tau + 1)}{\Gamma (\tau - k + 1) \Gamma (k)} \int_0^{+\infty} w^{k-1} \frac{\ln (1 + \xi w)}{(1 + w)^{\tau + 1}} \, dw
\]
\[
\overset{(c)}{=} \frac{\Gamma (\tau + 1)}{\Gamma (\tau - k + 1) \Gamma (k)} \int_0^1 \int_0^{+\infty} \frac{\xi w^k}{(1 + w)^{\tau + 1} (1 + \xi uw)} \, dw \, du
\]
\[
\overset{(d)}{=} \frac{k \xi}{\tau + 1} \int_0^1 \left( \frac{\xi - 1}{\tau + 1} \right) \frac{\Gamma(1, k + 1; 2, \tau + 2; 1 - \xi)}{3F_2(1, 1, k + 1; 2, \tau + 2; 1 - \xi)}
\]
\[
= \frac{k (\xi - 1)}{\tau + 1} \frac{\Gamma(1, k + 1; 2, \tau + 2; 1 - \xi)}{3F_2(1, 1, k + 1; 2, \tau + 2; 1 - \xi)} + \psi (\tau + 1) - \psi (\tau - k + 1).
\]

In equality (a) we switched to hyperspherical coordinates and integrated out the angles [34], in (b) we performed the change of variables \(r^2 \rightarrow w\), in (c) we used the identity

\[\ln (1 + a) = \int_0^1 \frac{a}{1 + au} \, du,\]

and exchanged the order of integration (this exchange is valid since the integrand is positive and thus Tonelli’s theorem holds [37]), in (d) we integrated using formula 3.197.5 of [43]. The final equality holds by a change of variables and using the definition of the \(3F_2\) function (4.1). 

\[
\blacksquare
\]

While the hypergeometric function \(3F_2\), in general, does not have a closed form expression, we observe that in the particular instance of Lemma 2, the function can be expressed as a simple one-dimensional integral over the interval \([0, 1]\). By using the integral representation (4.1) of the hypergeometric function \(3F_2\),

\[
\frac{k (\xi - 1)}{\tau + 1} \frac{\Gamma(1, k + 1; 2, \tau + 2; 1 - \xi)}{3F_2(1, 1, k + 1; 2, \tau + 2; 1 - \xi)}
\]
\[
\begin{align*}
\Gamma (\tau + 1) \Gamma (\tau - k + 1) & \int_0^1 t^{k-1} (1 - t)^{\tau-k} \ln (1 + (\xi - 1) t) \, dt, \\
\end{align*}
\]

since \(2F_1 (1, 1; 2; t (1 - \xi)) = -\frac{1}{t(1-\xi)} \ln (1 + (\xi - 1) t)\) in (4.1). This integral above is over a bounded interval and can be numerically evaluated in a simple manner. We will now proceed to give a formula for the ergodic rate that can be used to easily compute the ergodic rate.

**Theorem 6.** The ergodic sum rate is given by the formula,

\[
R = \frac{p \Gamma (n + 1)}{\pi^{1/2} \Gamma (n - p + 1) \Gamma (p - 1/2)} \int_{\mathbb{R}} \int_{\mathbb{R}^{p-1, \text{UR}}} \ln (1 + \alpha w + \alpha (y + \beta)^2) \frac{dw \, dy}{(1 + w + y^2)^{n+1}} \\
+ p \psi (n - p + 1) - p \psi (n) - \frac{p(p - 1)(\alpha - 1)}{n} 3F_2 (1, 1; p; 2, n + 1; 1 - \alpha) \\
\]

**Proof.** The first term in the rate expression (4.3) can be expressed as,

\[
\begin{align*}
\mathbb{E} \{ \log_2 (1 + \alpha |g_{kk} + \beta|^2 + \alpha \ell_k) \} \\
= & \frac{\Gamma (n + 1)}{\pi^p \Gamma (n - p + 1) \Gamma (p - 1/2)} \int_{\mathbb{R}^{p-1, \text{UR}}} \ln (1 + \alpha \|\mathbf{u}\|^2 + \alpha (y + \beta)^2) \frac{d\mathbf{u}}{(1 + \|\mathbf{u}\|^2 + y^2)^{n+1}} dy \\
= & \frac{2\Gamma (n + 1)}{\pi^{1/2} \Gamma (n - p + 1) \Gamma (p - 1/2)} \int_{\mathbb{R}} \int_{\mathbb{R}^{p-2}} \ln (1 + \alpha r^2 + \alpha (y + \beta)^2) \frac{dr \, dy}{(1 + r^2 + y^2)^{n+1}} \\
= & \frac{\Gamma (n + 1)}{\pi^{1/2} \Gamma (n - p + 1) \Gamma (p - 1/2)} \int_{\mathbb{R}} \int_{\mathbb{R}^{p-1}} \ln (1 + \alpha w + \alpha (y + \beta)^2) \frac{dw \, dy}{(1 + w + y^2)^{n+1}}.
\end{align*}
\]

The second term in equation (4.3) can be evaluated by using Lemma 2. 

The integral representation of the ergodic rate consists of a two dimensional integral of a certain function over the positive half-plane. Software for numerically integrating such integrals is widely available, for example, Mathematica and MATLAB are able to handle such integrals. It is important to note that the formula is valid for any number of users \(p\) and transmit antennas \(n\) and can be used to plot the ergodic sum rate for arbitrarily large systems using fixed computational effort.
4.1.1 Limiting Cases: Feedback Variance and the Ergodic Sum Rate

One expects that the ergodic sum rate is a monotonic function of the CSI noise variance, i.e. as the CSI noise power increases the performance will decrease. Thus, considering the two limiting cases of no CSI noise ($\sigma_f^2 \to 0$) and CSI noise with infinite power ($\sigma_f^2 \to +\infty$) will provide natural upper and lower bounds, respectively. The case of no CSI noise is the easier of the two. By taking the limit, one can compute the sum rate to be,

$$\lim_{\sigma_f^2 \to 0} R = p \log_2 \left( 1 + P \sigma_c^2 (n - p) \right).$$

This sum rate corresponds to the situation where the transmitter has perfect knowledge of the channel matrix and hence is able to invert the channel exactly. This perfect inversion of the channel removes all interference.

The limit of infinite feedback noise variance is more complicated. One can observe from equation (3.6) that as $\sigma_f^2 \to +\infty$, the deterministic component due to the parameter $\mu$ disappears since $\mu \to 0$ as $\sigma_f^2 \to +\infty$. Instead, the effective channel is distributed as a central complex T distribution. Thus, to compute the ergodic sum rate for infinite CSI noise variance, one must evaluate the ergodic-rate when the effective channel is T distributed. Mathematically, as $\sigma_f^2 \to +\infty$, the ergodic rate converges to an ergodic sum rate,

$$\lim_{\sigma_f^2 \to +\infty} R = p \mathbb{E} \left\{ \log_2 \left( 1 + \tilde{\alpha}^2 \sum_l |g_{k,l}|^2 \right) \right\} - p \mathbb{E} \left\{ \log_2 \left( 1 + \tilde{\alpha}^2 \sum_{l \neq k} |g_{k,l}|^2 \right) \right\}, \quad (4.4)$$

where $\tilde{\alpha} = \sigma_c^2 (n - p) \frac{P}{\sigma_c^2}$. Using previous results, we can evaluate the limiting ergodic sum rate in terms of an Euler hypergeometric function.
Theorem 7. As $\sigma_f^2 \to +\infty$, the ergodic sum rate converges to the limit,

$$
\lim_{\sigma_f^2 \to +\infty} R = \frac{1}{n} + \frac{(\alpha - 1)(n - p + 1)}{n(n + 1)} 2F_1(1, p; n + 2; 1 - \hat{\alpha}).
$$

Explicitly, the limiting rate can be written as

$$
\lim_{\sigma_f^2 \to +\infty} R = \frac{1}{n} + \frac{(n - p + 1)}{n(n - p)} \left[ \sum_{k=1}^{n-p+1} \frac{(n - p - k + 1)!}{(n - k + 1)!} \left( \frac{\hat{\alpha}}{\alpha - 1} \right)^{k-1} \right]
$$

$$
- \frac{\hat{\alpha} - 1}{n + 1} \left( \frac{\hat{\alpha}}{\alpha - 1} \right)^{n-p+1} \left( \frac{\ln(\hat{\alpha})}{(\alpha - 1)^{n+2}} + \sum_{k=1}^{n+1} \frac{(\hat{\alpha} - 1)^{-k}}{n - k + 2} \right)
$$

Proof. Applying Lemma 2 to equation (4.4), we obtain the expressions,

$$
\mathbb{E} \left\{ \log_2 \left( 1 + \alpha \sum_l |g_{k,l}|^2 \right) \right\}
$$

$$
= \psi(n + 1) - \psi(n - p + 1) + \frac{p(\alpha - 1)}{n + 1} \, 3F_2(1, 1, p + 1; n + 2; 1 - \alpha),
$$

$$
\mathbb{E} \left\{ \log_2 \left( 1 + \alpha \sum_{l \neq k} |g_{k,l}|^2 \right) \right\}
$$

$$
= \psi(n) - \psi(n - p + 1) + \frac{(p - 1)(\alpha - 1)}{n} \, 3F_2(1, 1, p; n + 1; 1 - \alpha).
$$

We can simplify,

$$
\hat{R} = \frac{1}{n} + (\alpha - 1) \left[ \frac{p}{n + 1} \, 3F_2(1, 1, p + 1; n + 2; 1 - \alpha) - \frac{p - 1}{n} \, 3F_2(1, 1, p; n + 1; 1 - \alpha) \right]
$$

$$
= \frac{1}{n} + (\alpha - 1) \left[ \frac{p}{n} \, 3F_2(1, 1, p + 1; n + 2; 1 - \alpha) - \frac{p - 1}{n} \, 3F_2(1, 1, p; n + 1; 1 - \alpha) \right]
$$

$$
- \frac{p}{n(n + 1)} \, 2F_1(1, p + 1; n + 2; 1 - \alpha)
$$

$$
= \frac{1}{n} + (\alpha - 1) \left[ \frac{1}{n} \, 3F_2(1, 2, p; n + 1; 1 - \alpha) - \frac{p}{n(n + 1)} \, 2F_1(1, p + 1; n + 2; 1 - \alpha) \right]
$$

$$
= \frac{1}{n} + (\alpha - 1) \left[ 2F_1(1, p; n + 1; 1 - \alpha) - \frac{p}{n + 1} \, 2F_1(1, p + 1; n + 2; 1 - \alpha) \right]$$
\[ \frac{1}{n} + \frac{(\alpha - 1)(n-p+1)}{n(n+1)} F_1(1, p; n+2; 1-\alpha). \]

All steps were done by using relations and properties of hypergeometric functions [38], [43] and by observing \( \psi(n+1) - \psi(n) = \frac{1}{n}. \)

### 4.1.2 Simulations

In Section 4.1.1, we derived the limiting values of the ergodic rate for both small and large CSI noise variance. Here, we use the formula of Theorem 6 to numerically study how the value of CSI noise variance affects the ergodic rate. We consider a system with two users and four transmit antennas, with channel fading variance \( \sigma^2_c = 1. \) Figure 4.1 plots the ergodic sum rate with respect to feedback variance \( \sigma^2_f, \) for three different power constraint values. As expected, the ergodic rate drops as the feedback noise variance increases. However, we can observe noticeable differences in the behavior of the ergodic rate for the low and high power regimes. The low power performance is represented by a curve with \( P = 1/2 \) which shows that the feedback noise variance has little impact on the ergodic rate. In contrast, for the other curves with values \( P = 5 \) and \( P = 10, \) one can observe a significant reduction in the ergodic rate due to increasing feedback noise variance.

