LINEAR PRECODING AND DECODING FOR MULTIPLE INPUT MULTIPLE OUTPUT (MIMO) WIRELESS CHANNELS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Hemanth Sampath

April 2001

© Copyright by Hemanth Sampath 2001

All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Arogyaswami Paulraj
(Principal Adviser)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

John M. Cioffi

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Andrea Goldsmith

Approved for the University Committee on Graduate Studies.

Abstract

Space-time coding and spatial multiplexing are prime candidates for achieving high data rates and link quality in multiple-input multiple-output (MIMO) wireless links. However, both the schemes assume no channel knowledge at the transmitter. In a number of applications, channel knowledge can be made available at the transmitter. A natural question to ask is how to use these channel estimates to further optimize the transmitter. There can be several ways to linearly or non-linearly optimize the transmitter and receiver depending on channel knowledge. In this thesis, we consider linear optimization schemes at the transmitter (precoding) and receiver (decoding) to improve performance of MIMO systems. We consider a number of designs depending on the performance criteria and degree of channel knowledge at the transmitter.

First, assuming perfect channel knowledge at the transmitter and receiver, and a flat-fading channel, we propose a generalized linear block precoding and decoding scheme based on the weighted mean square error criteria, assuming a total power constraint at the transmitter. The optimum design forces transmission only on the

eigenmodes of the MIMO channel, for any set of error weights. The power allocation on the eigenmodes depends on the error weights, which can be varied depending on the application. The weighted MMSE criteria thus provides a unified framework for designing jointly optimal linear precoders and decoders, assuming perfect channel knowledge at the transmitter.

Next, we derive optimal linear precoders and decoders for other optimization criteria such as the pairwise error probability (PEP) criteria, and constraints such as the peak power constraint. In the latter case, we find that the eigenmode transmission need not be the optimum strategy.

Next, we show how to extend the linear precoder and decoder framework to delay spread channels and multicarrier systems employing OFDM modulation. In the former case, we employ block transmission with guard symbols inserted between data blocks to prevent inter-block interference. In OFDM modulation, redundancy is added in the form of a cyclic prefix to handle delay spread. We then present a novel Finite Impulse Response (FIR) precoder structure to pre-equalize the MIMO channel with delay spread assuming a zero-forcing constraint and subject to a transmit power constraint.

Finally, we consider the design of precoders assuming partial channel knowledge at the transmitter and perfect channel knowledge at the receiver. In this context, we develop an optimum linear precoder for a space-time coded system, assuming knowledge of only the transmit antenna correlations.

Acknowledgments

I would like to thank my PhD advisor Professor Arogyaswami J. Paulraj for his research support, generosity, and sharp feedback on research problems. He inspired me to think critically and build physical intuition on research problems. I will forever be indebted to Professor Paulraj for agreeing to supervise my PhD research in February 1998. Prof. Paulraj has been a great mentor and friend.

I am grateful to Prof. Petre Stoica at Uppsala University, Sweden, for his guidance and for streamlining my mathematical thinking. His precise and in-depth feedback on research manuscripts helped me tremendously. Working with Prof. Stoica on math problems was always fun! I enjoyed several long discussions, mostly when he was a visiting Professor at Stanford and some over phone calls to Sweden at 2 AM in the morning.

I thank Dr. Rajiv Laroia and Lucent Technologies for financial support through the Fellow-Mentor-Advisor (FMA) Fellowship. I had the opportunity to work as a Member of Technical Staff at the Digital Communications Research group of Wireless

Research Laboratory under Dr. Rajiv Laroia. He inspired me to think creatively, positively and confidently during my internship at Lucent Technologies. He instilled in me the spirit of team work, dynamism and tireless enthusiasm. I also want to thank the other members of the research group at Lucent - Junyi Li, Michaela Vanderveen, Pramod Viswanath (now an Assistant Professor at the University of Illinois at Urbana Champaign), Sandeep Rangan, Haris Vikalo and Satya Uppala.

I would like to acknowledge the post-doctoral fellows at the Smart Antennas Research Group - Dr. Helmut Bolcskei (now an Assistant Professor at the University of Illinois at Urbana Champaign) and Dr. Erik Lindskog (now at Array Comm). Their sharp observations and feedback on my thesis research was valuable. I also enjoyed many social dinner and coffee break conversations with Erik and Helmut during the past two years. They have been great mentors and friends.

I want to recognize my friends and colleagues at Iospan Wireless, a wireless start-up company in San Jose, CA, where I have been working full-time since my PhD defense in July 2000. My special thanks to Jose Tellado, Rajeev Krishnamurthy, John Fan, and Shilpa Talwar for being very considerate and allowing me time to write my thesis, while working full-time.

I am grateful to Professors John Cioffi and Andrea Goldsmith for agreeing to be on my reading committee. Prof Goldsmith provided an excellent feedback on my thesis and helped improve the presentation aspects. I would also like to thank Professors

Donald Cox and Lambertus Hesselink for serving on my PhD Orals Committee. I also thank Professor Umran Inan for his generous support during my early years at Stanford, when I was pursuing an MSEE.

I thank my friends and colleagues in the Smart Antenna Research Group: Tareq Al-Nafouri, Tushar Moorti, Sumeet Sandhu, Sriram Mudulodu, Robert Heath, Dhananjay Gore, Rohit Nabar and Wonil Roh. They were a fun group to work with and learn from. They have helped at every stage of my PhD thesis.

I would always cherish my social and academic interactions with my friends in the EE department: Pawan Kapur, Atul Salvekar, Carlos Aldana, Aaron Flores, Wei Yu, Sunil Kandukuri, Ozge Koymen, Volkan Rodoplu, Lalit Nathawad, Mayur Joshi, Dinkar Singh, Arvindh Krishnaswamy, Aaron Flores, Peter Chou and many others who I may have inadvertently left out.

I thank my close friends Pawan Kapur (my big brother on campus), Debjit Mukerji (Bengal Tiger), Neha Jain, Atul Salvekar, Lalit Nathawad and Raj Kuppuswami for making my stay at Stanford and in the bay-area, a memorable one. I will always cherish their friendship, the countless dinners, movie nights and fun trips to San Francisco on friday evenings.

I will fondly remember the Bhangra dance teams from Sanskriti (a south-Asian cultural organization at Stanford) of which I was a part of, during my stay at Stanford. I also thank friends from the Stanford Ultimate Frisbee team, with whom I played for a year.

Finally, my sincerest thanks to my family for their love and encouragement. My brothers, Suman and Sachin, and my parents, Indira and Sampath have been a source of constant love and support. They have helped me in every possible way to pursue and achieve my goals. They took care of all the little things, and made sure that I had the time, space and resources to work hard with an open mind. It is to them I dedicate this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Wireless Communications

The birth of wireless communications can be traced back to the invention of the radio telegraph by Guglielmo Marconi 100 years ago. In the last 20 years, the rapid progress in radio and electronics technology has triggered a communications revolution. Wireless networks are being deployed throughout the world to meet increasing consumer demand. These include (a) cellular mobile networks (b) digital cordless telephony and (c) wireless LANS. Wireless mobile services are currently growing at about 40 % per year and these trends are likely to continue for several years.

A cellular mobile network is primarily designed to maximize bandwidth efficiency and capacity in a macro and micro cellular environment. A cellular network (see

Figure 1.1) uses cells with radii ranging from 2 to 20 kilometers. Each cell has a base-station (BTS) that services multiple users. In the forward link, the BTS transmits to the users, while in the reverse link, the users transmit to the BTS. The BTS is connected to the back-haul network through the Mobile Telephone Switching Office (MTSO). Cells are grouped in clusters. Each cluster has N cells ($N = 3$ in Figure 1.1) that are alloted distinct frequency blocks, where N is the cluster size or equivalently the *frequency reuse factor*. Reuse factor is an important measure of spectral efficiency in a wireless network.

Radio links suffer from cochannel interference (CCI) from cells that share the same frequency channel. CCI degrades system performance by lowering the signal to interference power ratio (C/I). In a wireless link, CCI depends upon the frequency reuse factor. A large reuse factor leads to low CCI. However, this means that the frequency bandwidth is not reused often, leading to an inefficient use of available bandwidth and low system capacity. Wireless networks are designed to tradeoff cochannel interference and system capacity.

There are several cellular radio systems that have been deployed on a large scale throughout the world, for fixed and mobile applications. In mobile systems, users can move from one cell to another, in which case the users must be *handed-off* from one cell to another. In fixed systems, the users are relatively stationary or geographically localized to one building. Examples of mobile cellular standards include the IS-54/136

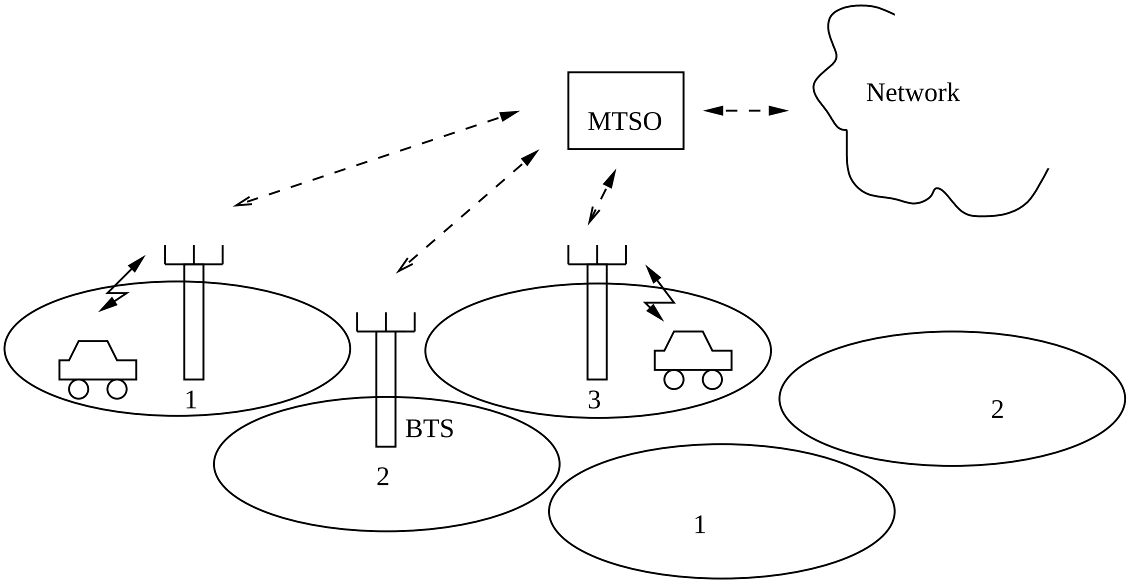

Figure 1.1: Wireless cellular network

and IS-95 systems in North America, the GSM and DCS systems in Europe and Asia, and the PDC system in Japan. Fixed wireless systems are an emerging market and there has been an active effort by several companies such as Cisco Systems, Nortel and Iospan Wireless, Inc, to develop such systems.

Digital cordless telephony is typically employed in local area/ low mobility applications; it is optimized for high quality speech with low complexities in the handset and base station. These systems allow subscribers to roam within a few tens or hundreds of feet from the base-station, which are connected to the public switched telephone network (PSTN). Examples include the 40 and 900 MHz cordless systems.

Wireless LANs provide all the functionality of wired LANs, without the physical constraints of the wire itself. Wireless LANs are easy to install, scalable and can have flexible configurations which range from simple peer-to-peer topologies to complex multiple access networks.

1.2 Space-Time Processing

Communicating over a wireless channel is highly challenging due to the complex, time-varying propagation medium. Consider a wireless link with one transmitter (Tx) and one receiver (Rx). The transmitted signal that is launched into the wireless environment arrives at the receiver along a number of distinct paths, referred to as multipaths (see Figure 1.2). These paths arise from scattering and reflection of radiated energy from objects such as buildings, hills and trees. Each of these paths has a distinct and time-varying path delay, angle of arrival, and signal amplitude. Due to constructive and destructive interference of these multipaths, the received signal can vary as a function of frequency, time and space. These variations are collectively referred to as *fading* and cause deterioration in link quality. Figure 1.3 shows the typical received signal amplitude variation across time, frequency and space. As seen from the figure, the signal amplitude can experience deep nulls leading to a highly unreliable wireless link.

Moreover, as discussed before, links suffer from *cochannel interference* (CCI) from cells that share the same frequency channel. Interference distorts the desired signal and hence leads to low system performance.

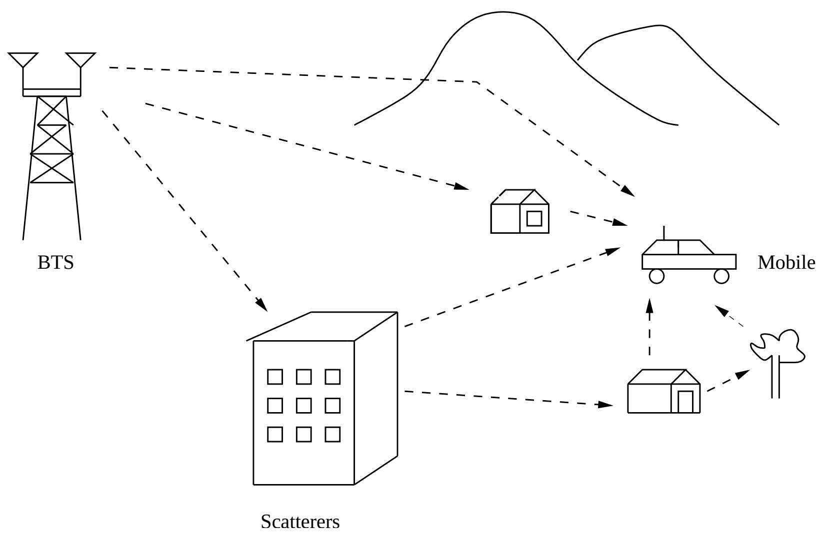

Figure 1.2: Wireless propagation environment

Wireless systems must be designed to mitigate fading and interference to ensure a reliable link.

Space-time processing is a useful tool to provide reliable communication over wireless channels. In this technology, multiple antennas and space-time modems are employed at the transmitter (Tx) and/or at the receiver (Rx) for *array gain*, *interference cancellation*, *diversity gain* and *multiplexing*. These will be explained in the following sections.

The concepts of *array gain*, *diversity gain*, and *interference reduction* have been studied for many decades. These schemes typically require channel knowledge at the receiver. In a typical setting, a known training symbol sequence is periodically transmitted to the receiver. The receiver uses the transmitted training sequence and

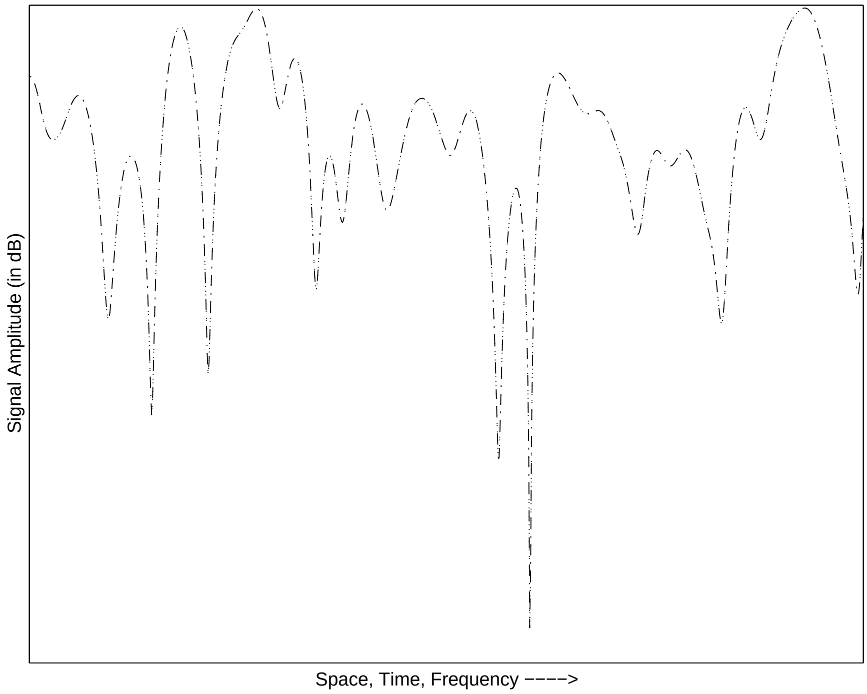

Figure 1.3: Illustration of Fading in a wireless link.

the received signal to estimate the channel. In a fading environment, the channel changes over time and hence the receiver must be trained periodically. For a detailed survey on receive space-time processing techniques, see for example [44, 82].

In the past decade, the concepts of *transmit diversity* [15, 16, 19, 20, 31, 38, 42, 39] and *spatial multiplexing* [43, 45] have attracted wide-spread attention. Both schemes do not require channel knowledge at the transmitter, but assume perfect channel knowledge at the receiver.

We now explain the above concepts in the sections to follow.

1.2.1 Array Gain

Figure 1.4 shows a mobile transmitting a signal that is received by two receive antennas. The signals (S_1 and S_2) received on the two antennas have different amplitudes and phases due to multipath propagation. These signals are appropriately combined by the receiver so that the resultant power of the output signal $\hat{S}$ is enhanced. The output signal to noise ratio (SNR) is the sum of the signal to noise ratios (SNR_1 and SNR_2) on the receive antennas, leading to an improvement in signal quality. Array gain is proportional to the number of receive antennas used.

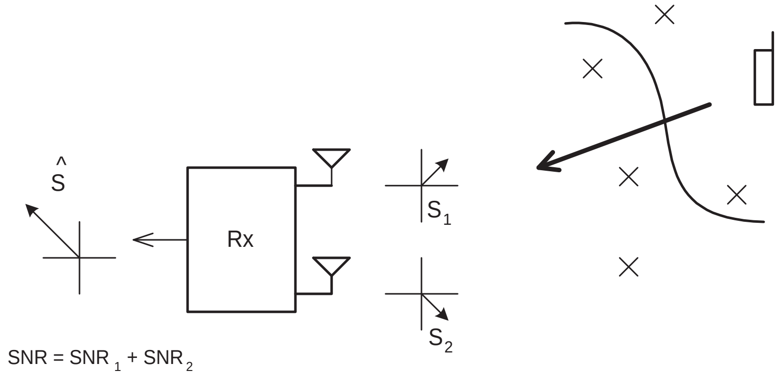

Figure 1.4: Array Gain

1.2.2 Interference Mitigation

Cochannel interference adds to the overall noise of the system and deteriorates performance. Interference reduction allows use of aggressive reuse factors and improves the system capacity.

Figure 1.5 illustrates the desired signal S and interference I from an adjacent cell, that is received by a receiver with two antennas. Typically, the desired signal (S) and interference (I) arrive at the receive antennas with different spatial signatures (S_1, S_2 and I_1, I_2), since they are coming from two different geographical locations. Interference mitigation algorithms can cancel or reduce the interference, boost the desired signal power and reduce the desired signal amplitude variability. The signal to interference ratio at the receiver output is thus enhanced, as shown in Figure 1.5.

Interference reduction can also be implemented in the transmitter, where the goal is to enhance the signal power at the intended receiver and minimize the interference

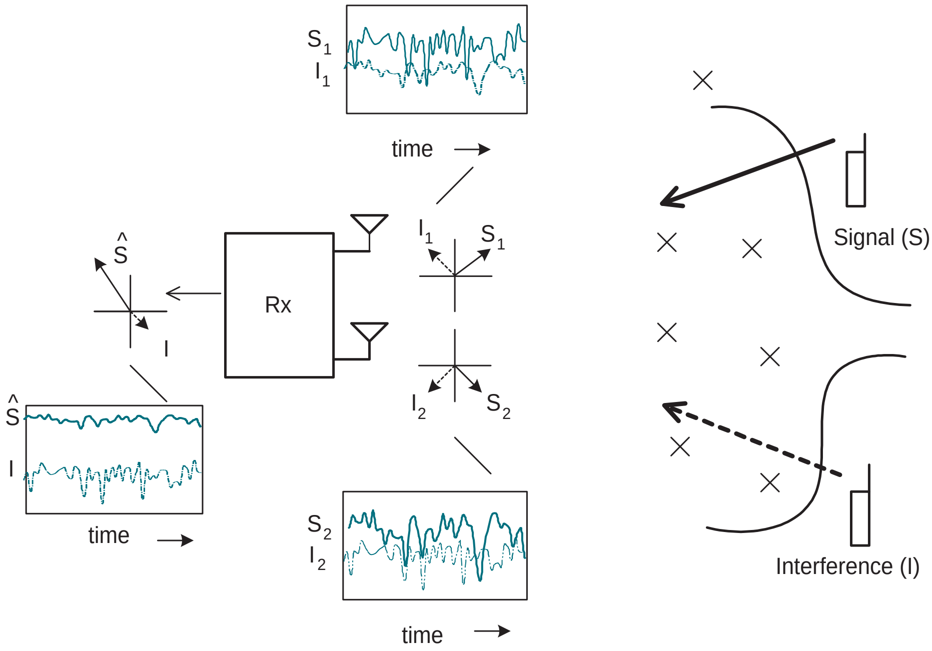

Figure 1.5: Interference Mitigation

energy sent towards cochannel users (users that communicate over the same frequency block as the intended user).

1.2.3 Receive Diversity

As explained before, multipath propagation in the wireless environment can cause the signal to fluctuate or fade as a function of time and spatial location. This causes a deterioration in link quality. Receive diversity is a well-known technique that combines (independently) fading signals on different antennas to mitigate fading and improve link quality.

Figure 1.6 illustrates receive diversity for a receiver with 2 antennas. The received

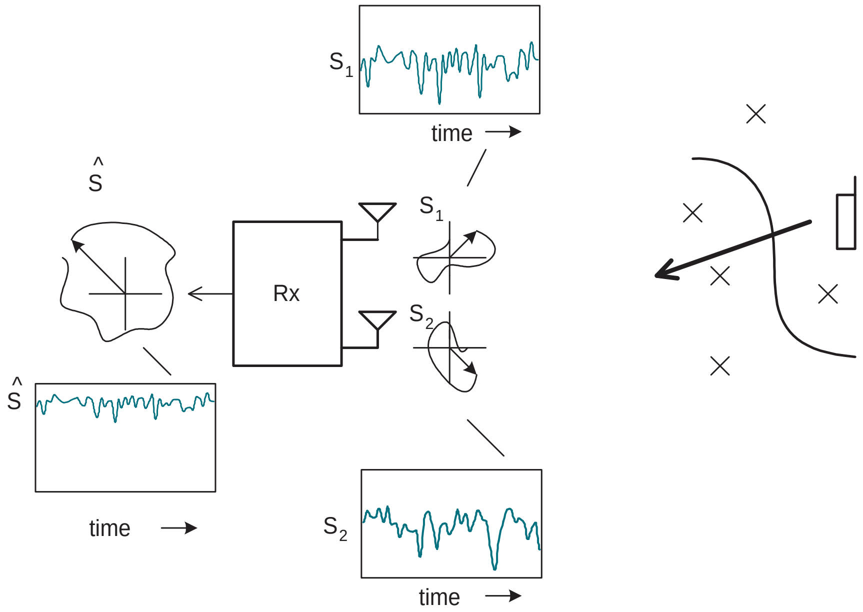

Figure 1.6: Receive Diversity

signals (S_1 and S_2) fade independently and are combined by the receiver so that the resultant signal exhibits much reduced fading. The received signal output, $\hat{S}$, has reduced signal amplitude variability when compared to S_1 and S_2. Well-known receive diversity schemes include maximum ratio combining, equal-gain combining and antenna selection.

1.2.4 Transmit Diversity

Transmit diversity was first introduced by Wittneben in 1993 and then popularized by Tarokh and Seshadri's space-time codes [15, 16]. Figure 1.7 illustrates transmit diversity for a system with 2 transmit antennas and 1 receiver antenna. At the transmitter, the original signal (S) is non-linearly or linearly mapped or encoded to generate two signals (S_1 and S_2), which are then transmitted on the two antennas. The received signals (S_{11} and S_{21}) are then processed by the receiver to reduce fading. The time waveform of the output signal, $\hat{S}$, has reduced signal amplitude variability when compared to S_{11} and S_{21}.

For a system with M_T transmit antennas and M_R receive antennas, the symbol stream is encoded by the space-time encoder to generate M_T parallel output symbol streams that are then launched into the MIMO channel simultaneously, using M_T transmit antennas. The signal is received by M_R antennas and processed at the receiver to recover the transmitted symbol stream. The transmitted symbol stream thus effectively sees $M_R \times M_T$ paths from the transmitter to the receiver, and hence reaps a maximum of $M_R M_T$-th order diversity gain.

Space-time diversity methods [15, 16, 19, 20, 31, 38, 42, 39] assume that the receiver has perfect channel knowledge and the transmitter has no channel knowledge. They are designed to perform well over averaged channel statistics, to provide diversity gain. Space-time codes can either be block codes [19, 20] or trellis codes [15]. Trellis codes designed for 2-4 transmit antennas perform well in slow-fading environments and come close to outage capacity computed by Telatar [21] and independently by Foschini and Gans [35]. The drawback of using trellis codes is the high decoding complexities at the receiver [19]. On the other hand, block codes employ a linear

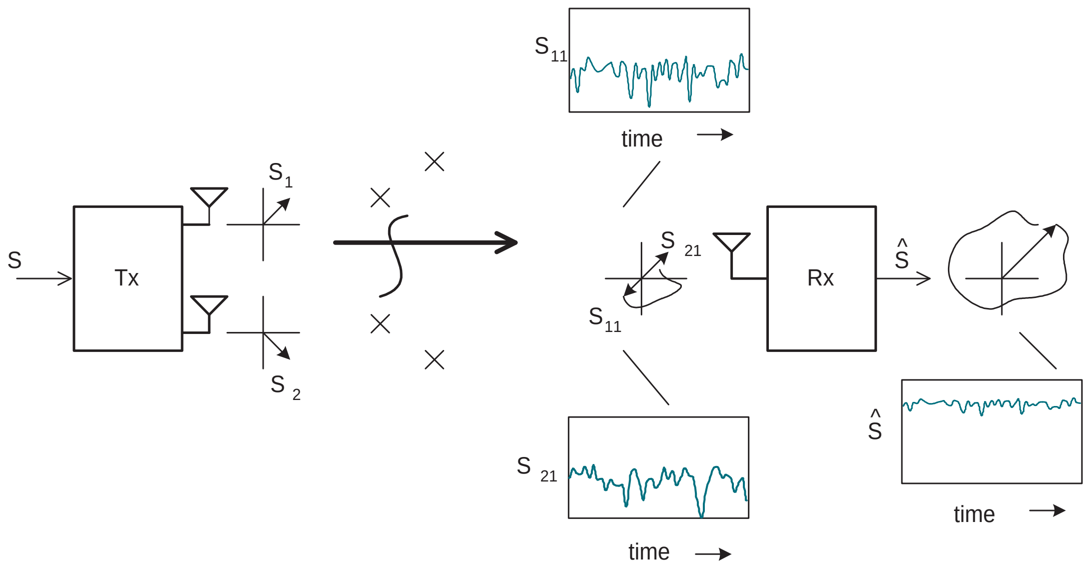

Figure 1.7: Transmit Diversity

(block) transmission scheme and enjoy a simpler decoding complexity, but suffer a loss in performance [19].