### 4.2 The Ergodic Rate at High SNR

A key element in understanding the impact of noisy CSI on system performance is its rate in the high power regime. It is clear that as we increase the transmitter power, the interference power will also be increased due to the imperfect interference cancelation. Indeed, if one were to fix all system parameters and increase the power, the system will become interference limited [41]. As observed in [40], in order to avoid such a situation, the quality of the CSI must be improved with increasing transmitted power.
In the following, we will use the affine approximation to the rate curve [39] to study the consequence of such scaling. The affine approximation consists of two figures of merit: the multiplexing gain and the power offset. The multiplexing gain signifies the asymptotic slope of the rate curve as power is increased while the power offset gives the additional power required to transmit at a given rate as compared to the standard AWGN channel. Mathematically, the sum rate \( R(P) \) is well approximated by the affine function,

\[
R(P) = K (\log_2 P - L) + o(1)
\]

where \( K \in \mathbb{R} \) is the multiplexing gain and \( L \in \mathbb{R} \) is the power offset. The multiplexing gain of the system is given as the limit

\[
K = \lim_{P \to \infty} \frac{R(P)}{\log_2 P},
\]
and the power offset is the limit,

\[ L = \lim_{P \to \infty} \left( \log_2 P - \frac{R(P)}{K} \right). \]  

(4.6)

In the large power regime, we use a polynomial scaling \( \sigma_f^2 = P^{-r} \) of the CSI variance as a function of the power where the constant \( r > 0 \) is the CSI scaling factor. To motivate this assumption, we mention [22] where the authors analyze the asymptotic sum rate loss of a broadcast system where the feedback mechanism is an orthogonal training sequence with an MMSE estimator. It is shown in [22] that the CSI variance scales as \( \sigma_f^2 = \frac{1}{P^{1+r}} \). In our setting, this corresponds to \( r = 1 \), an inverse scaling. Therefore, due to its practical significance, we pay special attention to inverse scaling.

It will be shown that for \( r > 1 \) the multiplexing gain is not reduced due to noisy CSI and there is no asymptotic rate penalty. However, when CSI variance is scaled inversely to power (\( r = 1 \)), a rate penalty does exist. We give this rate loss in closed form by expressing it in terms of the \( 3F_2 \) special function. We show that for nearly square channels (i.e. \( n = p + 1 \)) there is a simple closed form related to the harmonic numbers. We conclude with some simulations which examine the relationship between the system dimensions \((n, p)\) and the rate gap. The rate of the noisy CSI system will be denoted as \( R_{\text{noisy}} \) with an associated multiplexing gain and power offset, denoted as \( K_{\text{noisy}} \) and \( L_{\text{noisy}} \) respectively. The multiplexing gain is given in the following theorem.

**Theorem 8.** If the CSI variance is scaled as \( \sigma_f^2 = P^{-r} \) where \( r > 0 \), the multiplexing gain is

\[ K_{\text{noisy}} = \begin{cases} 
 rp & \text{if } 0 < r \leq 1 \\
 p & \text{if } r > 1 
\end{cases} \]

where \( p \) is the number of users in the system.
Proof. Writing the multiplexing gain using definition (4.5),

\[ K_{\text{noisy}} = \lim_{P \to \infty} p \mathbb{E} \left\{ \frac{\log_2 \left( 1 + \frac{\alpha |g_{kk} + \beta|^2}{1 + \alpha \ell_k} \right)}{\log_2 P} \right\} \]

\[ = \lim_{P \to \infty} p \mathbb{E} \left\{ \frac{\log_2 \left( 1 + \alpha \ell_k + \alpha |g_{kk} + \beta|^2 \right)}{\log_2 P} \right\} - \lim_{P \to \infty} p \mathbb{E} \left\{ \frac{\log_2 \left( 1 + \alpha \ell_k \right)}{\log_2 P} \right\} \]

\[ = p \mathbb{E} \left\{ \lim_{P \to \infty} \frac{\log_2 \left( 1 + \alpha \ell_k + \alpha |g_{kk} + \beta|^2 \right)}{\log_2 P} \right\} - p \mathbb{E} \left\{ \lim_{P \to \infty} \frac{\log_2 \left( 1 + \alpha \ell_k \right)}{\log_2 P} \right\} \]

The exchange of limit and expectation is justified in Section 4.5.1. Invoking l’Hôpital’s rule, one can observe the limit of the first term is always equal to 1 while the limit of the second term is \((1 - r)\) for \(0 < r \leq 1\) and 0 for \(r > 1\).

Theorem 8 states that the sum rate asymptotically increases with power so long as the feedback variance is scaled downward with respect to power. Moreover, the scaling factor plays a role in the rate of increase in the sum rate. If the scaling factor is less than one, then the asymptotic rate of growth will be reduced. In contrast, if the feedback variance is decreased at least linearly with power, then a multiplexing gain of \(p\) is preserved.

4.2.1 The Rate Penalty

Thus far we showed that the multiplexing gain is equal to the number of users so long as we scale the CSI quality at least linearly with transmitter power. However, if we scale in a sublinear fashion \((r < 1)\), then the multiplexing gain will be reduced by a factor of \(r\). Another key quantity of interest is the rate penalty due to the noisy CSI. In examining the rate penalty, we must first establish an ideal system by which the penalty will be measured. Thus, we will measure performance against an identical zero-forcing system but whose feedback variance approaches zero, \(\sigma_f^2 \to 0\). The ergodic sum rate can be
easily computed,

\[ R_{\text{ideal}} = p \log_2 \left( 1 + P \frac{\sigma^2_c (n - p)}{\sigma^2_e} \right) \]

\[ = p \left[ \log_2 (P) + \log_2 \left( \frac{\sigma^2_c (n - p)}{\sigma^2_e} \right) \right] + o(1). \quad (4.7) \]

The final equation is an asymptotic expansion for large power, implying that the multiplexing gain of the ideal system is \( K_{\text{ideal}} = p \) and its power offset is \( L_{\text{ideal}} = \log_2 \left( \frac{\sigma^2_c (n - p)}{\sigma^2_e} \right) \).

Moreover, in order to remove the dependence of the rate on the number of users, we will consider a normalized form of the rate penalty. The normalized asymptotic rate penalty is the rate penalty divided by the number of users \( p \), denoted \( \Delta R_{n,p} \). Assuming \( r \geq 1 \), the asymptotic rate penalty can be written,

\[ \Delta R_{n,p} = \lim_{P \to +\infty} \frac{1}{p} (R_{\text{ideal}} - R_{\text{noisy}}) \]

\[ = \frac{K_{\text{ideal}}}{p} (\log_2 P - L_{\text{ideal}}) - \frac{K_{\text{noisy}}}{p} (\log_2 P - L_{\text{noisy}}) + o(1) \]

\[ = L_{\text{noisy}} - L_{\text{ideal}} + o(1) \quad (4.8) \]

where the last equality follows since \( K_{\text{ideal}} = K_{\text{noisy}} = p \) for \( r \geq 1 \) as stated in Theorem 8. This measure can be understood as the asymptotic rate penalty incurred by any given user. We now proceed to show that the normalized rate penalty is zero for a CSI scaling factor \( r > 1 \).

**Theorem 9.** Given that the CSI quality is scaled as in Theorem 8 and \( r > 1 \), then \( \Delta R_{n,p} = 0 \).

**Proof.** Similar to Theorem 8, exchange the limit and expectation operations (see justification in Section 4.5.2). Then one needs the limit

\[ L_{\text{noisy}} = \lim_{P \to +\infty} \left( \log_2 P - \log_2 \left( 1 + \alpha |g_{kk} + \beta|^2 + \alpha \ell_k \right) + \log_2 (1 + \alpha \ell_k) \right) \quad (4.9) \]
Substituting the CSI quality policy $\sigma_f^2 = P^{-r}$, the limit is evaluated to be $L_{\text{noisy}} = \log_2 \left( \frac{\sigma_f^2}{\sigma_f^2(n-p)} \right)$. Since $L_{\text{noisy}}$ is equal to the power offset of the ideal system (see equation (4.7)), then equation (4.8) implies the theorem statement.

Of course, scaling the CSI such that $r > 1$ may be quite taxing in terms of system resources, thus the next theorem addresses the situation where the CSI quality is scaled linearly. It is shown that in the linear case, there is indeed a rate penalty.

**Theorem 10.** Assume that CSI variance is scaled linearly (i.e. $r = 1$), then at high SNR, the rate loss is,

$$
\Delta R_{n,p} = \psi(n) - \psi(n-p+1) + \frac{(p-1)(n-p-1)}{n} F_2(1,1,p;2,n+1;1+p-n).
$$

(4.10)

**Proof.** Using definition (4.6) and Theorem 8, write the power offset of the noisy CSI scheme,

$$
L_{\text{noisy}} = \lim_{P \to \infty} \log_2 P - \mathbb{E}\left\{ \log_2 \left( 1 + \frac{\alpha |g_k + \beta|^2}{1 + \alpha \ell_k} \right) \right\}
$$

$$
= \mathbb{E}\left\{ \lim_{P \to \infty} \log_2 P - \log_2 \left( 1 + \frac{\alpha |g_k + \beta|^2}{1 + \alpha \ell_k} \right) \right\}
$$

$$
= \mathbb{E}\left\{ \log_2 \left( \frac{\sigma_f^2 (1 + (n-p) \ell_k)}{(n-p) \sigma_c^2} \right) \right\}
$$

$$
= \mathbb{E}\left\{ \log_2 (1 + (n-p) \ell_k) \right\} + \log_2 \left( \frac{\sigma_c^2}{\sigma_c^2 (n-p)} \right).
$$

The exchange of expectation and limit is justified in Section 4.5.2. By (4.7) and (4.8), the asymptotic normalized rate penalty is

$$
\mathbb{E}\left\{ \log_2 (1 + (n-p) \ell_k) \right\},
$$

which can be evaluated using Lemma 2 to arrive at (4.10).
A special case when the rate gap is expressible in simpler terms is when \( p = n - 1 \), i.e. there is one additional transmit antenna than the number of users. It can be shown that the rate penalty is equal to the \( p \)th harmonic number minus 1,

\[
\Delta R_{p+1,p} = \sum_{k=1}^{p} \frac{1}{k} - 1.
\]

Moreover, using the asymptotic expansion of harmonic numbers,

\[
\Delta R_{p+1,p} \sim \ln(p) + \gamma - 1 + O\left(\frac{1}{p}\right). \tag{4.11}
\]

Therefore, in the limit of a large number of users, the rate penalty is logarithmic in the number of users.

### 4.2.2 Simulations

We plot the asymptotic rate penalty given in Theorem 10 in Figure 4.2. The evaluation is performed for three scenarios: two users, four users and six users. For each scenario, the rate penalty is evaluated for up to twenty-five transmit antennas. First, we observe that as the number of users increases, the penalty incurred by each user is increased. It can be inferred from Figure 4.2 that the rate penalty is approximately linear in the number of users for a small number of users. This is in contrast to the behavior of the rate penalty for a large number of users, see equation (4.11), where the rate penalty increases logarithmically with the number of users. Also, we note that increasing the number of transmit antennas for a fixed number of users will increase the rate penalty but this penalty will quickly saturate. Moreover, the additional penalty due to additional transmit antennas becomes more severe with a larger number of users.
4.3 Outage Probability

Here, we study the outage probability associated with the zero forcing broadcast system. We define the outage probability to be the probability of an outage for a single user,

\[ P_{\text{out}}(R) = \mathbb{P} \left[ \ln (1 + \text{SINR}) \leq R \right]. \]

Recall that due to the symmetry of the system, all receivers have identical SINR, hence \( k \) does not appear on the left hand side of the equation above. We begin with a formula that allows for a simple numerical computation of the outage probability.

**Theorem 11.** The outage probability is given as the integral,

\[
P_{\text{out}}(R) = \int_{0}^{+\infty} \int_{-1}^{1} \sqrt{px + \rho \alpha^{-1}} D \left( x, y \sqrt{px + \rho \alpha^{-1}} - \beta \right) \, dy \, dx
\]
where,
\[
D(x, y) = \frac{\Gamma(n + 1)}{\pi^{1/2} \Gamma(n - p + 1) \Gamma(p - 1/2)} \frac{x^{p-3/2}}{(1 + x + y^2)^{n+1}},
\]
and \(\rho = 2^R - 1\).

**Proof.** The outage probability can be expressed as,
\[
P_{out}(\rho) = \mathbb{P} \left[ \frac{\alpha(g_{k,k} + \beta)^2}{1 + \alpha \ell_k} \leq 2^R - 1 \right].
\]
Moreover, it is straightforward to show that the joint probability density of \(g_{k,k}\) and \(\ell_k\) is given as,
\[
D(\ell_k, g_{k,k}) = \frac{\Gamma(n + 1)}{\pi^{1/2} \Gamma(n - p + 1) \Gamma(p - 1/2)} \frac{\ell_k^{p-3/2}}{(1 + \ell_k + g_{k,k}^2)^{n+1}}.
\]

We can express the outage probability in a simple form with the following steps. Note that the outage probability can be expressed as,
\[
P_{out}(\rho) = \mathbb{P} \left[ (g_{k,k} + \beta)^2 \leq \rho \ell_k + \frac{\rho}{\alpha} \right],
\]
where \(\rho = 2^R - 1\). The outage probability can be manipulated to yield,
\[
P_{out}(\rho) = \int_0^{+\infty} \int_{(y+\beta)^2 \leq \rho x + \frac{\rho}{\alpha}} D(x, y) \ dy \ dx
\]
\[
= \int_0^{+\infty} \int_{-1}^{1} \sqrt{\rho x + \rho \alpha^{-1}} D\left(x, y\sqrt{\rho x + \rho \alpha^{-1}} - \beta\right) \ dy \ dx
\]

\[
\square
\]

### 4.3.1 Outage Probability at High SNR

In a similar fashion to the ergodic rate, one is interested in studying the asymptotic behavior of the outage probability at high SNR. The primary figure of merit associated
with the asymptotic characterization of the outage probability is the *outage rate of decay*, defined as,

\[ d = - \lim_{P \to +\infty} \frac{\log P_{\text{out}}}{\log P}. \]

The number \( d \) is the polynomial rate of decay of the outage probability to zero, as the SNR grows large. As in the ergodic rate case, we will study the situation where the feedback variance is scaled as the SNR grows and compare it to the case where no feedback noise scaling is present. In the following theorem, we calculate the value of \( d \) and give some additional information.

**Theorem 12.** Assume the feedback variance is scaled as \( \sigma_f^2 = P^{-r} \) for \( r \geq 0 \). The outage rate of decay is, \( d = r(n - p + 1) \). The outage probability behaves asymptotically as,

\[ P_{\text{out}} \approx \Lambda(n, p, \beta, \sigma_c^2) P^{-d}. \]

The constant \( \Lambda(n, p, \beta, \sigma_c^2) \) is equal to the integral,

\[ \frac{\sigma_c^{-2(n-p+1)} \Gamma(n+1)}{\pi^{1/2} \Gamma(n-p+1) \Gamma(p-1/2)} \int_{-1}^{1} \int_{0}^{+\infty} \frac{\sqrt{\rho u^{p-1}}}{(u + (y \sqrt{\rho u} - 1)^2)^{n+1}} dy du. \]

**Proof.** See Section 4.5.3.

The theorem above states that the outage probability decays to zero as the power is increased and the feedback variance is scaled appropriately. Moreover, the outage probability decays to zero polynomially with an order \( r(n - p + 1) \). This order of decay is controlled by the number of additional transmit antennas as compared to the number of users, a kind of transmit diversity. One may regard this result as analogous to the multiplexing gain given in Theorem 8 which states that the multiplexing gain is \( rp \). Next we examine the case where the feedback variance is fixed with respect to power.

**Corollary 13.** Assuming a fixed feedback variance, \( \sigma_f^2 \) and letting \( P \to +\infty \), the outage
probability will converge to a fixed value

\[ \lim_{P \to +\infty} P_{\text{out}} = \Lambda(n, p, \beta, \sigma_c^2) \]

Proof. Set \( r = 0 \) in Theorem 12.

The corollary states that the outage probability will converge to a fixed value \( \Lambda(n, p, \beta, \sigma_c^2) \) that depends on various system parameters. Note that this value is not zero and this suggests that the system will become interference limited, in other words, increasing power does not increase performance.

### 4.3.2 Simulations

Here we present a numerical study of the outage probability. Consider Figure 4.3 where the logarithm of the outage probability is plotted against the number of transmit antennas. The system parameters are fixed at the values \( \sigma_c^2 = 1, \sigma_f^2 = 0, \gamma = 10, \rho = 0.5 \).

The number of transmit antennas varies from 3 to 25 antennas for three different scenarios of \( p = 2, p = 4 \) and \( p = 6 \). We observe a clear linear relationship between the logarithm of outage and the number of transmit antennas. This suggests the outage probability decreases exponentially with the number of transmit antennas. Moreover, we note that for a fixed value of outage, increasing the number of users requires increasing the number of transmit antennas by the same amount. Thus, the outage probability strongly depends on the difference of the number of transmit antennas and number of users. Figure 4.4 plots the outage probability versus the feedback variance for three different for three different power constraint values, as expected the outage probability increases with increasing CSI noise variance.
Figure 4.3: $\log(P_{out})$ vs. number of transmit antennas.

Figure 4.4: Outage probability vs. feedback variance.
4.4 Summary

We investigated the performance of the broadcast system introduced in Chapter 3 in terms of the ergodic sum rate and outage probability. For the ergodic rate, we give a formula that can be easily evaluated numerically. The limiting cases of zero feedback noise variance and infinite feedback noise variances are discussed and the ergodic rates are given in closed form. The ergodic rate in the regime of infinite feedback noise variance is expressed in terms of an Euler hypergeometric function. In the high SNR regime, it was proven that the multiplexing gain is maintained so long as the CSI quality increases linearly with increased transmitter power. Also, we show that there is no rate loss when the CSI quality is scaled super-linearly with respect to transmitter power. We find that at linear scaling, there is a rate loss. The rate penalty for linear scaling was given in closed form in terms of special functions.

Similarly, the outage probability is expressed as an integral expression that can be evaluated numerically. Also, we showed that for a fixed feedback variance, the system will become interference limited. However, if the feedback noise variance is scaled appropriately, the outage probability will go to zero. The rate at which the outage probability goes to zero, a quantity analogous to the diversity order of single user systems, is found in closed form.