1.2.5 Spatial Multiplexing

In the past, multiple antennas have been employed only at the base-station and not at the subscriber unit, to form a SIMO (single-input multiple-output) or a MISO (multiple-input single-output) wireless link. Recently, it has been shown that multiple antennas at both ends of the wireless link can significantly increase capacity [43, 45, 35, 21]. This has motivated the research in MIMO (Multiple-input multiple-output) wireless links. MIMO algorithms and architectures are being considered in next generation fixed and mobile wireless systems.

Spatial Multiplexing is a revolutionary concept invented by Paulraj in 1994, to improve data rates over MIMO wireless links.

Figure 1.8 illustrates a 2 transmit and 2 receive antenna spatial multiplexing system. At the transmitter, the input symbol stream S is split into 2 independent lower rate substreams. These streams are modulated to form signals S_1 and S_2 , which are transmitted on separate transmit antennas. If there is a lot of multipath in the wireless channel and if the antennas are separated in space sufficiently enough to fade independently, then each transmitted symbol stream induces a different spatial signature on the receive antenna array. If the spatial signatures of these signals (denoted by (S_{11}, S_{12}) and (S_{21}, S_{22}) respectively) induced at the receiver antennas are well separated, the receiver can separate the two transmitted signals to yield the sub-streams ($\hat{S}_1$ and $\hat{S}_2$). The receiver can use linear (zero-forcing, MMSE) or non-linear (Maximum Likelihood, Successive Interference Cancelling) receiver to separate the transmitted symbol streams.

By transmitting and receiving parallel independent symbol streams in the same frequency bandwidth, spatial multiplexing obtains a linear increase in data rates, with increase in the number of antennas. We note that spatial multiplexing does not require channel knowledge at the transmitter.

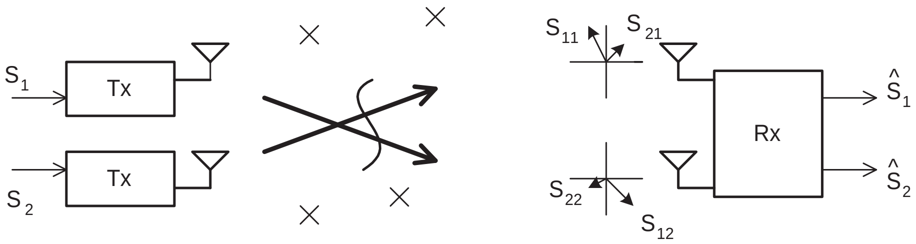

Figure 1.8: Spatial multiplexing

1.3 Thesis Overview

Space-Time coding and spatial multiplexing are prime candidates for achieving high data rates and link quality in wireless links. However, both the schemes assume no channel knowledge at the transmitter and hence are not optimized over instantaneous or fixed channel realizations.

In a number of applications, channel knowledge can be made available at the transmitter via feedback from the receiver to the transmitter (assuming the feedback delay is much less than the channel coherence time). Alternatively, in time-division duplexing (TDD) systems, where the transmission and reception proceeds in a ping-pong fashion, the channel that was estimated in the receiving mode can be used in the transmitting mode (assuming the ping-pong lag time is much less than the channel coherence time). A natural question to ask is how to use these channel estimates to optimize the transmitter. There can be several ways to linearly or non-linearly optimize the transmitter and receiver depending on channel knowledge. Linear processing techniques have smaller implementation complexity but suffer from a loss in performance when compared to non-linear processing. In this thesis, we will focus only on linear processing techniques.

1.3.1 Related Work in Literature

If the channel is fixed and known at the transmitter and receiver, jointly optimal linear processing of transmit signals (precoding) and receive signals (decoding) has been shown to improve system performance [47, 48, 56]. In this thesis, we will confine ourselves to linear processing of signals unless otherwise specified.

The following linear precoder and decoder designs have been addressed in the literature. The optimum design that *maximizes SNR* was shown to signal only on the strongest eigen-mode of the MIMO channel [78]. The optimum design that *maximize information rate* or that *minimize the sum of output symbol estimation errors* (MMSE), also decouples the channel into parallel channels and allocates the power on these sub-channels according to a water-pouring [47, 70] or inverse water-pouring policy [56, 48], respectively. Recently, the optimum precoder and decoder that maximizes SNR subject to a zero-forcing constraint on the channel has also been considered [48].

While the MMSE design minimizes the sum of symbol estimation errors across symbol streams, it does not tell us anything about the mean square error on each symbol stream. Some symbol streams may have larger errors compared to the others. In order to get a better handle on the errors on each symbol stream, we chose to design an optimum precoder and decoder that *minimize any weighted sum of symbol estimation errors*, subject to a total power constraint across all transmit antennas.[1]

The solution to the weighted MMSE design also yielded solutions to well-known designs (solved in the literature and mentioned above) for different choices of error weights. Moreover, we were able to find new applications that were not considered in the literature. A nice feature of our generalized design is that we are now able to provide a unified framework to solve the above class of linear precoder and decoder designs.

In a number of applications, channel knowledge is not perfectly available at the transmitter, either due to a slow feedback link or a fast fading channel. In such cases, one can typically feedback some statistic of the channel that is relatively stationary. When the transmitter has knowledge of only the *antenna correlations*, Shiu and Kahn [73, 74], have derived the optimum linear transmission scheme that maximizes the expected capacity of the MIMO channel. More recently, linear precoding for space-time coded systems has been considered for the case when the transmitter has knowledge of only the *angle of departures* of the channel multipaths[36, 37].

1.3.2 Thesis Contributions

Our main thesis contributions are as follows.

[1] We note that the derivation of the weighted MMSE solution involves considerable analysis to show that the optimum solution diagonalize the MIMO channels into eigen sub-channels for any set of error weights. Such an analysis is fairly straightforward in the (unweighted) MMSE design.

- First, assuming a fixed and known channel, we derive a linear block precoding and decoding scheme that minimizes any *weighted sum of output symbol estimation errors*, assuming a total transmit power constraint across all antennas. The optimum design is shown to force transmission only on the eigenmodes of the MIMO channel, for any set of error weights. The power allocation on the eigenmodes depends on the error weights, which can be varied depending on the application. The weighted MMSE criteria thus provides a unified framework for designing jointly optimal linear precoders and decoders, assuming perfect channel knowledge at the transmitter. We can cast many of the previously known work such as the SNR maximizing [78] and information rate maximizing [47] design as well as new applications into a unified framework.

- Next, we present optimal block linear precoder/decoder designs (assuming a fixed and known channel) based on other constraints and criteria, that has not yet been solved in the literature. We consider (a) Pairwise error probability (PEP) criteria assuming a total power constraint.[2] (b) Geometric mean square error criteria assuming the total power constraint and (c) Maximum information rate criteria assuming a peak power constraint. In the latter case, we find that the eigenmode transmission need not be the optimum strategy.

[2]Note that the space-time coded designs considered by [15, 19] are based on *average* PEP criterion, in which case the PEP is averaged over all channel realizations. We however optimize the transmitter assumed a fixed and known channel realization.

In formulating and solving these optimization problems, we introduce several novel mathematical techniques that can potentially be very useful in solving other problems in MIMO systems.

- Next, we show how to extend the linear precoder and decoder framework to delay spread channels and multicarrier systems employing OFDM modulation. In the former case, we employ block transmission with guard symbols inserted between data blocks to prevent inter-block interference. In OFDM modulation, redundancy is added in the form of a cyclic prefix to handle delay spread. We then present a novel Finite Impulse Response (FIR) precoder structure to pre-equalize the MIMO channel with delay spread assuming a zero-forcing constraint and subject to a transmit power constraint.

- Finally, we consider a linear precoder design assuming partial channel knowledge at the transmitter and perfect channel knowledge at the receiver. In this context, we design an optimal precoder for a space-time coded system in a flat fading channel, assuming knowledge of only the transmit antenna correlations. The precoder minimizes the average PEP subject to a total transmit power constraint.

The rest of the thesis is organized as follows. In Chapter 2, we provide an overview of the MIMO wireless channel and discuss MIMO signal models. Next, we show that existing space-time coded and spatial multiplexing systems lose capacity and performance due to lack of channel knowledge at the transmitter. This provides a motivation to research signaling schemes that make use of channel knowledge at the transmitter, which is the crux of the thesis.

In Chapters 3,4,5 and 6 we present our main thesis contributions. We conclude in Chapter 7 and present suggestions for future research.

Chapter 2

MIMO Channel and System Overview

In this chapter we will present an overview of the MIMO wireless channel model and describe parameters of interest such as fading, delay-spread, MIMO matrix eigenvalues and antenna correlation. Next, we discuss MIMO signal models for flat-fading and frequency selective channels.

Finally, we show that existing spatial multiplexing and space-time coded schemes that do not assume channel knowledge at the transmitter lose in capacity and performance. This provides a motivation to research signaling schemes that make use of channel knowledge at the transmitter, which is the crux of the thesis.

2.1 MIMO Wireless Channel

Communicating over a wireless channel is highly challenging due to the complex, time-varying propagation medium. Numerous efforts have been made to characterize the behavior of wireless channel, to help better design wireless systems. First, we will review the *scalar channel*, that captures the time and frequency selective nature of the channel, path loss and cochannel interference. Scalar channels are applicable to wireless links with one transmit and one receive antenna. Next, we will discuss the *matrix channel*, which builds upon the scalar channel model, and is applicable to wireless links with multiple antennas at the transmitter and receiver. The matrix channel model adds several new parameters.

2.1.1 The Scalar Channel

Consider a wireless link with a transmitter (Tx) and a receiver (Rx) (see Figure 2.1). The transmitted signal that is launched into the wireless environment arrives at the receiver along a number of distinct paths, referred to as multipaths. These paths arise from scattering and reflection of radiated energy from objects such as buildings, hills and trees. Each of these paths has a distinct and time-varying path delay, angle of arrival, and signal amplitude. Due to constructive and destructive interference of these multipaths, the received signal can vary as a function of frequency and time. These variations are collectively referred to as *fading* and cause deterioration in link

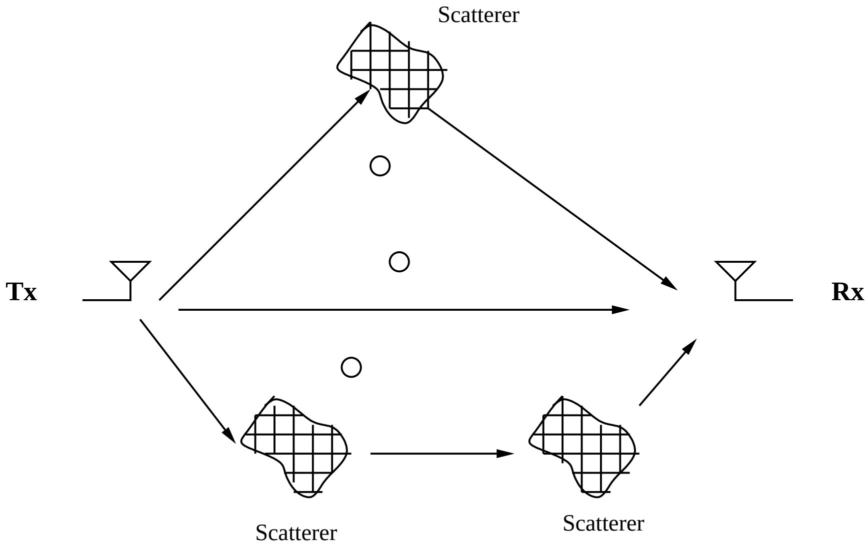

Figure 2.1: Scalar channel model

quality. We now describe fading across time and frequency, in a scalar channel.