4.5 Appendices

4.5.1 Appendix: Theorem 8

The multiplexing gain is written as,

$$\lim_{P \to \infty} P \mathbb{E} \left\{ \log_2 \left( 1 + \frac{\alpha |g_{k,k} + \beta|^2}{1 + \alpha \ell_k} \right) \right\}$$

$$\log_2 P$$
\[= \lim_{P \to \infty} p\mathbb{E}\left\{ \frac{\log_2 \left( 1 + \alpha \ell_k + \alpha |g_k + \beta|^2 \right)}{\log_2 P} \right\} - \lim_{P \to \infty} p\mathbb{E}\left\{ \frac{\log_2 \left( 1 + \alpha \ell_k \right)}{\log_2 P} \right\} \]

We show the exchange of limit and integral is valid by applying the dominated convergence theorem [37] to each term. Substituting \( \sigma_f^2 = P - r \), we bound the second term by an integrable function as follows,

\[
\frac{\log_2 (1 + \alpha \ell_k)}{\log_2 P} = \frac{1}{\log_2 P} \log_2 \left( 1 + \frac{(n - p) \sigma_c^2 P^{1-r}}{\sigma_f^2 + P^{1-r} \ell_k} \right) < \frac{1}{\log_2 P} \log_2 \left( 1 + (n - p) P^{1-r} \ell_k \right) \leq (a) \log_2 \left( 1 + (n - p) 2^{1-r} \ell_k \right) < (n - p) 2^{1-r} \ell_k
\]

where \((a)\) is true since the function is decreasing with respect to \( P \) everywhere except \( P = 1 \) (uniformly in \( \ell_k \)). The expectation of the final bound is but the sum of second moments of \( T \) random variables, these are finite for \( n > p \). We conclude the function is bounded by an integrable function for \( n > p \). The first term in the ergodic rate expression is bounded by an identical argument,

\[
\frac{\log_2 \left( 1 + \alpha \ell_k + \alpha |g_{kk} + \beta|^2 \right)}{\log_2 P} < \frac{\log_2 \left( 1 + \alpha \left(|g_{kk}|^2 + \ell_k \right) \right)}{\log_2 P}
\]

and we simply repeat the argument above to show it is bounded by an integrable function. Thus the dominated convergence theorem applies [37] and we are done.
4.5.2 Appendix: Theorems 10 & 9

Lemma 3. For $r \geq 1$, the following bound holds,

$$\log_2 \left( x^{-r-1} + (kx^{-r/2} + 1)^2 \right) \geq \log_2 \left( \frac{1/2}{k^4 + 1} \right).$$

Proof. Consider the change of variables $x = y^{-2/r}$, then the expression above becomes,

$$f_d(y) = \log_2 \left( y^d + (ky + 1)^2 \right).$$

where $d = 2(r + 1)/r$. Note that $2 \leq d \leq 4$ since $d$ is monotonic in $r$ and $\lim_{r \to +\infty} 2(r + 1)/r = 2$ while $2(r + 1)/r = 4$ when evaluated at $r = 1$. Therefore,

$$f_d(y) \geq \min \left\{ \min_y (f_2(y)), \min_y (f_4(y)) \right\}.$$

By taking a derivative, it is straightforward to show,

$$\min_y (f_2(y)) = \ln \left( \frac{1}{k^2 + 1} \right).$$

Also, one can find the minimum of $f_4(y)$ exactly by equating the derivative to zero and solving the resulting quartic equation. We do not give the details for the calculation due to its length, but by using the formula for the exact minimum and some elementary bounds, one can show,

$$f_4(y) \geq \ln \left( \frac{1/2}{k^4 + 1} \right).$$

The bound in the lemma statement follows.

As before, we show the validity of exchanging limit and integration by appealing to
the dominated convergence theorem. Consider the following inequalities,

\[
\log_2 P - \log_2 \left( 1 + \frac{\alpha |g_{k,k} + \beta|^2}{1 + \alpha \ell_k} \right) \\
= \log_2 \left( \frac{\sigma_c^2 + P^{-r} + P^{1-r} (n-p) \sigma_c^2 \ell_k}{P^{-r-1} + P^{-1} \sigma_c^2 + P^{-r} (n-p) \ell_k + P^{-r} (n-p) \sigma_c^2 (g_{k,k} + \sigma_c^2 P^{r/2})^2} \right) \\
< - \log_2 \left( P^{-r-1} + P^{-1} \sigma_c^2 + P^{-r} (n-p) \ell_k + P^{-r} (n-p) \sigma_c^2 (g_{k,k} + \sigma_c^2 P^{r/2})^2 \right) \\
+ \log_2 \left( 1 + \sigma_c^2 + (n-p) \sigma_c^2 \ell_k \right) \\
< \log_2 \left( 1 + \sigma_c^2 + (n-p) \sigma_c^2 \ell_k \right) - \log_2 \left( P^{-r-1} + (n-p) \sigma_c^2 (g_{k,k} P^{-r/2} + \sigma_c^2)^2 \right) \\
= \log_2 \left( 1 + \sigma_c^2 + (n-p) \sigma_c^2 \ell_k \right) - \log_2 \left( (n-p) \sigma_c^3 P^{-r-1} + \left( \frac{g_{kk}}{\sigma_c^2} P^{-r/2} + 1 \right)^2 \right) \\
- \log_2 \left( (n-p) \sigma_c^3 \right) \\
< \log_2 \left( 1 + \sigma_c^2 + (n-p) \sigma_c^2 \ell_k \right) + \log_2 \left( g_{kk}^4 \sigma_c^{-8} \left( (n-p) \sigma_c^3 \right)^{\frac{r+1}{r}} + 1 \right) \\
- \log_2 \left( \frac{1}{2} (n-p) \sigma_c^3 \right) \\
< \sigma_c^2 + (n-p) \sigma_c^2 \ell_k + g_{kk}^2 \sigma_c^{-4} \left( (n-p) \sigma_c^3 \right)^{\frac{r+1}{r}} - \log_2 \left( (n-p) \sigma_c^3 \right) + \log_2 2.
\]

The final bound is integrable since the second moment of the T distribution exists for \( n > p \). This concludes the proof.

### 4.5.3 Appendix: Theorem 12

The outage probability can be written,

\[
P_{out} = \int_0^{\infty} \int_{-1}^1 \sqrt{px + \rho x^{-1}} D \left( x, y \sqrt{px + \rho x^{-1}} - \beta \right) dy \, dx \\
= \frac{\Gamma (n+1)}{\pi^{1/2} \Gamma (n-p+1) \Gamma (p-1/2)} \int_0^{\infty} \int_{-1}^1 \frac{dy \, dx \sqrt{px + \rho x^{-1} x^{p-3/2}}}{\left( 1 + x + (y \sqrt{px + \rho x^{-1}} - \beta)^2 \right)^{n+1}} \\
= \frac{\beta^{-2(n-p+1)} \Gamma (n+1)}{\pi^{1/2} \Gamma (n-p+1) \Gamma (p-1/2)} \int_0^{\infty} \int_{-1}^1 \frac{dy \, du \sqrt{pu + \frac{u}{\beta^2} u^{p-3/2}}}{\left( \frac{1}{\beta^2} + u + (y \sqrt{pu + \frac{u}{\beta^2}} - 1)^2 \right)^{n+1}}.
\]
where the final equality was a change of variables $x \rightarrow \beta^2 u$. Note that,

$$\beta = \frac{\sigma_c}{\sigma_f} = \sigma_c P^{r/2}$$

and,

$$\alpha\beta^2 = \sigma_c^3 (n - p) P^{1+r/2},$$

this implies $1/ (\alpha\beta^2) \to 0$ and $1/\beta^2 \to 0$ as $P \to +\infty$. Therefore,

$$\lim_{P \to +\infty} P^{r(n-p+1)} P_{\text{out}}$$

$$= \frac{\sigma_c^{-2(n-p+1)} \Gamma (n+1)}{\pi^{1/2} \Gamma (n-p+1) \Gamma (p-1/2)} \int_{1}^{+\infty} \int_{-1}^{1} \frac{\sqrt{\rho w^{p-1}}}{(u + (y\sqrt{\rho w} - 1)^2)^{n+1}} dy du$$

$$\triangleq \Lambda (\beta, n, p).$$

where the limit is evaluated by exchanging the order of integration and limit operations.

We can make this calculation rigorous by first making a change of variables to obtain the integral,

$$\int_{0}^{+\infty} \int_{1}^{1} \frac{1}{\left( \frac{1}{\alpha^2} + w - \frac{1}{\alpha^2} + (y\sqrt{\rho w} - 1)^2 \right)^{n+1}} dy dw.$$
for sufficiently large $\gamma$. The final bound is absolutely integrable over $[0, +\infty) \times [-1, 1]$ and thus the dominated convergence theorem holds [37].
In this chapter, we consider the design of linear precoders for MISO BC channels with imperfect information on the channel. As discussed in Chapter 2, the transmitter uncertainty due to the flawed CSI is modeled as a probability distribution on the channel. The randomness in the channel induces a non-zero probability, from the perspective of the transmitter, that the channel cannot support a desired rate, i.e. a non-zero outage probability, just as in a fading channel. This is regardless of whether the physical channel is in fact time-varying. The transmitter’s belief in the value of the channel matrix $H$ is captured in the probability distribution it assigns to $H$, hence the outage probability computed based on that distribution is its own view of how likely it is to encounter an outage event. If it does transmit at the desired rate, the outage probability observed will be different from the transmitter’s computed value as long as the actual channel $H$ varies over time in a manner different from the assumed distribution. In this sense, outage probability due to imperfect channel knowledge is not the same as that due to channel variations with perfect channel knowledge at both ends of the link – the latter outage probability will be observed when transmitting at the desired rate, while the former will
not. For that reason, outage probability due to channel uncertainty is more appropriately seen as a *risk* undertaken by the transmitter, given the information available to it. Like a gambler whose bet is risky only because he does not know the cards in the dealer’s hand, our transmitted rate is risky only because the transmitter does not know the channel.

First, we formulate the design problem as trading off rate versus risk in a manner similar to the outage formulation in fading channels. Thus, the solution is no longer unique but consists of the efficient frontier of all precoding matrices that achieve maximal rate for a given risk or minimal risk for a given rate. We study some properties of the risk function and show that for each tradeoff point, minimal risk or maximal rate can be achieved using a full power precoder. However, finding the optimal precoder is still difficult due to the high dimensional integrations required. Therefore, we opt to solve a related mean-variance optimization problem. This is a standard approach in financial mathematics that is also termed *modern portfolio theory*. In this context, one may regard the rate-risk formulation as a financial return versus a risk of investment loss. The quantity that is being ”invested“ is a fixed amount of power that is allocated to each entry of the precoder matrix (rather than a sum of money).