- **Multipath and Fading**:

  Let the transmitted signal be $s(t) = s_0(t)e^{j2\pi ft}$ where $s_0(t)$ is the complex baseband signal with bandwidth B, f is the carrier frequency and t denotes time. Ignoring the receiver additive white Gaussian noise (AWGN), the corresponding received signal, $x(t)$ is the sum of all multipath components given as

$$x(t) = \sum_{i=1}^{L} \alpha_i(t)s(t-\tau_i)e^{j2\pi[f_d cos(\theta_i)t - f\tau_i(t)]} \tag{2.1}$$

where L is the number of multipaths; $\alpha_i(t)$ is the received power, θ_i is the direction of the i-th scatterer from the receiver and $\tau_i(t)$ is the delay of the i-th multipath; finally, f_d is the peak Doppler shift, which is a function of the receiver speed and carrier frequency, as will be explained later.

The multipath delay spread is $\delta_\tau = \max_i [\tau_i(t)] - \min_i [\tau_i(t)]$. If this delay spread is much less than the symbol period (T_0) of the Tx signal, i.e., $\delta_\tau \ll T_0, \;\; \tau_i \simeq \tau_0$, then equation (2.1) can be re-written as

$$x(t) \simeq s(t - \tau_0) \, [\sum_{i=1}^{L} \alpha_i(t) e^{j\phi_i(t)}] = h(t) s(t - \tau_0) \tag{2.2}$$

where $\phi_i(t) = 2\pi(f_d \, cos\theta_i \, t - f\tau_i)$ and $h(t) = \sum_{i=1}^{L} \alpha_i(t) e^{j\phi(t)}$ is the channel at time t. The channel is *frequency-flat*, i.e., is invariant across the bandwidth $B(= 1/T_0)$ of the signal. When there are a large number of multipath with similar gains and the phase is uniformly distributed between 0 and 2π, it can be shown that the received signal amplitude $|h(t)|$ is Rayleigh distributed [57].

In practice, there may also be a few line-of-sight (LOS) paths from the transmitter to the receiver. These paths do not fade as much when compared to the other diffuse multipath, and are also referred to as *dominant paths*. If there are dominant paths, $|h(t)|$ is Rician distributed [57]. An important parameter in wireless system design is the K-factor, which is the ratio between the dominant path and diffuse multi-path signal power, and is defined as:

$$K = \frac{A^2}{2\sigma_h^2} \tag{2.3}$$

where A is the amplitude of the dominant path and σ_h^2 is the multipath signal power. When $K \gg 1$, the power of the dominant LOS signal is very strong, and the channel does not suffer from fading. Channel measurements at the 2.4 GHz MMDS frequency band in the bay-area have indicated that typical values of K ranges from 0 dB at the edge of a cell, to as high as 10 dB in near line of sight conditions [64].

When the signal bandwidth (B) increases (or alternatively, when the symbol time period (T_0) decreases), $\delta_\tau \leq T_0$ is no longer true. The received signal is a sum of copies of the original signal, where each copy is delayed in time by τ_i. The channel $h(t, \tau)$ at time t to an impulse at time $t - \tau$ is now given as:

$$h(t, \tau) = \sum_{i=1}^{L} \delta(\tau - \tau_i(t))\alpha_i(t)e^{j\phi_i(t)} \tag{2.4}$$

and referred to as a delay-spread channel with L taps. In the frequency domain, the channel response fluctuates across the signal bandwidth and hence such a channel is also referred to as a *frequency selective* channel. In the absence of delay spread, the channel response is flat across the signal bandwidth and is

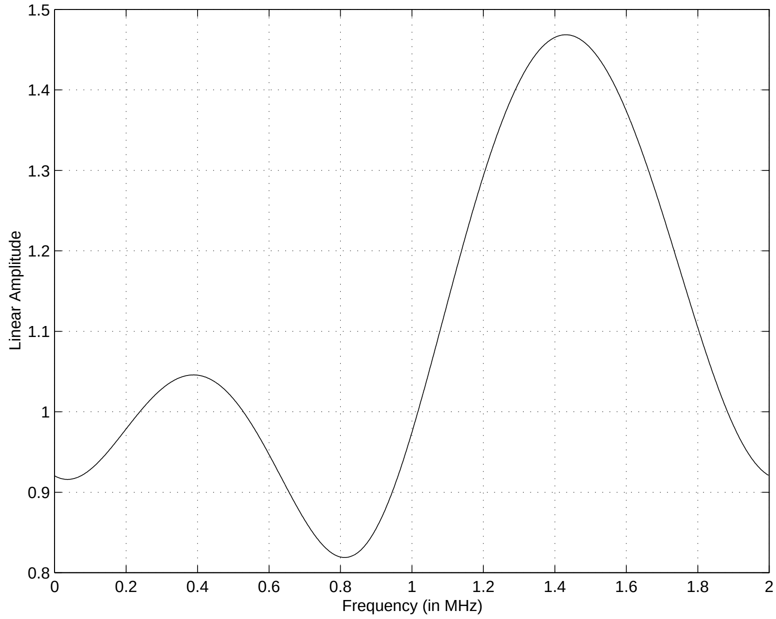

Figure 2.2: Frequency Selective Channel

referred to as a *flat channel*. *Coherence bandwidth* (C_B) is (approximately) inversely proportional to the delay spread, i.e., $C_B \simeq \alpha \frac{1}{\delta_\tau}$.

Figure 2.2 illustrates the frequency response of a delay spread channel over a 2 MHz bandwidth, obtained from a channel measurement campaign at Stanford University [64]. The measurements were made at the 2.4 GHz MMDS frequency band. The time domain channel had $L = 3$ taps, with the taps separated by 0.5 micro secs. The second and third paths had a relative gain of -5dB and -10 dB with respect to the first path.

Coherence bandwidth and frequency selectivity are important parameters that affect system performance. In multicarrier systems employing Orthogonal Frequency Division Multiplexing (OFDM) modulation, the total bandwidth is divided into smaller frequency bins over which signals are transmitted. A frequency selective channel with low coherence bandwidth implies that difference frequency bins experience different fades. Techniques such as coding and interleaving across these frequency bins can be employed to get frequency diversity and improve system performance. In single carrier systems, frequency selectivity introduces inter-symbol interference (ISI). This causes distortion of the received signal and lowers system performance. Time domain equalization of the channel taps is typically employed to mitigate inter-symbol interference and get multipath diversity.

- **Doppler Spread**:

  The wireless channel is usually time-varying because either the user or the scatterers are in motion. Motion introduces frequency shift and consequently generates carrier frequency spreading, termed *Doppler Spread*. The peak Doppler shift, f_D, is defined as: $f_D = 2\pi \frac{v}{\lambda}$, where v is the effective velocity (m/sec), and λ is the wavelength (m).

  In mobile wireless systems, when a moving Tx with velocity v transmits a continuous-wave (CW) signal with carrier frequency f_c, the Doppler spectrum at the stationary Rx is given as [57]:

$$S(f) = \frac{2\sigma_h^2}{\pi\sqrt{f_D^2 - (f - f_c)^2}}, \qquad |f - f_c| \leq f_D \tag{2.5}$$

where $2\sigma_h^2$ is the average signal power. Figure 2.3 plots the Doppler (Jakes) spectrum for a peak Doppler spread of 50 Hz. The *Coherence time* (C_t) of a channel is (approximately) inversely proportional to the peak Doppler shift, i.e., $C_t \simeq \frac{1}{f_D}$.

Coherence time is a measure of how fast the channel is changing across time. A small coherence time implies that the channel is changing rapidly across time. It is an important parameter that affects the design and performance of wireless system, as explained below.

Most wireless systems require an accurate estimate of the channel at the receiver to decode the transmitted information. To facilitate channel estimation, the transmitter sends training information to the receiver. If the coherence time is low, the transmitter must re-train the receiver periodically. For example, with a 50 Hz peak Doppler shift, the coherence time is 20 milli seconds and the transmitter must re-train after about 2 milli seconds.

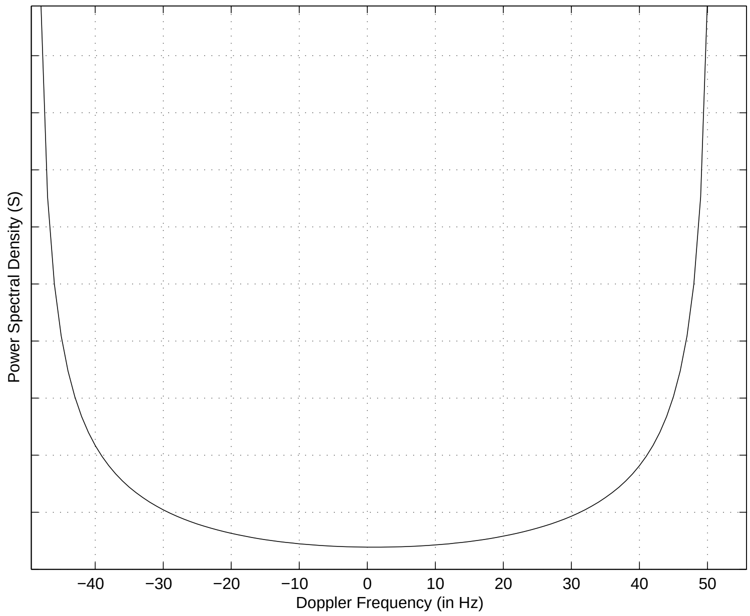

Figure 2.3: Doppler (Jakes) spectrum in mobile environment

2.1.2 The Matrix Channel

The matrix channel model adds several new parameters to the scalar channel model.

Consider a MIMO system with M_T transmit antennas and M_R receive antennas (see Figure 2.4). Let there be L scatterers in the environment. The M_T transmitted signals are reflected from these scatterers and arrive at the M_R receive antennas as multi-paths. Each multi-path has an angle of departure $\theta_{T,i}$ w.r.t the transmitter, angle of arrival $\theta_{R,i}$ w.r.t the receiver, path gain $\alpha_i(t)$, and delay τ_i.

We will first develop the MIMO channel model assuming no delay spread, i.e., assuming $\delta_\tau \ll T_0$. The signal generated by the ith scatterer will induce an array response vector $\mathbf{a}_{T,i}(t)$ at the M_T transmit antennas and $\mathbf{a}_{R,i}(t)$ at the M_R receive

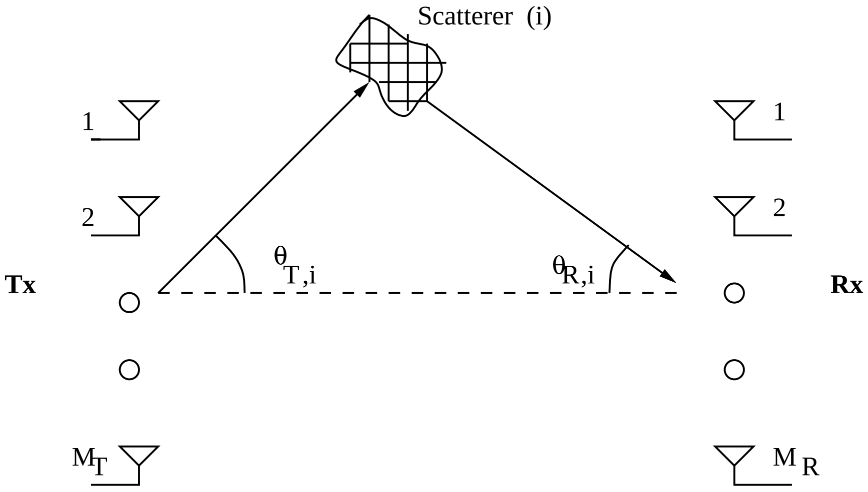

Figure 2.4: MIMO channel model

antennas.

The $M_T \times 1$ vector channel $\mathbf{h}_{T,i}(t)$ from the ith scatterer to the M_T transmit antennas is given by

$$\mathbf{h}_{T,i}(t) = \mathbf{a}_{T,i}(t)\, \alpha_{T,i}(t)\, e^{j\phi_{T,i}(t)} \tag{2.6}$$

where $\phi_{T,i}(t)$ is the phase shift and $\alpha_{T,i}(t)$ is the path loss. Likewise, the $M_R \times 1$ vector channel $\mathbf{h}_{R,i}(t)$ from the ith scatterer to the M_R receive antennas is given by

$$\mathbf{h}_{R,i}(t) = \mathbf{a}_{R,i}(t)\alpha_{R,i}(t)e^{j\phi_{R,i}(t)}, \tag{2.7}$$

where $\phi_{R,i}(t)$ is the phase shift and $\alpha_{R,i}(t)$ is the path loss.

Hence, the channel constructed by the ith scatter, between the M_T transmit and

M_R receive antennas, is given as:

$$\mathcal{H}_i(t) = \mathbf{h}_{R,i}(t)\mathbf{h}^*{}_{T,i}(t) \tag{2.8}$$

Using (2.6) and (2.7) in (2.8), and defining $\alpha_i(t) := \alpha_{R,i}(t)\alpha_{T,i}(t)$ and $\phi_i(t) := \phi_{R,i}(t) - \phi_{T,i}(t)$, we have:

$$\mathcal{H}_i(t) = \mathbf{a}_{R,i}(t)\mathbf{a}^*_{T,i}(t)\alpha_i(t)e^{j\phi_i(t)} \tag{2.9}$$

Adding up the contributions due to all L scatterers, we get the MIMO channel as:

$$\mathbf{H}(t) = \sum_{i=1}^{L} \mathcal{H}_i(t) = \sum_{i=1}^{L} \mathbf{a}_{R,i}(t)\mathbf{a}^*_{T,i}(t)\alpha_i(t)e^{j\phi_i(t)} \tag{2.10}$$

When the MIMO channel matrix has contributions from a lot of multipaths (large L), the matrix channel elements will tend to be Gaussian random variables from the Central Limit Theorem. Such a channel is referred to as an *iid matrix channel* or a Rayleigh flat-fading matrix channel with elements having a mean 0 and variance σ_h^2. Such a channel is typically common in urban areas with lots of scatterers like buildings, trees etc [64].

In practice, there may also be a few line-of-sight (LOS) paths from the transmitter to the receiver. These paths do not fade as much when compared to the other diffuse multipath, and are also referred to as *dominant paths*. In such cases, the channel matrix is not iid. As explained before, the K-factor indicates the ratio between the dominant path and diffuse multi-path signal power. Typical values of the K-factor ranges from 0 dB at the edge of a cell, to as high as 10 dB in near line of sight conditions [64].

When the signal bandwidth increases so that $\delta_\tau \simeq T_0$, then the delay spread is no longer negligible. In this case, we can index the channel contribution from each scatterer by a delay τ_i, where i is the scatterer index. The MIMO channel response at time t to an impulse at time $t-\tau$ is given as:

$$\mathbf{H}(t,\tau) = \sum_{i=1}^{L} \mathbf{a}_{R,i}(t)\mathbf{a}_{T,i}^{*}(t)\alpha_i(t)e^{j\phi_i(t)}\delta(\tau-\tau_i(t)) \tag{2.11}$$

The MIMO channel frequency response is given as:

$$\tilde{\mathbf{H}}(f,t) = \int_{\tau} \mathbf{H}(t,\tau)e^{j2\pi f\tau}d\tau \tag{2.12}$$

We now investigate some key parameters of a MIMO channel

- **Eigenvalues, rank and condition number**

  Perhaps the single most important parameter in a MIMO channel is the eigenvalue distribution of the channel matrix.

The rank of a matrix equals the number of non-zero eigenvalues of a matrix. It is hence a measure of the number of available spatial subchannels. (In this thesis, we will hence use the term *subchannel* and *eigenmode* interchangeably). The number of spatial subchannels indicates the number of parallel symbol streams that can be transmitted through the MIMO channel, using the same frequency bandwidth, and is hence a measure of the capacity of the MIMO channel. This fact is exploited by spatial multiplexing systems that send independent data streams on multiple transmit antennas to improve data rates.

The spread of the channel matrix eigenvalues is a measure of the orthogonality of the MIMO channel. A large eigenvalue spread means that the channel matrix is highly non-orthogonal and vice-versa. Orthogonal channels are desirable since they have no null-spaces and hence do not lose transmitted information. Moreover, orthogonal channels can be inverted in the receiver without noise amplification, leading to good system performance.

The condition number captures the spread of the eigenvalues of a matrix, as explained below. Consider the $M_R \times M_T$ MIMO channel frequency response $\tilde{\mathbf{H}}(f)$ (where we have dropped the time index t for simplicity), and let $M = min(M_R, M_T)$. The singular value decomposition (SVD) of $\tilde{\mathbf{H}}(f)$ is given as:

$$\tilde{\mathbf{H}}(f) = \mathbf{U}(f)\mathbf{\Lambda}(f)\mathbf{V}^*(f) \tag{2.13}$$

where $\mathbf{U}(f)$ is the $M_R \times M$ orthogonal left-singular matrix, $\mathbf{V}^*(f)$ is the $M \times M_T$ orthogonal right-singular matrix and $\mathbf{\Lambda}(f)$ is the $M \times M$ diagonal matrix of singular values, $\lambda_i(f)$, $i = 1, 2.., M$. We will assume $\lambda_i(f)$ are arranged in decreasing order of magnitude from top-left to bottom-right, i.e., $\lambda_1(f) \geq \lambda_2(f) \geq ... \geq \lambda_M(f)$. The *condition number* of the matrix is defined as:

$$\kappa[\tilde{\mathbf{H}}(f)] = \frac{\lambda_1(f)}{\lambda_M(f)} \tag{2.14}$$

A condition number of unity implies that the channel matrix is orthogonal. A large condition number means that the channel matrix is highly non-orthogonal or *ill-conditioned*.

- **Antenna Correlations** - In many applications, the transmit and/or the receive antennas can be correlated. For example, in cellular systems, the base-station antennas are typically unobstructed and have no local scatterers. This induces correlation across the base-station antennas, as a result of which the MIMO channel matrix entries do not fade independently.

  Antenna correlations tell us about the available spatial diversity in a MIMO channel. If the antennas are highly correlated, very little spatial diversity gain can be extracted from the channel, and vice-versa if the antennas are uncorrelated.

Consider an $M_R \times M_T$ flat-fading MIMO channel matrix $\mathbf{H}(t)$. Denote the $M_T \times M_T$ transmit antenna correlation matrix as

$$\mathbf{R}_{a,T} = \begin{pmatrix} 1 & \rho_{1,1} & \cdots & \rho_{1,M_T} \\ \rho_{1,1} & 1 & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \rho_{M_T,1} & \cdots & \cdots & 1 \end{pmatrix}_{M_T \times M_T} \tag{2.15}$$

where $\rho_{i,j}$ is the correlation coefficient between the ith and jth transmit antennas. The $M_R \times M_R$ receive antenna correlation matrix $\mathbf{R}_{a,R}$ can be defined similarly.

If the receiver antennas are correlated and the transmit antennas are uncorrelated, then only the rows of $\mathbf{H}(t)$ are correlated with respect to each other. The channel correlation matrix is then given as $cov\,[vec\,(\mathbf{H}(t)\,)\,] = E(\mathbf{R}_{a,R} \otimes \mathbf{I}_{M_R})$ where $\otimes$ is the matrix Kronecker product[1] [74]. Similarly, if only the transmit antennas are correlated, we have $cov[vec(\mathbf{H}(t))] = E(\mathbf{I}_{M_T} \otimes \mathbf{R}_{a,T})$. If both transmit and receive antennas are correlated, we have $cov[vec(\mathbf{H}(t))] =$

[1] For an $m \times n$ matrix $\mathbf{A}$ and $p \times q$ matrix $\mathbf{B}$, we define the Kronecker product as

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix} \tag{2.16}$$

$E(\mathbf{R}_{a,R} \otimes \mathbf{R}_{a,T})$, in which case the statistical properties of $\mathbf{H}(t)$ are identical to those of the product matrix $\mathbf{R}_{a,R}^{1/2}\mathbf{H}_w(t)\mathbf{R}_{a,T}^{1/2}$ where $\mathbf{H_w}(t)$ is an iid matrix channel.

We can hence write down the following channel model developed in [74] to include antenna correlations:

$$\mathbf{H}(t) = \mathbf{R}_{a,T}^{1/2}\mathbf{H}_w(t)\mathbf{R}_{a,R}^{1/2} \tag{2.17}$$

Note that if $\mathbf{R}_{a,T} = \mathbf{R}_{a,R} = \mathbf{I}$, then the channel model reduces to the iid matrix channel. Note that the eigenvalues of the $\mathbf{H}(t)$ are dictated by the eigenvalues of the antenna correlation matrices. Hence, if the antenna correlations are high, then the correlation matrix and hence $\mathbf{H}(t)$ is ill-conditioned. As explained before, ill-conditioned matrices have null-spaces or eigenmodes with low gains and lead to a deterioration in system performance.

2.2 Signal Model

Consider a MIMO system with M_T transmitter antennas and M_R receiver antennas as shown in Figure 2.5. At the transmitter the input bit stream is coded/modulated to generate M_T symbol streams that are launched into the MIMO channel using M_T transmit antennas. The signal is then received by M_R receive antennas and decoded/demodulated to recover the bit stream.

In this section, we will develop a signal model for (a) single-carrier systems with flat-fading and frequency selective fading channels and (b) multi-carrier systems that employ OFDM (Orthogonal Frequency Division Multiplexing) modulation, and show that they have the same form.

2.2.1 Signal Model: Single Carrier Modulation

Flat Fading Channel:

If the symbol time-period (T_0) is much larger than the channel delay-spread (δ_τ), the channel is essentially flat-fading and the system equation is given as

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{2.18}$$

where $\mathbf{H}$ is an $(M_R \times M_T)$ matrix whose (i, j)-th entry denotes the gain from transmit antenna j to receive antenna i, $\mathbf{x}$ is the $M_R \times 1$ received symbol vector and $\mathbf{s}$ is the $M_T \times 1$ transmitted symbol vector, and finally $\mathbf{n}$ is the $M_R \times 1$ noise vector (at the given symbol time). Note that that we have dropped the time index t for simplicity.

Frequency Selective Channel:

In some applications, the delay spread (δ_τ) is much larger or comparable to the symbol time period (T_0). In such a case, the channel is frequency selective.

For a MIMO channel with finite delay-spread that spans L symbol periods, we

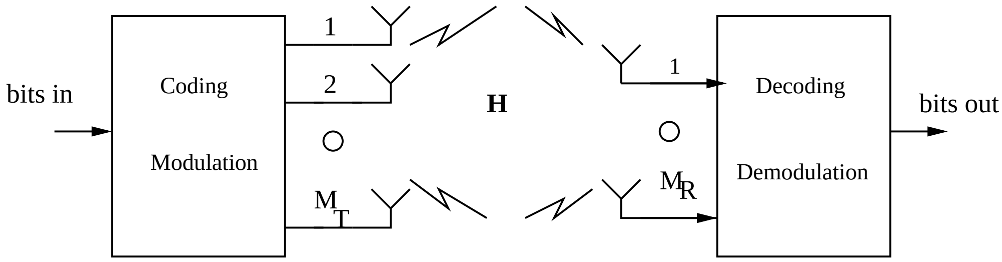

Figure 2.5: MIMO communication system

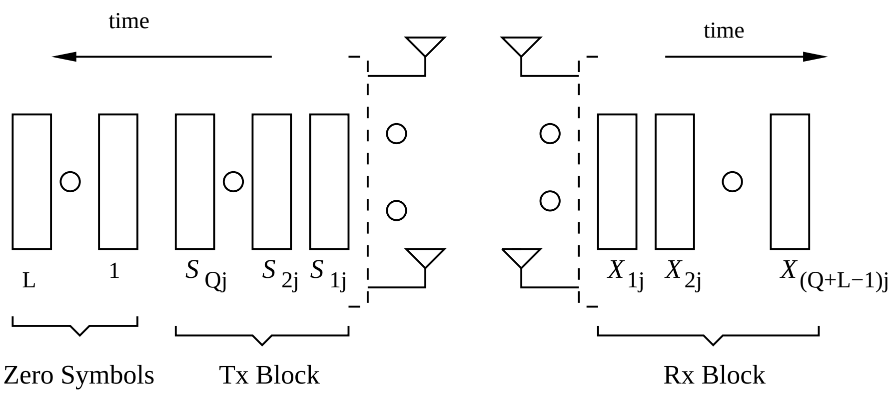

Figure 2.6: System model: Delay spread channel

can write the system equation for a finite number of vector symbols in block form.

Consider a data *block* with Q input symbol vectors (in time succession). These symbol vectors are denoted as $\mathcal{S}_{ij}$ where i is the symbol time index within a block and j is the block index. The Q symbol vectors are launched into the MIMO channel in succession, at the end of which $L-1$ zero guard symbol vectors are transmitted to prevent inter-block interference. At the receiver, the $Q+L-1$ received symbol vectors, $\mathcal{X}_{ij}$, $(i = 1, 2, ..., Q+L-1)$, corresponding to the block j are processed to recover the transmitted bit stream.

Mathematically, we can represent the above process as follows. We stack Q input symbol vectors, $\mathcal{S}_{1j}, \mathcal{S}_{2j}, ..., \mathcal{S}_{Qj}$ in a $(Q\, M_T \times 1)$ vector $\mathcal{S}_j = vec([\mathcal{S}_{1j} \ldots \mathcal{S}_{jQ}])$, and the $Q+L-1$ output symbol vectors, $\mathcal{X}_{1j}, \mathcal{X}_{2j}, ..., \mathcal{X}_{(Q+L-1)j}$ in a $[(Q+L-1)M_R \times 1]$ vector, $\mathcal{X}_j = vec([\mathcal{X}_{1j} \ldots \mathcal{X}_{j(Q+L-1)}])$.

The system output for the j-th data block can be written as (cf Fig. 2.6):

$$\mathcal{X}_j = \mathcal{H}\mathcal{S}_j + \mathcal{N}_j \tag{2.19}$$

where $\mathcal{H}$ is given by:

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \mathcal{H}_L & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \mathcal{H}_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathcal{H}_L \end{pmatrix}_{(Q+L-1)M_R \times QM_T} \tag{2.20}$$

and $\mathcal{H}_j$, $1 \leq j \leq L$, denotes the j-th tap of the MIMO channel impulse response matrix and where we have assumed that the channel is stationary over the input block length.