Next, we assume a Gaussian mixture model (GMM) probability distribution for the channel. This is a flexible model, often used in signal processing tasks, that allows one to fit many different distributions by varying the weights and parameters of the various mixture components. Theoretically, it has been shown [42] that GMMs are universal approximators in the sense that they may be made close (in a suitable sense) to any continuous function for a sufficient number of mixture components. By using the GMM, we rewrite the mean and variance of the rate, which are given as integrals over the channel density, to simple integral expressions that can be computed efficiently. The optimization procedure itself can then be implemented using standard numerical optimization techniques.
5.1 System Model

We study a multiple input single output (MISO) broadcast (BC) channel, where the channel outputs at all receivers are collected in the vector

\[ y = Hx + n, \]

\( x \in \mathbb{C}^p \) is channel input, and \( H \in \mathbb{C}^{p \times p} \) is the complex valued channel matrix. The noise term, \( n \) is distributed as a zero-mean uncorrelated \( p \)-dimensional, circularly symmetric, complex Gaussian random variable with variance \( \sigma_e^2 \). The weighted sum-rate for a fixed channel \( H \) and weighting vector \( \omega \) is denoted,

\[ R_\omega (H) = \sum_{i=1}^p \omega_i \ln \left( 1 + \frac{h_{ii}^2}{\sigma_e^2 + \sum_{j \neq i} h_{ij}^2} \right). \]

for a \( p \)-vector \( \omega = [\omega_1, \ldots, \omega_p]^T \) whose entries sum to one. When we work with the standard rate function, where all user rates are equally weighted, we will use the notation \( R (H) = pR_\omega (H) \) when \( \omega = [1/p, \ldots, 1/p]^T \). Suppose that, in order to enhance system performance, we introduce the linear precoder, \( D \in \mathbb{C}^{p \times p} \), which operates on the symbol vector \( u \),

\[ y = HDu + n. \]

The weighted sum-rate for this system is \( R_\omega (HD) \). In the following we are interested in the case where the channel matrix \( H \) is not exactly known at the transmitter, however to introduce some ideas suppose the channel matrix \( H \) is exactly known at the transmitter. As mentioned in Chapter 2 one should choose \( D \) to maximize the rate \( R (HD) \), subject to transmitter power restrictions. Thus, we can write the fixed channel precoder design problem as,

\[ \max_{D \in \mathbb{C}^{p \times p}} R_\omega (HD) \tag{5.1} \]
where $\mathcal{F}_P$ denotes the set of precoders that satisfy the power constraint,

$$\mathcal{F}_P = \left\{ D \in \mathbb{C}^{n \times n} \text{ such that } \text{tr} (D^* D) \leq P \right\}.$$ 

Note that a solution exists since $R_\omega (H D)$ is a continuous function of $D$ and $\mathcal{F}_P$ is a non-empty compact set.

5.1.1 Stochastic Channel Information - A Risk-Rate Formulation

When the transmitter does not know the channel exactly, one cannot formulate the optimization problem (5.1). Nonetheless, in most cases of interest, some information on the channel is available, e.g. a channel estimate, quantized channel information or some other prior information. Indeed, in practice it is quite reasonable to assume some prior information is available. By using this CSI, one can obtain a probability distribution over the channel that reflects the transmitter’s knowledge of the channel.

Suppose that the physical channel is contained in a set $\mathcal{S}$. Using the CSI, we obtain a probability measure on $\mathcal{S}$ and let $q = \mathbb{P} [H \in \mathcal{A}]$. One can consider $q$ as the weight of all possible channels in the set $\mathcal{A}$ as compared to the channel ensemble $\mathcal{S}$. In this context, we are interested in the overall weight of all channels that cannot support a weighted rate $\varphi$ for a given precoder $D$,

$$I_{\varphi} (D) = \mathbb{P} [R_\omega (H D) < \varphi].$$

This expression can also be regarded as the risk that the weighted rate $\varphi$ cannot be supported by the physical channel. An alternative measure of performance is the quantile function,

$$W_{\alpha} (D) = \inf \{ \varphi : I_{\varphi} (D) \geq \alpha \},$$
which can be interpreted as giving the rate achieved with risk $\alpha$ as a function of the precoder. We are interested in finding the precoder that will maximize the rate with risk $\alpha$, i.e.

$$D^\dagger_\alpha = \arg \max_{D \in \mathcal{F}_P} W_\alpha(D).$$

(5.2)

where $\mathcal{F}_P$ is the set of possible precoders, determined by the power constraint, $\mathcal{F}_P = \{D \in \mathbb{C}^{p \times p} : \text{tr}(D^*D) \leq P\}$. We are also interested in the precoder that minimizes risk for a given rate, i.e.

$$D^\dagger_\phi = \arg \min_{D \in \mathcal{F}_P} I_\phi(D).$$

(5.3)

We begin by establishing some continuity and monotonicity properties,

**Theorem 14.** Assume the random channel $H$ has a probability density function, then the following statements hold,

1. $W_\alpha(D)$ is strictly increasing with respect to $\alpha$ for $D$ invertible.

2. $I_\phi(D)$ is strictly increasing with respect to $\phi$, so long as the pdf of $H$ is strictly positive.

3. $W_\alpha (D)$ is continuous in both $\alpha$ and $D$ (with respect to the product topology), so long as the pdf of $H$ is strictly positive on $\mathbb{R}^{p^2}$.

4. $I_\phi(D)$ is continuous on $\phi$ for $D$ invertible.

*Proof.* See Appendix 5.5.1.

We further note that $W_\alpha(D)$ is often everywhere differentiable with respect to $D$, for example when the pdf is compactly supported\(^1\) (see Lemma 2.30, [44]). In the next theorem, we show that the precoder that achieves the minimal risk $\alpha$ for a given weighted rate $\phi$ is identical to the one that achieves maximal weighted rate $\phi$ for a given risk $\alpha$.

\(^1\)A common example would be to restrict the channel to lie in an ellipsoidal region; this region models the uncertainty in the channel knowledge. This would correspond to a uniform distribution on the ellipsoidal region (and a compactly supported pdf).
Theorem 15. Full power precoders achieve all tradeoff points between risk and rate. Also, a full power precoder which maximizes rate for a given risk will also minimize risk for some fixed rate.

Proof. See Appendix 5.5.2.

The risk can only be computed when the transmitter knows the CSI probability distribution. We proceed to show in the next section that given two types of channel information sources, the CSI distribution can be found.

5.1.2 Constructing the CSI probability distribution

Thus far, the framework given above is very general and many physical mechanisms that would induce channel uncertainty could be mentioned. However, it is not clear how one can obtain the probability distribution that specifies the channel uncertainty. Here, we discuss two concrete systems that induce channel uncertainty and show how the probability distribution of the channel can be obtained. The two scenarios are: finite rate channel feedback and receiver channel estimation by pilot symbols. In each case, there is no natural probability distribution on the channel and we must make a choice or introduce a selection principle to obtain a channel probability distribution. In the case of a finite rate feedback, we select the maximum entropy distribution while in the scenario of channel estimation by pilot symbols, we introduce a prior distribution for the channel that leads us to a probability distribution for the channel.

Finite Rate Feedback

In this situation, the transmitter has access to a quantization of the channel matrix. This is done by assuming that the receivers have exact knowledge of their own channel and are able to feed back only a quantized version of it. Moreover, we assume that the \( i \)th receiver’s channel lies in a compact subset \( Z_i \) which is a subset of the space \( \mathbb{C}^p \), \( Z_i \subset \mathbb{C}^p \).
For example, assuming that each channel’s energy is at most $P_i$, then $Z_i$ would be the ball that contains all channels whose energy is at most $P_i$. For simplicity, we assume that all subsets $Z_i$ are identical and denote the set in which all individual channels must lie by $Z$. Each user quantizes its own channel in the following manner: consider a partitioning of the set $Z$ into $2^k$ disjoint sets and number them, $Q_1, \ldots, Q_{2^k}$. Hence, $k$ bits are sufficient to number all partitions. The quantized channel vector of the $i$th user, denoted as $b_i$, is the label of the partition inside of which the channel $h_i$ lies. It is this label that is fed back to the transmitter by all receivers. The quantized channel matrix $B$, is formed at the transmitter by stacking all quantized user channels $b_i$, $i = 1, \ldots, p$. This matrix can be used by the transmitter in designing the precoder.

Such a system does not immediately induce a probabilistic distribution on the channel, but we will use the maximum entropy principle to construct one. Since the quantized channel of the $i$th user, $b_i$, is fixed and known to the transmitter, the probability distribution of the channel will be parameterized by it. Moreover, the quantized channel indicates that the channel must lie in the partition $Q_{b_i}$, where $b_i$ is the label represented by the vector $b_i$. The maximum entropy distribution that satisfies this condition is the uniform distribution over $Q_{b_i}$,

$$\mathbb{P}[h_i \in A; b_i] = \frac{\int_{A \cap Q_{b_i}} dX}{\int_{Q_{b_i}} dX},$$

for every appropriate subset $A \subset \mathbb{C}^p$. Since each user quantizes its channel independently of all other users’ channels, it is natural to assume the corresponding probability distributions will also be independent of each other. Thus, the distribution of the full channel matrix is given as the product of the marginal distributions,

$$\mathbb{P}[H \in A; B] = \prod_{i=1}^{p} \mathbb{P}[h_i \in A_i; b_i],$$

for every appropriate subset $A \subset \mathbb{C}^{p \times p}$ and where $A_i$ is the image of projection function...
which takes a matrix in $\mathbb{C}^{p \times p}$ to its $i$th row.