Making the change of variables: $\mathcal{X}_j \rightarrow \mathbf{x}$, $\mathcal{S}_j \rightarrow \mathbf{s}$, $\mathcal{N}_j \rightarrow \mathbf{n}$ and $\mathcal{H} \rightarrow \mathbf{H}$, we get

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{2.21}$$

which has the same form as (2.18).

The optimal choice of Q is a system design parameter. A large value of Q is more efficient, since the redundancy $(L-1)/Q$ is low. The drawback is a large receiver complexity, since the receiver now has to process a large number of samples to recover the transmitted stream.

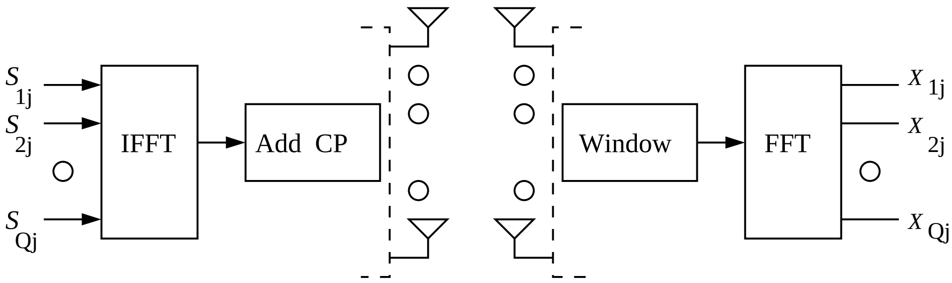

Figure 2.7: System model: OFDM modulation

2.2.2 Signal Model: Multicarrier Modulation

Recently, OFDM (Orthogonal Frequency Division Multiplexing) or DMMT (Discrete Matrix Multi Tone) modulation has received considerable attention for MIMO transmission systems to combat delay spread [49, 70, 71].

We now develop the signal model for an OFDM system (see Figure 2.7). The total frequency bandwidth is divided into Q frequency tones. A data block j consists of Q symbol vectors, $\mathcal{S}_{ij}$, $i = 1, ...Q$. Over each tone i, symbol vector $\mathcal{S}_{ij}$ is transmitted. The time waveform is generated by taking the Q point IFFT of the data block. A cyclic prefix (CP) of length L symbol vectors is then added to the beginning of the time waveform and launched into the MIMO channel.

At the receiver, the first and last L symbol vectors in the received time waveform are ignored. A Q point FFT is taken to get the received data block in the frequency domain. The received data block j consists of Q symbol vectors, $\mathcal{X}_{ij}$, $i = 1, .., Q$.

Mathematically, the system equation becomes:

$$\mathcal{X}_j = \mathcal{G}\mathcal{H}\mathcal{F}\mathcal{S}_j + \mathcal{G}\mathcal{N}_j \tag{2.22}$$

where $\mathcal{S}_j = vec([\mathcal{S}_{1j} \dots \mathcal{S}_{jQ}])$, $\mathcal{X}_j = vec([\mathcal{X}_{1j} \dots \mathcal{X}_{jQ}])$, $\mathcal{F}$ and $\mathcal{G}$ are the transmit IFFT and receive FFT operations given as:

$$\mathcal{F} = (\bar{\mathbf{\Theta}}_{\mathbf{Q}} \otimes \mathbf{I}_{M_T \times M_T}) \tag{2.23}$$

$$\mathcal{G} = (\mathbf{\Theta}^*_{\mathbf{Q}} \otimes \mathbf{I}_{M_R \times M_R}) \tag{2.24}$$

and

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_L & \cdots & \mathcal{H}_1 & 0 & \cdots & 0 \\ 0 & \mathcal{H}_L & \cdots & \mathcal{H}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots 0 & \mathcal{H}_L & \cdots & \cdots & \mathcal{H}_1 \end{pmatrix}_{QM_R \times (Q+L)M_T} \tag{2.25}$$

In the expression for $\mathcal{F}$ and $\mathcal{G}$, $\bar{\mathbf{\Theta}}_{\mathbf{Q}}$ is the $(Q+L) \times Q$ matrix obtained by taking a $Q \times Q$ IFFT matrix $\mathbf{\Theta}_{\mathbf{Q}}$ and adding the last L rows on top, creating the so-called cyclic prefix.

Using (2.23), (2.24) in (2.22) and defining $\tilde{\mathbf{H}} := (\mathbf{\Theta}^*_{\mathbf{Q}} \otimes \mathbf{I}_{M_R \times M_R})\mathcal{H}(\bar{\mathbf{\Theta}}_{\mathbf{Q}} \otimes \mathbf{I}_{M_T \times M_T}) = diag(\tilde{\mathbf{H}}_1, ..., \tilde{\mathbf{H}}_Q)$ to be a $QM_R \times QM_T$ block diagonal matrix made from the MIMO transfer function values at the frequency bins $(1, 2, ..., Q)$, we can write:

$$\mathcal{X}_j = \tilde{\mathbf{H}}\mathcal{S}_j + \tilde{\mathbf{N}}_j$$

Making the change of variables, $\mathcal{X}_j \rightarrow \mathbf{x}$, $\mathcal{S}_j \rightarrow \mathbf{s}$, $\mathcal{N}_j \rightarrow \mathbf{n}$, and $\tilde{\mathbf{H}} \rightarrow \mathbf{H}$, we get:

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{2.26}$$

making it identical to equation (2.18).

The optimum choice of Q is a system design parameter. The number of frequency bins (Q) should be large enough to accurately sample the channel response in the frequency domain. A small Q leads to undersampling the frequency channel response, leading to deterioration in system performance. On the other hand, a large Q increases system complexity due to the size of the FFT and IFFT. Furthermore, a large Q causes each frequency bin to span a very small bandwidth. This makes the carriers on frequency bins more susceptible to local oscillator (LO) drifts, RF phase noise, Doppler spread and inter- carrier distortion.

2.2.3 Generalized Signal Model

In the previous two sub-sections, we observed that the generic MIMO system equation (for single-carrier and OFDM case) can be written in the form

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{2.27}$$

In both the cases, only the dimensionality of the system variables (i.e. $\mathbf{x}$, $\mathbf{s}$ and $\mathbf{n}$) as well as the structure of $\mathbf{H}$ will change as discussed in the previous section.

2.3 Channel Knowledge at the Transmitter

Existing space-time coded and spatial multiplexing schemes assume no channel knowledge at the transmitter and are hence not optimized over instantaneous channel realizations.

They assume an iid matrix channel for good system performance. As explained before, such a channel model assumes that the MIMO channel matrix has contributions from a number of multipaths, each with near-equal power. In practice, line-of-sight (LOS) paths from the transmitter to the receiver, and the transmit and receive antenna correlations, will result in a MIMO channel matrix that is far from iid.

In this section, we will show that spatial multiplexing and space-time coded systems lose capacity and performance in such channels, since they are not optimized over channel realizations. This motivates the design of MIMO transmission schemes that assume channel knowledge at the transmitter, which is the main contribution of the thesis.

2.3.1 Capacity

In this section, we review a well-known result on MIMO capacity for *flat-fading* channels and infer the following:

- When the channel matrix is not orthogonal, lack of channel knowledge at the transmitter, leads to a loss in channel capacity. Existing space-time coded schemes assume no channel knowledge and are hence sub-optimal from the point of view of achieve Shannon capacity, when channel knowledge is available at the transmitter. This provides a motivation for researching MIMO transmission schemes that make use of the channel knowledge at the transmitter.

The information capacity of the MIMO channel is defined as the maximum of the mutual information between the input and output over all the distributions on the input that satisfy the transmit power constraint.

Definition: The *information capacity* of a MIMO system with a fixed channel $\mathbf{H}$ that is perfectly known at the receiver and a total transmit power constraint, p_0 can be computed as [70]:

$$C = \max_{p(\mathbf{s})} I(\mathbf{x};\mathbf{s}) \tag{2.28}$$

$$subject\ \ to: \quad Tr(E[\mathbf{s}\mathbf{s}^*]) = p_0$$

where $I(\mathbf{x};\mathbf{s})$ is the mutual information between $\mathbf{s}$ and $\mathbf{x}$, and given as

$$I(\mathbf{x};\mathbf{s}) = log_2\, det(\, \mathbf{R_{ss}}^{*1/2}(\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H})\mathbf{R_{ss}}^{1/2} + \mathbf{I}) \tag{2.29}$$

where we have assumed that

$$E(\mathbf{ss}^*) = \mathbf{R_{ss}};\ \ E(\mathbf{nn}^*) = \mathbf{R_{nn}};\ \ E(\mathbf{sn}^*) = 0. \tag{2.30}$$

with the superscript * denoting the conjugate transpose.

Denoting $B = rank(\mathbf{H})$, we can define the eigenvalue decomposition (EVD):

$$\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H} = (\mathbf{V}\ \ \tilde{\mathbf{V}}) \begin{pmatrix} \mathbf{\Lambda} & 0 \\ 0 & \tilde{\mathbf{\Lambda}} \end{pmatrix} (\mathbf{V}\ \ \tilde{\mathbf{V}})^* \tag{2.31}$$

where $\mathbf{V}$ is an $M_T \times B$ orthogonal matrix which forms a basis for the range space of $\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the B nonzero eigenvalues arranged in a decreasing order from top-left to bottom-right. In addition, $\tilde{\mathbf{\Lambda}}$ contains the zero eigen-values, and $\tilde{\mathbf{V}}$ is a $M_T \times (M_T - B)$ orthogonal matrix which constitutes a basis for the null-space of $\mathbf{H}^* \mathbf{R_{nn}}^{-1} \mathbf{H}$.

We have the following two scenarios:

- *Perfect channel knowledge at the transmitter* :

  The mutual information in (2.29) is maximized when [47, 70]:

$$\mathbf{R_{ss}} = \mathbf{V} \mathbf{\Lambda_s} \mathbf{V}^*$$
$$\Lambda_s = \left(\mu \mathbf{I} - \mathbf{\Lambda}^{-1}\right)_+ \qquad (2.32)$$

  where the $(.)_+$ means that the matrix Λ_s has to be positive semi-definite, i.e., the diagonal entries of Λ_s have to be greater than or equal to zero; and μ is a constant whose value is computed from the transmitter power constraint, $Tr(\mathbf{R_{ss}}) = Tr(\mathbf{\Lambda_s}) = p_0$. The resultant Shannon capacity is obtained by substituting the optimal $\mathbf{R_{ss}}$ in (2.29).

  The optimum transmission scheme pours power on the eigenmodes of the MIMO channel according to the *waterpouring* strategy given in equation (2.32). According to this solution, the transmitter will pour more power on stronger eigenmodes and less or no power on the weaker eigenmodes. By selectively weighting the eigenmodes, the transmitter avoids the nullspaces of the channel.

- *No channel knowledge at the transmitter*:

In such a case, it has been shown [21, 35] that the optimum transmit strategy is to pour power (uniformly) *blindly* in all spatial directions, i.e., $\mathbf{R_{ss}} = \frac{p_0}{M_T}\mathbf{I}$. This will lead to a reduced channel capacity (C_{blind}), given as:

$$C_{blind} = \frac{1}{2} log \left[det \left(\frac{p_0}{M_T} \mathbf{H}^* \mathbf{R_{nn}^{-1}} \mathbf{H} + \mathbf{I} \right) \right] \tag{2.33}$$

We note the following:

- If the channel matrix is *orthogonal* with $M = M_T = M_R$ (say) then it can be verified from equation (2.32) that the optimum $\mathbf{R_{ss}} = \mathbf{I}\frac{p_0}{M}$, even with channel knowledge at the transmitter. This is because orthogonal channels have equal gain on all eigenmodes, and hence distributing power uniformly on all eigenmodes is indeed the optimum transmit strategy. Hence, the resulting capacities, C and C_{blind} are also the same, given as:

$$C_{orth} = \frac{1}{2} log_2 \left[det \left(\frac{p_0 \mathbf{I}}{M\sigma^2} + \mathbf{I} \right) \right] = M \times \frac{1}{2} log \left(1 + \frac{p_0}{M\sigma^2} \right) \tag{2.34}$$

where we have assumed $\mathbf{R_{nn}} = \sigma^2 \mathbf{I}$. Furthermore, it can be seen from (2.34) that C_{orth} increases with increase in M. Hence, capacity increases with the increase in transmit and receive antennas, and this effect is is referred to as the *spatial multiplexing* effect [43]. Figure 2.8 plots the channel capacity, C_{orth}, for

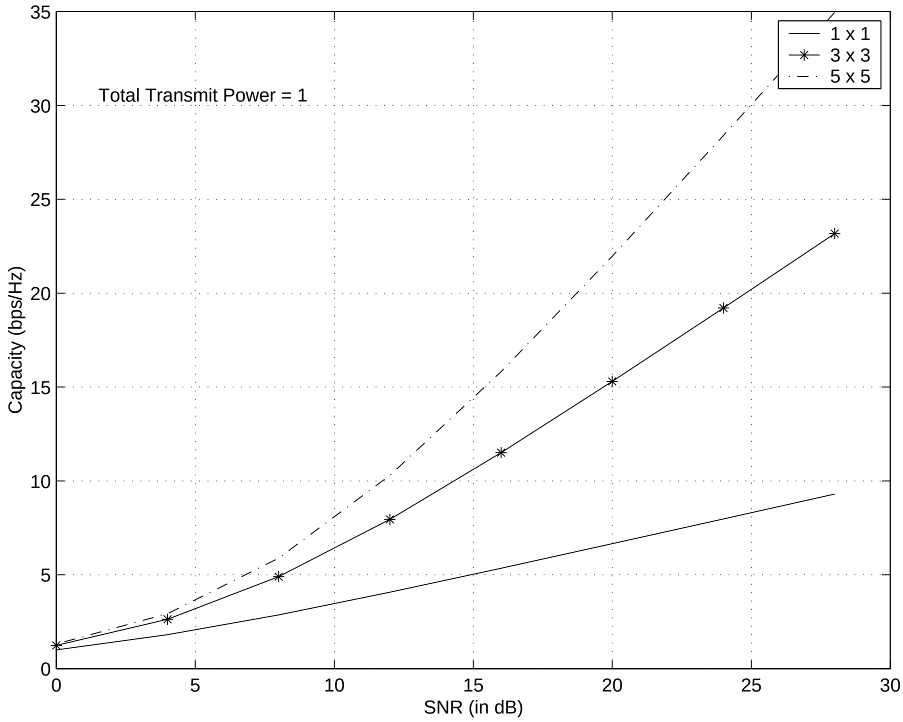

Figure 2.8: Channel capacity as a function of antennas

a 1×1, 3×3, and a 5×5 MIMO system.

- When the channel $\mathbf{H}$ is not perfectly orthogonal, we have $C \geq C_{blind}$. This is because, for non-orthogonal channels, eigenmodes have unequal gains. The optimal transmitter (with channel knowledge at the transmitter) will pour more power on the stronger eigenmodes and increase capacity, when compared to a transmitter that pours equal power on all eigenmodes. The latter wastes power in the null modes of the channel, leading to a lower capacity. This is illustrated in Figures 2.9 and 2.10 where we have plotted C and C_{blind} for a 5×5 flat-fading channel with antenna correlations of 0 and 1. It can be verified that as

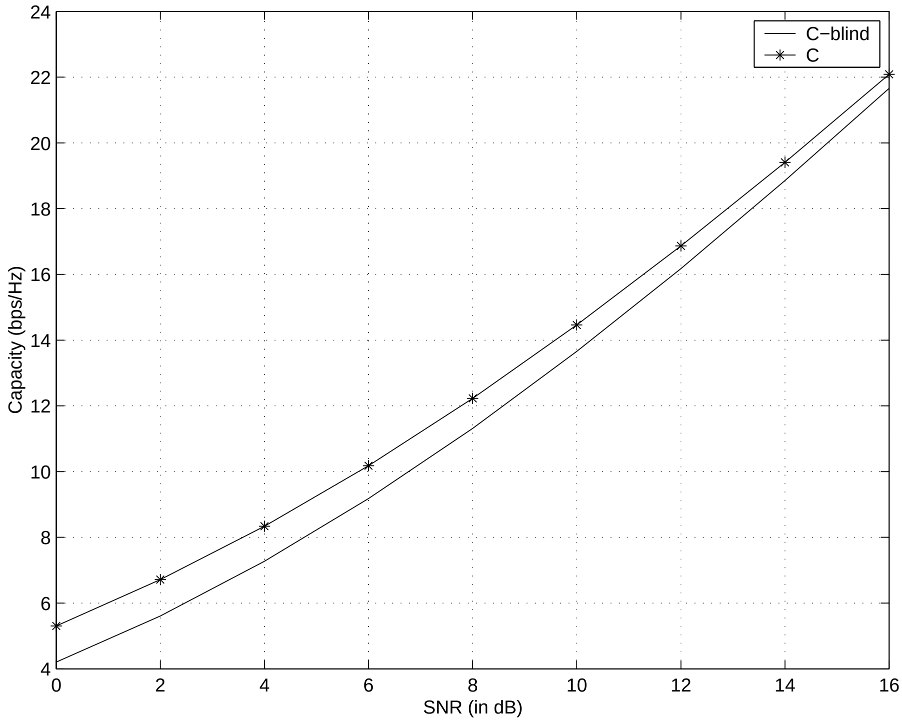

Figure 2.9: C vs. C_{blind}, for antenna correlation coefficient of 0.0

the antenna correlation increases, C is much larger than C_{blind}. Hence, there can be a significant loss in capacity if channel knowledge is not assumed at the transmitter.

2.3.2 Performance

In this section, we will demonstrate via Monte-Carlo simulations that existing space-time coded systems and spatial multiplexing systems that assume no channel knowledge at the transmitter lose in performance, when the MIMO channel matrix is not orthogonal. This is because non-orthogonal channels have eigenmodes that have low

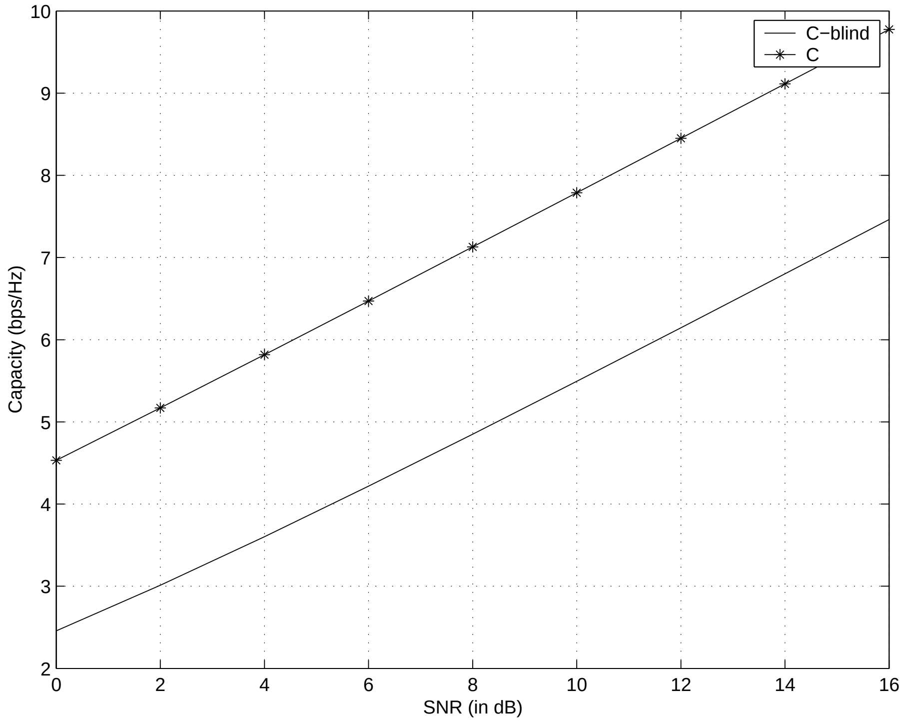

Figure 2.10: C vs. C_{blind}, for antenna correlation coefficient of 1.0

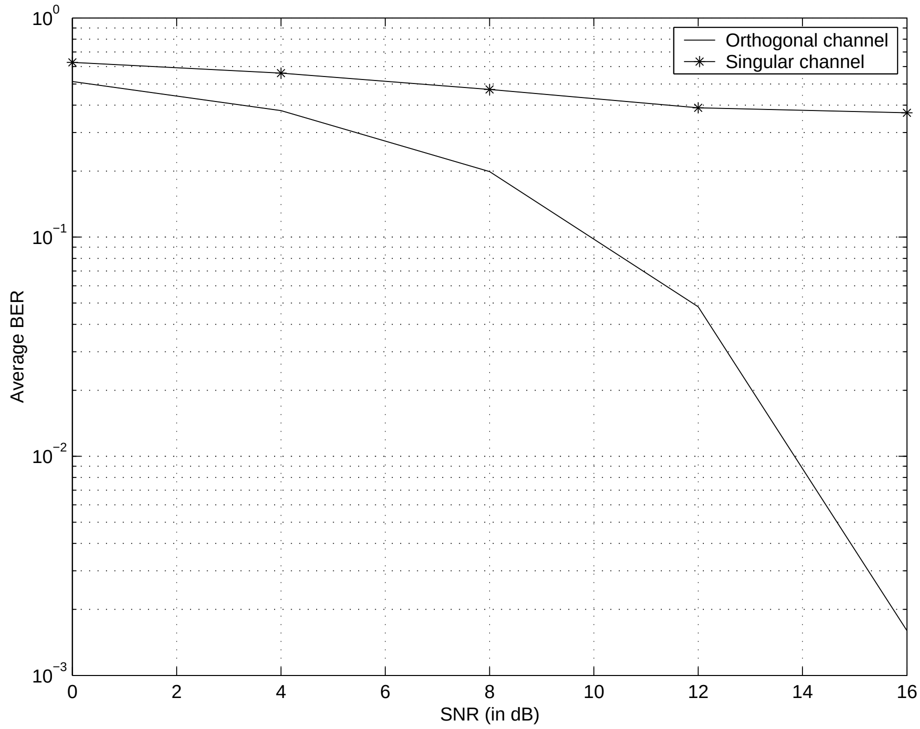

Figure 2.11: Performance degradation of spatial multiplexing system in rank-deficient channels

gains and can null out the transmitted signal leading to poor performance.

Figure 2.11 that plots the performance of a spatial multiplexing system that transmits 2 independent streams of data over a flat-fading MIMO channel with $M_T = 2$ and $M_R = 2$ antennas, and employing 4-QAM modulation. We assume uncorrelated noise on the receive antennas and that the total transmit power across all transmit antennas is normalized to unity. We consider the performance over both an orthogonal channel and a rank 1 channel. It can be seen that the the performance degrades in the latter case, due to the finite null-space in the rank 1 channel.

Existing space-time coded systems that do not use make use of the channel knowledge at the transmitter, also lose performance in non-orthogonal channels. However, the performance degradation is not as severe as that of spatial multiplexing systems, due to the spatial diversity benefits of the space-time code. This is illustrated in Figure 2.12 that plots the performance of a rate 1 space-time (Alamouti) block code developed in [39] for a $M_T = 2$ and $M_R = 2$ MIMO system employing 4-QAM modulation. We assume uncorrelated noise on the receive antennas and that the total transmit power across all transmit antennas is normalized to unity. We consider the performance for both an orthogonal channel and a rank 1 channel, and show that the performance degrades in the latter case.

This motivates the design of MIMO transmission schemes that assume channel knowledge at the transmitter, which is the main contribution of the thesis. Chapters 3,4,5 and 6 describe linear transmit and receive processing schemes that use channel knowledge at the transmitter and receiver to improve system performance.

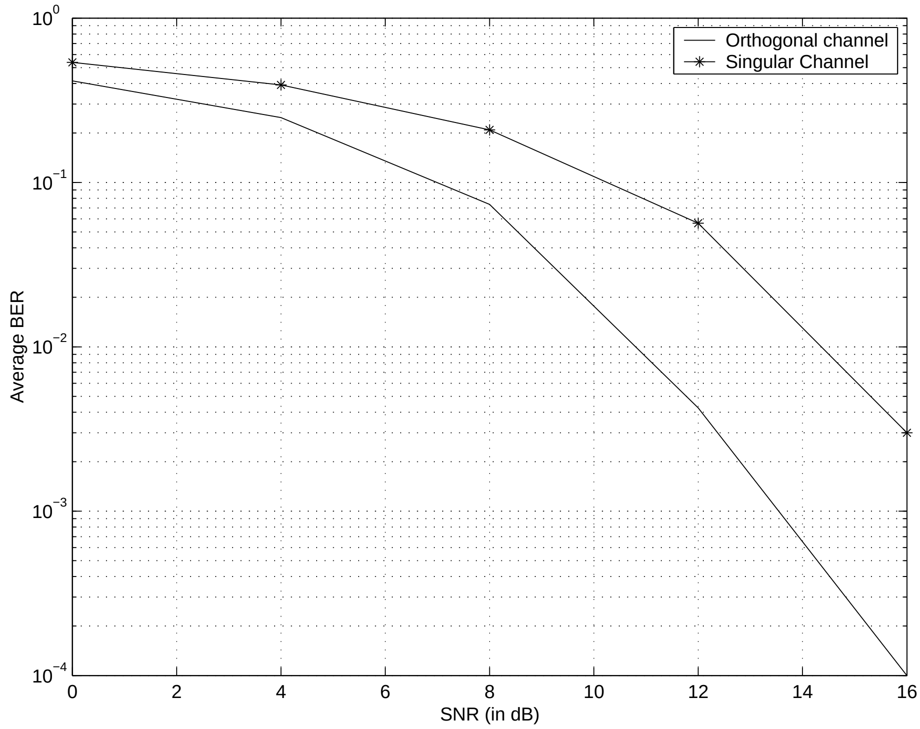

Figure 2.12: Performance degradation of space-time coded system in rank-deficient channels

Chapter 3

Generalized Linear Precoding and Decoding

3.1 Introduction

Multiple antennas at both ends of the wireless link can be used for spatial multiplexing [43, 45] and transmit diversity [15, 16, 31, 38, 39]. Spatial multiplexing involves transmitting independent streams of data across multiples antennas to maximize throughput, whereas transmit diversity schemes such as space-time coding involves mapping the same input symbol stream across space (multiple antennas) and across time, to get diversity gain. Both schemes assume no channel knowledge at the transmitter.