**Channel Estimation with Pilot Symbols**

Here, we assume the communication is divided into two distinct stages. First, the transmitter sends pilot symbols for $t$ channel uses and the $i$th receiver forms an estimate, $\hat{h}_i$, of its own channel by using these pilot symbols and the received symbols. Following [22], we consider a collection of $t$ pilot-symbol vectors, and collect them in an $p \times t$ matrix $X$. The symbols received by the $i$th receiver during the training stage are written,

$$z_i = X^T h_i + n$$

where $z_i$ is a $t$-dimensional vector of all received symbols. Assuming the channel estimation is performed by a maximum mean square error (MMSE) estimator, the channel estimator is given by the formula,

$$\hat{h}_i = X^* (I + X^T X^*)^{-1} z_i.$$  \hspace{1cm} (5.4)

All $p$ estimates (from all $p$ receivers) are then noiselessly fed back to the transmitter and used in constructing a precoder. We assume the channel estimation is performed by a maximum likelihood (ML) estimator. For simplicity, let us assume the training symbols are orthogonal,

$$X^T X^* = P I,$$

where $P$ is the training power. Then we obtain the channel estimate,

$$\hat{h}_i = \frac{P}{P + 1} h_i + \frac{1}{1 + P} X^* n.$$  \hspace{1cm} (5.4)

We wish to obtain a probability distribution $\mathbb{P}[h_i | \hat{h}_i]$ and use this distribution as the channel probability distribution associated with channel uncertainty. In order to do so
we must introduce a prior distribution on the channel $h_i$, $p[h_i]$ and then use Bayes’ theorem,

$$f(h_i | \tilde{h}_i = r) = f_{\tilde{h}_i|h_i}(r | h_i) \frac{f(h_i)}{f_{\tilde{h}_i}(r)}. \tag{5.5}$$

where we have $\tilde{h}_i | h_i \sim \mathcal{N}\left(\frac{P}{p+1} h_i, \frac{P}{(1+p)^2} I\right)$ and $f_{\tilde{h}_i}(r)$,

$$f_{\tilde{h}_i}(r) = \int_{C^N} f_{\tilde{h}_i|h_i}(r | h_i = x) f_{h_i}(x) \, dx. \tag{5.6}$$

The overall probability distribution of the channel is then the product of distributions given in (5.5). Note that in this case, we took a prior probability distribution on the channel that assumes different users’ channels are independent and thus we can use the marginal distribution of every user’s channel in Bayes’ formula. One can easily allow for any prior distribution of the channel by taking a channel prior distribution to be over the matrix space $\mathbb{C}^{p \times p}$ and applying Bayes’ theorem in an identical manner to equation (5.5). Also, the integration in equation (5.6) will then be over the matrix space $\mathbb{C}^{N \times N}$.

### 5.2 Finding Good Precoders

Even assuming the CSI probability distribution is known, the risk-rate optimization problem is still difficult due to two separate issues. First, the optimization is non-convex over $2N^2$ variables, though nonlinear optimization algorithms are applicable (assuming $p$ is not too large). The more pressing issue is that of numerically evaluating the risk, which is critical since it is a prerequisite to applying any numerical nonlinear optimization techniques. Unfortunately, directly evaluating the risk seems intractable as it would require integration over $p^2$-dimensional space. In the following, we propose to solve a related (heuristic) mean variance optimization problem that enables one to compute good precoders.
5.2.1 Mean Variance Optimization

In the introduction, we alluded to a connection between the risk-rate problem and the investment problem in modern portfolio theory (MPT). The standard solution used in MPT is to formulate a mean variance optimization problem. Heuristically, in order for one to maximize rate subject to a fixed risk constraint, one desires to shift the center of the rate distribution to the right while constraining the variance of the distribution (so that the rate distribution does not flatten out as it is being shifted). Thus, the mean-variance optimization equivalent of problem (5.2) is,

\[
D^*_\varphi = \arg\max \mathbb{E}\{R(HD)\} \text{ such that } \text{Var}\{R(HD)\} \leq \delta \text{ and } D \in \mathcal{F}_P
\]

for some rate variance constraint \(\delta > 0\). Conversely, in order to minimize risk for a fixed rate we should minimize the variance while fixing the mean rate. Thus, the mean-variance optimization equivalent of problem (5.3) is,

\[
D^*_\varphi = \arg\min \text{Var}\{R(HD)\} \text{ such that } \mathbb{E}\{R(HD)\} \geq \upsilon \text{ and } D \in \mathcal{F}_P,
\]

for some mean rate constraint \(\upsilon > 0\). The parameters \(\upsilon\) and \(\delta\) should be chosen by the system designer. The rate constraint \(\upsilon\) can be interpreted as the minimal average rate that is required. Furthermore, \(\delta\) is a measure of maximal rate volatility that is tolerable.

5.2.2 Gaussian Mixture Models

The mean variance optimization as described is still a difficult problem for arbitrary channel distributions since in order to evaluate the mean and variance of the rate, one must still perform a \(2N^2\)-dimensional integral. We resolve this issue by assuming that the channel density is a Gaussian mixture model (GMM). The GMM family of distributions is a flexible class of probability densities that can be used as an approximation for many
non-GMM probability distributions (see [42]). The assumption of a GMM distributed channel implies that the density \( f(H|U) \) of the channel can be written,

\[
f(H|U) = \sum_{i \in I} a_i \phi(H; M_i, \sigma^2 I)
\]

where \( \phi(H; M_i, \sigma^2 I) \) denotes the density of a Gaussian random matrix with mean \( M_i \) and covariance matrix \( \sigma^2 I \). In practice, one must find the parameters \( M_i, \sigma^2 \) and \( a_i \) by fitting the GMM to the known CSI probability distribution. There are a variety of algorithms that can do this, some of the standard techniques are expectation-maximization (EM) and maximum a posteriori (MAP) [51].

Using the GMM structure, one is able to analytically evaluate the \( 2N^2 \)-dimensional integral mentioned above. Consequently, we now state simple expressions for the mean rate and mean variance that allow one to compute them with a one-dimensional integration (rather than a \( 2N^2 \)-dimensional integration). First, we give a formula for the mean rate,

**Theorem 16.** Assuming \( H \) is distributed as a GMM, the expected rate can be computed by the formula,

\[
\mathbb{E}\{R(HD)\} = \sum_{i,k} \Psi(i)(M_k) - \Pi(i)(M_k).
\]

where,

\[
\Psi(k)(M) \triangleq \lim_{\epsilon \to 0} \int_0^{+\infty} G_k(t, M) \frac{\partial^2}{\partial t^2} t^\epsilon dt,
\]

\[
\Pi(k)(M) \triangleq \lim_{\epsilon \to 0} \int_0^{+\infty} F_k(t, M) \frac{\partial^2}{\partial t^2} t^\epsilon dt.
\]

The functions \( F_k(t, M) \) and \( G_k(t, M) \) are given in the appendix. The index \( k \) is over mixture components and index \( i \) is over users.

**Proof.** See Appendix 5.5.3. \( \square \)
Note that in practice, the integral can be evaluated by picking a sufficiently small value for $\epsilon$ and numerically integrating. Next, a formula and a bound for the rate variance are given,

**Theorem 17.** Assuming $\mathbf{H}$ is distributed as a GMM, the rate variance can be computed with two dimensional integrals. Moreover, it is bounded by an expression that can be computed by one dimensional integrals,

$$
\text{Var}\{R(\mathbf{HD})\} - \mathbb{E}\{R(\mathbf{HD})\}^2 \geq \sum_{i \neq j} \Pi(i)\Pi(j) + \Psi(i)\Psi(j) - \Pi(i)\Psi(j) + \sum_i \Pi(i,2) + \Psi(i,2) - \sqrt{\Pi(i,2)\Psi(i,2)}.
$$

In the above,

\begin{align*}
\Pi(i,2) &\triangleq 
\gamma^2 - \frac{\pi^2}{6} + \lim_{\epsilon \to 0} \int_0^{+\infty} F_k(t, \mathbf{M}) \left[ \frac{\partial^3}{\partial \epsilon \partial t^2} t^\epsilon + 2\gamma \frac{\partial^2}{\partial \epsilon \partial t} t^\epsilon \right] dt, \\
\Psi(i,2) &\triangleq 
\gamma^2 - \frac{\pi^2}{6} + \lim_{\epsilon \to 0} \int_0^{+\infty} G_k(t, \mathbf{M}) \left[ \frac{\partial^3}{\partial \epsilon \partial t^2} t^\epsilon + 2\gamma \frac{\partial^2}{\partial \epsilon \partial t} t^\epsilon \right] dt.
\end{align*}

$\gamma$ is the Euler constant.

**Proof.** See Appendix 5.5.4.

Using these two expressions, one can use general nonlinear optimization routines such as an interior point optimization technique [46]. Moreover, because the objective and constraint functions are available in closed form, we can compute their gradients in closed form. These gradients can be used by the optimization algorithm to enhance the accuracy and speed of the optimization. The objective and constraint functions are rather lengthy and their gradients even more so. Therefore, we only give the computation
of the gradient of \( G_k(t, M) \) and remark that an identical calculation can be made to find the gradient of \( F_k(t, M) \). Note that the optimization problem is dependent on the precoder only through those two functions, and thus the gradients of these functions can be substituted into the expressions in Theorems 16 and 17 to obtain the gradients of the mean and variance. To compute the gradient (with respect to the precoder) of the function \( G_k(t, M) \), we introduce some auxiliary functions,

\[
F_1(D) = \det(\sigma^2 t D^T D + I)
\]

\[
F_2(D) = \text{tr}\left( (D^T D)^{-1} D^T m m^* D \right)
\]

\[
F_3(D) = \text{tr}\left( D^T m m^* D \left( \sigma^2 t D^T D D^T D + D^T D \right)^{-1} \right).
\]

The function \( G_k(t, M) \) depends on the precoder and several other parameters, we rewrite the function as \( G(D) \) for simplicity and to emphasize its dependency on the precoder. We can express the function \( G(D) \),

\[
G(D) = \frac{e^{-N_0 t}}{F_1(D)} e^{F_2(D) - F_3(D)}.
\]

Using the chain rule and product rule for derivatives, the derivative of \( G(D) \) can be expressed,

\[
\frac{dG(D)}{dD} = -\frac{G(D)}{F_1(D)} \frac{dF_1(D)}{dD} + G(D) \left[ \frac{dF_2(D)}{dD} - \frac{dF_3(D)}{dD} \right]. \tag{5.7}
\]

We can now compute the derivatives of the auxiliary functions with respect to the precoder matrix \( D \) by using the matrix differential formalism of [47] (see also [48]).

\[
\frac{d}{dX} F_1(X) = \sigma^2 t \det(\sigma^2 t X^T X + I) \left( \sigma^2 t X^T X + I \right)^{-1} X^T.
\]

\[
\frac{d}{dX} F_2(X) = 2 \left( X^T X \right)^{-1} X^T m m^* \left[ I - X \left( X^T X \right)^{-1} X^T \right].
\]
\[
\frac{d}{dX} F_3(X) = 2 (\sigma^2 t X^T X + I)^{-1} (X^T X)^{-1} \left[ X^T mm^* - (X^T X)^{-1} X^T \right] \\
- 2 \sigma^2 t (X^T X)^{-1} X^T mm^* X (\sigma^2 t X^T X + I)^{-2} X^T.
\]

The explicit derivative of \( G(D) \), as a function of \( D \), can be obtained by substituting the three matrix derivatives into equation (5.7). As mentioned previously, these derivatives can be used to aid an interior point type nonlinear optimization algorithm in the search for a locally optimal precoder.

### 5.3 Simulations

We simulate a two-user broadcast channel, where the channel is modelled as a two-component (equally likely) Gaussian mixture having mean matrices,

\[
M_1 = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}
\]

and variance \( \sigma^2 = 1/2 \). The AWGN noise has unit variance and the SNR is constrained to be at most unity. We study the performance of three precoders in the range \( 0 < \alpha < 0.01 \).

The first precoder is the optimal risk-rate (RR) precoder for \( \alpha = 0.001 \), it was found by randomly drawing millions of precoders from the set of power constrained precoders and choosing the precoder that achieves the maximum rate. Note that one is able to find this precoder only because the dimensionality of the problem is still small, a similar random search when there are three users is no longer feasible. We use the mean-variance framework to obtain the mean-variance (MV) optimal precoder where the variance was constrained to be less than 0.2. Finally, we consider the performance of a precoder that ignores all randomness in the channel and maximizes the sum rate of the mean channel \( \frac{1}{2} (M_1 + M_2) \). Figure 5.1 plots the risk versus the rate curve of all precoders, we note
that the MV precoder is close in performance to optimal RR precoder while the precoder that assumes a deterministic channel has very poor performance. We remark that other precoders that assume deterministic channels, for example a precoder that will maximize the sum rate of only one of the components, have nearly identical poor performance. In the very low risk region, the MV-optimal precoder has slightly better performance than the RR-optimal precoder since the RR precoder was designed for $\alpha = 0.01$ and thus for lower risk a different precoder will be optimal (not necessarily the MV-optimal precoder plotted above).

### 5.4 Summary

We formulated the problem of linear precoder design for broadcast channels as a risk-rate optimization tradeoff. First, we showed full powered precoders achieve all tradeoff points.
Next, we used the mean-variance optimization heuristic together with an assumption of GMM distributed channel to derive efficient formulae to evaluate the optimization problem. Simulations indicate that the MV approach can find high performing precoders in the risk-rate sense.

5.5 Appendices

5.5.1 Appendix: Theorem 14

To prove statements (1) and (4), we will show that

\[ P[H \in A_{\varphi, D}] = 0 \quad A = \{ \mathbf{H} \in \mathbb{R}^{2 \times 2} : R_D(\mathbf{H}) = \varphi \}, \]

for any \( \varphi \in \mathbb{R} \) and \( D \) invertible. By Lemma 2.7 of [44], this is sufficient to conclude statements (1) and (4). First, note that for \( \varphi < 0 \), \( A = \emptyset \) since \( R_D(\mathbf{H}) \geq 0 \) for any \( D \), and thus \( P[H \in A] = 0 \). Consider the case where \( \varphi > 0 \), we will show \( A \) is a 3-dimensional manifold and thus its measure is zero. Let the vector \( h = \text{vec}(\mathbf{H}) \), the derivative of the rate function can be written,

\[
\frac{d}{dh} R\left((D^T \otimes I) \mathbf{h}\right) = \left[\frac{d}{dh} (D^T \otimes I) \mathbf{h}\right] \left[\frac{d}{dh} R\left(\tilde{h}\right) \big|_{\tilde{h} = (D^T \otimes I)h}\right] \\
= (D^T \otimes I) \left[\frac{d}{dh} R\left(\tilde{h}\right) \big|_{\tilde{h} = (D^T \otimes I)h}\right].
\]

Since \( D \) is invertible, the only case where all derivatives are zero is when,

\[
\frac{d}{d\text{vec}(\mathbf{H})} R\left(\tilde{\mathbf{H}}\right) \big|_{\tilde{\mathbf{H}} = HD} = 0.
\]
Taking derivatives of the rate function, one may observe this occurs only when all diagonal entries of $\tilde{H}$ are zero,

$$HD = \begin{pmatrix} 0 & \ast & \ldots & \ast \\ \vdots & \vdots & \ddots & \vdots \\ \ast & \ast & \ldots & 0 \end{pmatrix} = \tilde{H}.$$ 

Note that the set of such $\tilde{H}$ is a subspace of $\mathbb{R}^{p \times p}$ of dimension $p^2 - p$. Since $\tilde{H}^T$ is linearly mapped to $H^T$ by $(D^T)^{-1}$ and $D$ is invertible, the set of matrices $H$ that satisfy the required constraint will also be a subspace of dimension $p^2 - p$. Consequently, by invoking a well known theorem [45], $A$ is shown to be an embedded submanifold of $\mathbb{R}^{p \times p}$ of dimension $p^2 - p$. This further implies the measure of $A$ is zero and thus $P[H \in A] = 0$ for $\varphi > 0$. Finally, consider the case $\varphi = 0$. Given the rate function, this occurs only when, 

$$\begin{pmatrix} D^T_1 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & D^T_p \end{pmatrix}^T vec(H^T) = \tilde{D}vec(H^T) = 0.$$ 

Thus, $A$ is the null space $N(\tilde{D})$. Since $D$ is not the zero matrix, $\tilde{D}$ will also be non-zero. Therefore, the rank of $\tilde{D}$ will be non-zero and by the rank-nullity theorem, the dimension of the null space $N(\tilde{D})$ will be strictly smaller than $p^2$. Thus, $\dim N(\tilde{D}) < p^2$ and its measure is zero.

(3) and (2), to prove continuity with respect to $\alpha$, using lemma 2.8 of [44], it is sufficient to show that $\mu \{ H : |R_D(H) - \varphi| < \epsilon \} > 0$ for every $\epsilon > 0$. Define $\mathcal{B} = \{ H : |R_D(H) - \varphi| < \epsilon \}$, we will show this set is non-empty and open. Consider the channel, 

$$\hat{H} = \sqrt{\sigma^2_e (\varphi^2/p - 1)} D^{-1},$$ 

then it is straightforward to show $R_D(\hat{H}) = \varphi$ and thus $\hat{H} \in \mathcal{B}$ for $\varphi \geq 0$. Moreover, the function $R_D$ is continuous for every $D$, this implies $\mathcal{B}$ is open as it is the preimage of the open set $\{ |R_D(H) - \varphi| < \epsilon \}$. Statement (3) is proven by noting that the Lebesgue
Chapter 5. Risk-Rate Precoder Design with Random CSI

measure is strictly positive on non-empty open sets. Continuity with respect to \( D \) is established by noting that \( R_D(H) \) is continuous on both \( D \) and \( H \) and appealing to lemma 2.13 of [44].

5.5.2 Appendix: Theorem 15

We prove this theorem in two steps: first we show that a full power precoder will achieve minimal risk (for a fixed rate) and conclude by arguing that every precoder that achieves minimal risk for some rate will also achieve maximal rate for some risk. Consider a precoding matrix \( D_1 \) with power \( \text{tr}(D_1^T D_1) = P_1 \). One can obtain another precoding matrix \( D_2 \) with a scaling \( D_2 = \sqrt{\frac{P_2}{P_1}} D_1 \) where \( P_2 > P_1 \), it is clear the precoder \( D_2 \) has power \( P_2 \). Now let \( H \) be a channel such that \( R_\omega(HD_2) \leq \varphi \). Using the definition of the function \( R_\omega \),

\[
\sum_{i=1}^N \omega_i \ln \left( 1 + \frac{r_{ii}^2}{1 + \sum_{j \neq i} \hat{r}_{ij}^2} \right) \leq \varphi,
\]

where \( r_{ij} = (HD_2)_{ij} = \sqrt{\frac{P_2}{P_1}} \hat{r}_{ij} \) and \( \hat{r}_{ij} = (HD_1)_{ij} \). This implies,

\[
\sum_{i=1}^N \omega_i \ln \left( 1 + \frac{\hat{r}_{ii}^2}{\frac{P_1}{P_2} + \sum_{j \neq i} \hat{r}_{ij}^2} \right) \leq \varphi.
\]

and \( 0 < P_1/P_2 \leq 1 \). The inequality \( \ln \left( 1 + \frac{x}{1+y} \right) \leq \ln \left( 1 + \frac{x}{\alpha+y} \right) \) for \( x > 0, y > 0, 0 < \alpha \leq 1 \) implies,

\[
R_\omega(HD_1) = \sum_{i=1}^N \omega_i \ln \left( 1 + \frac{\hat{r}_{ii}^2}{1 + \sum_{j \neq i} \hat{r}_{ij}^2} \right) \leq R_\omega(HD_2) \leq \varphi.
\]

Now, define the sets,

\[
\mathcal{H}_2 = \left\{ H \in \mathbb{C}^{p \times p} : R_\omega(HD_2) \leq \varphi \right\},
\]
\[
\mathcal{H}_1 = \left\{ H \in \mathbb{C}^{p \times p} : R_\omega(HD_1) \leq \varphi \right\}.
\]
then the inequality above implies that if $H \in \mathcal{H}_2$ then $H \in \mathcal{H}_1$, i.e. $\mathcal{H}_2 \subseteq \mathcal{H}_1$. This implies $P[H \in \mathcal{H}_2] \leq P[H \in \mathcal{H}_1]$. Thus, the risk is no larger when using a higher power precoder. In other words, it is sufficient to consider only full powered precoders when minimizing risk subject to a fixed rate constraint $\varphi$. We now proceed to show that every risk minimal precoder subject to a rate constraint is also rate maximal subject to some risk constraint. First, let the maximal rate envelope function be defined,

$$V(\alpha) = \max_{D \in \mathcal{F}_P} W_\alpha(D).$$

then $V(\alpha)$ is clearly non-decreasing with respect to $\alpha$ and $V(\alpha) \geq W_\alpha(D)$ for any $D \in \mathcal{F}_P$ and $\alpha \in [0, 1]$. Let $\varphi \in [0, +\infty)$ be a rate for which $D^*$ obtains the minimal risk $\alpha^*$, $\alpha^* = I_\varphi(D^*)$. Also, by the previous argument, we can take $D^*$ to have full power. This implies $\alpha^* \leq V^{-1}(\varphi) = \alpha$ and $V(\alpha^*) \leq V(\alpha)$. Also, $V(\alpha^*) \geq W_{\alpha^*}(D^*)$. These inequalities together imply,

$$W_{\alpha^*}(D^*) \leq V(\alpha^*) \leq V(\alpha),$$

but $W_{\alpha^*}(D^*) = V(\alpha) = \varphi$ which further implies $V(\alpha^*) = \varphi$. Thus, we observe that maximum rate $\varphi$ can be obtained for risk $\alpha^*$. Moreover, this risk-rate pair is achieved by a full power precoder $D^*$.

5.5.3 Appendix: Average Rate Formula

We start with a lemma,

**Lemma 4.**

$$\ln(\eta)^d = (-1)^d \eta \sum_{i=0}^d \binom{d}{i} \left[ \frac{d}{d\mu^{d-i}} \frac{1}{\Gamma(\mu)} \right]_{\mu=1}^{+\infty} \int_0^\infty \ln(v)^i e^{-\eta v} dv.$$
Proof. Consider the Laplace transform,

\[
\frac{1}{\Gamma(\mu)} \int_0^{+\infty} e^{-\eta v^\mu - 1} dv = \frac{1}{\eta^\mu},
\]

differentiating \(d\) times under the integral sign with respect to \(\mu\),

\[
(-1)^d \ln(\eta)^d \eta^{-\mu} = \frac{d}{d\mu^d} \frac{1}{\Gamma(\mu)} \int_0^{+\infty} e^{-\eta v^\mu - 1} dv
\]

\[
= \sum_{i=0}^d \binom{d}{i} \left( \frac{d}{d\mu^{d-i}} \frac{1}{\Gamma(\mu)} \right) \int_0^{+\infty} \ln(v)^i e^{-\eta v^\mu - 1} dv.
\]

Setting \(\mu = 1\) we obtain the theorem. \(\square\)

The rate function can be written,

\[
R(\mathbf{H} \mathbf{D}) = \sum_{i=1}^K \ln \left( 1 + \frac{\langle \mathbf{h}_i, \mathbf{d}_i \rangle^2}{N_0 + \sum_{j \neq i} \langle \mathbf{h}_i, \mathbf{d}_j \rangle^2} \right)
\]

\[
= \sum_{i=1}^K \ln (N_0 + \mathbf{v}_i^\ast \mathbf{v}_i) - \ln (N_0 + \mathbf{u}_i^\ast \mathbf{u}_i)
\]

\[
= \sum_{i=1}^K \psi(i) - \xi(i),
\]

where \(\mathbf{u}_i = \mathbf{D}^T_{(i)} \mathbf{h}_i\), \(\mathbf{v}_i = \mathbf{D}^T \mathbf{h}_i\) and \(\mathbf{D}_{(i)}\) denotes the matrix \(\mathbf{D}\) with the \(i\)th column removed. Using the identity [43],

\[
\ln(A) = -\gamma - \int_0^{+\infty} \ln(t) A e^{-At} dt.
\]  

(5.8)

we have,

\[
\mathbb{E}\{\ln(N_0 + \mathbf{u}_k^\ast \mathbf{u}_k)\} = -\gamma
\]

\[
+ \int_0^{+\infty} \ln(t) \frac{\partial}{\partial t} \mathbb{E}\{e^{-(N_0 + \mathbf{u}_k^\ast \mathbf{u}_k)t}\} dt.
\]
It is a straightforward calculation to show,

\[
\mathbb{E}\left\{ e^{-\left(N_0 + u_k^* w_k\right)t} \right\} = \frac{e^{-N_0 t}}{\det(\sigma^2 t W(k) + I)}
\]

\[
\times \text{etr}\left( \frac{1}{\sigma^2} Z(k) \left( \sigma^2 t W^2(k) + W(k) \right)^{-1} - \frac{1}{\sigma^2} W^{-1}(k) Z(k) \right)
\]

\[\triangleq F_k(t, M),\]

where

\[
W(k) = D^T(k) D(k),
\]

\[
Z(k) = D^T(k) mm^* D(k).
\]

We simplify with an integration by parts step,

\[
\int_0^{+\infty} \ln(t) \frac{\partial}{\partial t} F_k(t, M) dt = \lim_{\epsilon \to 0} \int_0^{+\infty} F_k(t, M) \frac{\partial^2}{\partial \epsilon \partial t} t^\epsilon dt.
\]

\(G_k(t, M)\) is identical to \(F_k(t, M)\) except the matrix \(D\) is used rather then \(D(k)\).

### 5.5.4 Appendix: Rate Variance Formula

The square of the rate function can be expanded,

\[
R(HD)^2 = \sum_{i,j} \psi(i) \psi(j) + \sum_{i,j} \xi(i) \xi(j) - \sum_{i,j} \psi(i) \xi(j).
\]

By independence, for terms such that \(i \neq j\), we have the following relationships,

\[
\mathbb{E}\{\xi(i)\xi(j)\} = \mathbb{E}\{\xi(i)\} \mathbb{E}\{\xi(j)\}
\]

\[= \Pi(i) \Pi(j),\]
\[ E\{\psi(i)\psi(j)\} = E\{\psi(i)\} \cdot E\{\psi(j)\} = \Psi_{(i)}\Psi_{(j)}, \]
\[ E\{\psi(i)\xi(j)\} = E\{\psi(i)\} \cdot E\{\xi(j)\} = \Psi_{(i)}\Pi_{(j)}. \]

Using the identity,
\[
\ln^2(A) = \gamma^2 - \frac{\pi^2}{6} + \int_0^{+\infty} \left( \ln^2(t) + 2\gamma \ln(t) \right) Ae^{-At} dt.
\]

and similar steps as done previously, one can derive the formulae,
\[
E\{\xi^2(i)\} - \left( \gamma^2 - \frac{\pi^2}{6} \right) = \lim_{\epsilon \to 0} \int_0^{+\infty} F_k(t, M) \left[ \frac{\partial^3}{\partial^2 \epsilon \partial t} t^\epsilon + 2\gamma \frac{\partial^2}{\partial \epsilon \partial t} t^\epsilon \right] dt.
\]
\[
E\{\psi^2(i)\} - \left( \gamma^2 - \frac{\pi^2}{6} \right) = \lim_{\epsilon \to 0} \int_0^{+\infty} G_k(t, M) \left[ \frac{\partial^3}{\partial^2 \epsilon \partial t} t^\epsilon + 2\gamma \frac{\partial^2}{\partial \epsilon \partial t} t^\epsilon \right] dt.
\]

The final term configuration that we must evaluate is \( E\{\psi(i)\xi(i)\} \), this one can be expressed as a two dimensional integral by applying identity (5.8) twice and proceeding with the integration. However, we can easily derive a bound by the Cauchy-Schwarz inequality,
\[
\left| E\{\psi(i)\xi(i)\} \right|^2 \leq E\{\psi^2(i)\} \cdot E\{\xi^2(i)\} = \Pi_{(i)}\Psi_{(i)}.2.
\]
Chapter 6

Conclusion

In this work, we examined the impact of random CSI on the operation of MISO broadcast systems. This was done by studying two fundamental and complementary questions:

- What is the impact of random CSI on the performance?
- How should the precoder be designed so as to account for random CSI?

The impact of the random CSI on system performance was studied by examining a simple model. We considered a MISO broadcast channel with a zero forcing precoder where the CSI is contaminated by additive white Gaussian noise and the channel is Rayleigh fading. We derived a striking distributional result on the overall channel between the symbols and their destination: the effective channel is distributed as a complex T-random matrix. Using this result, the ergodic sum rate and the outage probability were analyzed. We confirmed the findings of other studies in the literature that the CSI quality should be scaled with SNR in order to prevent system performance saturation. Moreover, we observe that this phenomenon of an interference-limited system holds for both slow and fast fading.

The precoder design problem was examined by considering a broadcast channel with a fixed non-fading channel where the transmitter has access to CSI. The relationship between the channel and the CSI was described by introducing a probability distribution.
This model is Bayesian as the CSI probability distribution models a fixed event rather than a repeatable experiment. We restricted ourselves to linear precoders and introduced a framework that quantifies the tradeoff between a rate and the risk associated with the precoder that may achieve it. Hence, the precoder optimization problem was formulated as maximizing the rate subject to a risk constraint or minimizing the risk subject to a rate constraint. Additional structure was introduced into the problem by assuming the CSI probability distribution is of a Gaussian mixture parametric form. This assumption permitted important analytic simplifications. By using ideas from modern portfolio theory, we developed a precoder design problem that can be solved by a gradient descent algorithm.

6.1 Future Work

There are several possible extensions to this thesis.

- **Cross Rate Precoders for an arbitrary number of users:** the cross rate precoder introduced in Chapter 2 applies only to a MISO broadcast channel with two users. It is of interest to extend the cross rate concept to multiple users. There are several possible extensions of the cross rate, for example, one could consider the sum of the cross rates of all user pairs. It is an interesting problem to formulate and motivate a definition of a generalized cross rate that can be used in an efficient manner to obtain good precoders.

- **MMSE Precoding with Noisy CSI:** A commonly used precoder is the MMSE precoder discussed in Chapter 2. The MMSE and ZF precoders have similar structure but certain key differences preclude the direct use of the results given in Chapters 3 and 4. It is of significant interest to extend the performance analysis given in this thesis to the MMSE precoder.
• **Mean Variance Precoder with an Arbitrary Kernel**: The MV precoder described in Chapter 5 used a Gaussian mixture model to approximate the CSI probability distribution. It is possible to replace the Gaussian mixture model with other mixture models. Using these mixture models should allow one to obtain better performance for some CSI probability distributions. For example, for a compactly supported CSI distribution, a mixture of uniform distributions may be more suitable.
Bibliography