In a number of applications channel knowledge is also available at the transmitter. The channel can be estimated at the receiver and fedback to the transmitter in frequency division duplex (FDD) systems or can be estimated at the transmitter during the receiving mode in time division duplex(TDD) systems. In this chapter, we consider the scenario when both the transmitter and receiver have perfect channel knowledge, and investigate the design of optimal linear transmit and receive signal processing schemes to improve system performance for different applications. The transmit processing is done by the linear *precoder* and receive processing is done by the linear *decoder*. Linear processing schemes have lower implementation complexity, but typically suffer in performance with respect to non-linear schemes. In this chapter, we will restrict ourselves to linear processing schemes.

Jointly optimal linear precoder and decoder (LPD) designs have been known to provide better performance, when compared to receive only optimization [56, 46]. Space-Time linear precoding (and decoding) of symbols enables optimal allocation of resources such as power and bits over the spatial and temporal domain, depending on the channel conditions. Linear precoding (and decoding) is scalable to any number of transmit and receive antennas, and can be used in lieu of, or in conjunction with, coding to improve system performance.

We now summarize some jointly optimal LPD designs in the literature. The optimum design that *maximizes SNR* diagonalizes the MIMO channel and transmits data over the eigenmode with the maximum gain [78]. The optimum design that *maximizes information rate* or that *minimizes the sum of output symbol estimation errors* also decouple the channel into parallel channels and allocate the power on these sub-channels according to the waterpouring [70, 47] or inverse waterpouring policy [56, 48], respectively. The optimum design that maximizes SNR subject to a zero-forcing constraint on the channel has also been considered recently [48].

While the MMSE design minimizes the sum of symbol estimation errors across symbol streams, it does not tell us anything about the mean square error on each symbol stream. Some symbol streams may have larger errors compared to the others. In order to get a better handle on the errors on each symbol stream, we chose to design an optimum precoder and decoder that *minimize any weighted sum of symbol estimation errors* [1], subject to a total power constraint across all transmit antennas.

The solution to the weighted MMSE design also yielded solutions to well-known designs (solved in the literature and mentioned above) for different choices of error weights. Moreover, we were able to find new applications that were not considered in the literature. A nice feature of our generalized design is that we are now able to provide a unified framework to solve the above class of linear precoder and decoder designs.

We summarize the main results of the chapter:

[1] It will be seen later that the derivation of the weighted MMSE solution is quite different from the (unweighted) MMSE design, due to the considerable analysis involved in showing that the optimum solution diagonalize the MIMO channels into eigen sub-channels for any set of error weights.

- We first introduce the linear block precoder and decoder (LPD) system model for flat-fading channels and derive a closed form solution for generalized linear block precoding and decoding scheme based on the weighted mean square error criteria, assuming a total power constraint at the transmitter and a fixed channel realization.

- Next, we formally prove that the optimum design forces transmission only on the eigenmodes of the MIMO channel, for any set of error weights.(As an interesting side note, we will show in Chapter 4 that eigenmode transmission is not optimum if we assume a peak power constraint at the transmitter instead.)

- Next, we show that the power allocation on the eigenmodes depends on the error weights, which can be varied depending on the application. We show how to select error weights to obtain both well-known and new designs.

3.2 System Model

Consider the MIMO communication model in Figure 3.1. The input bit-stream is coded and modulated to generate one or more output symbol streams. The precoder linearly processes a block of these symbols and outputs a symbol block that is launched into the MIMO channel using M_T separate antennas. The precoder is channel dependent, operates in the complex field (rather than the Galois field) and

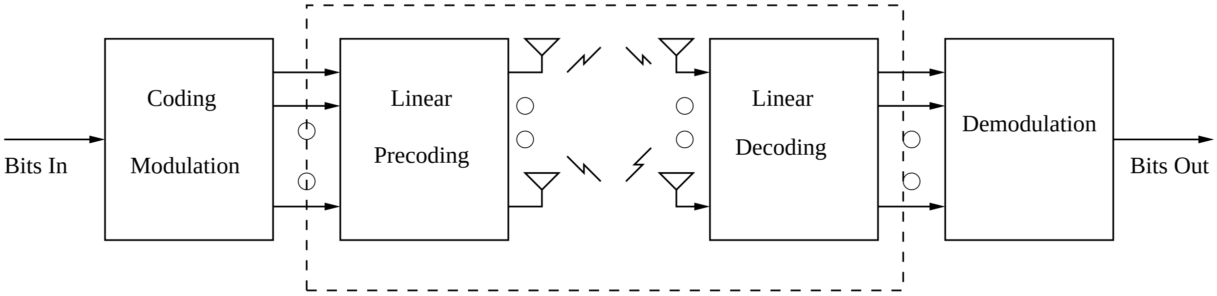

Figure 3.1: MIMO System with Linear Precoder and Decoder(LPD)

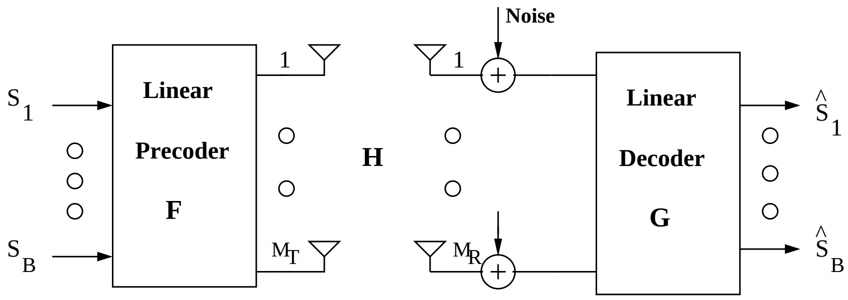

Figure 3.2: System Model: flat fading channel

can add redundancy to the input symbol stream to improve system performance (as will be explained later). The received signal is linearly processed by the decoder. The decoder, which is also channel-dependent, operates in the complex field and removes any redundancy that has been introduced by the precoder.

In the present chapter, we will focus on the *boxed* section of the communication system shown in Figure 3.1, i.e, we will not consider coding and modulation design, but rather focus on the design of the precoder and decoder. We will assume that the input symbol stream have been coded and modulated according to well-known schemes.

Flat-Fading Channel:

For a MIMO channel without any delay-spread (see Figure 3.2), the system equation is

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn} \tag{3.1}$$

where $\mathbf{H}$ is an $M_R \times M_T$ channel matrix, $\hat{\mathbf{s}}$ is the $B \times 1$ received vector and $\mathbf{s}$ is the $B \times 1$ transmitted vector, where $B = rank(\mathbf{H}) \leq min(M_R, M_T)$ is the number of parallel streams of data that is to be transmitted; $\mathbf{n}$ is the $M_R \times 1$ noise vector (at the given symbol time); finally, $\mathbf{G}$ is the $B \times M_R$ decoder matrix and $\mathbf{F}$ is the $M_T \times B$ precoder matrix. The latter adds a redundancy of $M_T - B$ across space, since it has B input symbols and M_T precoded output symbols that are launched simultaneously through M_T separate antennas. Note that linear precoding and decoding are implemented by simple matrix multiplications. We assume that

$$E(\mathbf{ss}^*) = \mathbf{I}; \;\; E(\mathbf{nn}^*) = \mathbf{R_{nn}}; \;\; E(\mathbf{sn}^*) = 0, \tag{3.2}$$

where the superscript * denotes the conjugate transpose. Note that, for simplicity of analysis, we have assumed uncorrelated input sources, each normalized to unit power, and that $B = rank(\mathbf{H})$. Based on the structure of the optimum precoder and decoder, we will later show how to handle the case when $B < rank(\mathbf{H})$. The case when $B > rank(\mathbf{H})$ is not possible, since the transmitter knows the channel and hence for any practical system will not transmit more than $rank(\mathbf{H})$ independent data streams.

3.3 Problem Formulation

Our goal is to design the $\mathbf{F}$ and $\mathbf{G}$ matrices to minimize any weighted combination of symbol estimation errors, namely $E[\mathbf{e}^*\mathbf{W}^{*1/2}\mathbf{W}^{1/2}\mathbf{e}]$ where $\mathbf{e} = \mathbf{s} - (\mathbf{GHFs} + \mathbf{Gn})$ is the $B \times 1$ error vector and and $\mathbf{W}^{1/2}$ is the $B \times B$ square root of a *diagonal* positive definite weight matrix $\mathbf{W}$. We assume that the channel matrix $\mathbf{H}$ is fixed and known at the transmitter and receiver. Mathematically, the problem statement is as follows:

$$\min_{\mathbf{G},\mathbf{F}} : c(\mathbf{G},\mathbf{F}) = E||\mathbf{W}^{1/2}\mathbf{e}||^2 \tag{3.3}$$
$$Subject\ \ to: \ \ tr(\mathbf{F}\mathbf{F}^*) \le p_0$$

where the expectation (E) is with respect to the distributions of $\mathbf{s}$ and $\mathbf{n}$, and p_0 is the total power available. Note that

$$c(\mathbf{G},\mathbf{F}) = E||\mathbf{W}^{1/2}\mathbf{e}||^2 = E(tr[\mathbf{W}^{1/2}\mathbf{e}\mathbf{e}^*\mathbf{W}^{*1/2}]) = tr[\mathbf{W} \times \mathbf{R_e}(\mathbf{G},\mathbf{F})] \tag{3.4}$$

where $\mathbf{R_e}(\mathbf{G},\mathbf{F})$ is the error covariance matrix, defined as

$$\mathbf{R_e}(\mathbf{G},\mathbf{F}) := E(\mathbf{ee}^*) \tag{3.5}$$

Using the expression for $\mathbf{e}$ in equation (3.5) we have

$$\mathbf{R_e}(\mathbf{G},\mathbf{F}) = E[\mathbf{s} - (\mathbf{GHFs} + \mathbf{Gn})][\mathbf{s} - (\mathbf{GHFs} + G\mathbf{n})]^*$$

or alternatively

$$\begin{aligned}\mathbf{R_e}(\mathbf{G},\mathbf{F}) &= E[\mathbf{ss}^*] + \mathbf{GHF}E[\mathbf{ss}^*]\mathbf{F}^*\mathbf{H}^*\mathbf{G}^* - \mathbf{GHF}E[\mathbf{ss}^*] - E[\mathbf{ss}^*]\mathbf{F}^*\mathbf{H}^*\mathbf{G}^* \\ &+\mathbf{GHF}E[\mathbf{sn}^*]\mathbf{G}^* + \mathbf{G}E[\mathbf{ns}^*]\mathbf{F}^*\mathbf{H}^*\mathbf{G}^* - \mathbf{G}E[\mathbf{ns}^*] + \mathbf{G}E[\mathbf{nn}^*]\mathbf{G}^* - \mathbf{G}E[\mathbf{sn}^*]\end{aligned} \tag{3.6}$$

Using the assumptions made in equation (3.2), the above equation can be simplified as:

$$\mathbf{R_e}(\mathbf{G},\mathbf{F}) = (\mathbf{GHF} - \mathbf{I})(\mathbf{GHF} - \mathbf{I})^* + \mathbf{GR_{nn}G}^* \tag{3.7}$$

We use the method of Lagrange Duality and the Karush-Kuhn-Tucker (KKT) conditions (see below) to solve the optimization problem in (3.3). We note that the KKT conditions were also used in [48, 56] to design the jointly optimal precoder and decoder using the (unweighted) MMSE criterion.[2].

We first form the Lagrangian (below μ is the Lagrange multiplier)

$$L(\mu, \mathbf{G}, \mathbf{F}) = c(\mathbf{G}, \mathbf{F}) + \mu[tr(\mathbf{F}\mathbf{F}^*) - p_0] \tag{3.8}$$

Using (3.4) and (3.7) in (3.8),we can write:

$$L(\mu, \mathbf{G}, \mathbf{F}) = tr[\mathbf{W}(\mathbf{GHF} - \mathbf{I})(\mathbf{GHF} - \mathbf{I})^* + \mathbf{W}\mathbf{G}\mathbf{R_{nn}}\mathbf{G}^*] + \mu[tr(\mathbf{F}\mathbf{F}^*) - p_0] \tag{3.9}$$

The following KKT conditions are necessary and sufficient for optimality: $\mathbf{G}$ and $\mathbf{F}$ are optimal if and only if there is a μ that together with $\mathbf{G}$ and $\mathbf{F}$ satisfy the conditions:

$$\nabla_{\mathbf{G}} L(\mu, \mathbf{G}, \mathbf{F}) = 0 \tag{3.10}$$

$$\nabla_{\mathbf{F}} L(\mu, \mathbf{G}, \mathbf{F}) = 0 \tag{3.11}$$

$$\mu \geq 0; \;\; tr(\mathbf{F}\mathbf{F}^*) - p_0 \leq 0 \tag{3.12}$$

$$\mu[tr(\mathbf{F}\mathbf{F}^*) - p_0] = 0 \tag{3.13}$$

Using equation (3.9) in equations (3.10) and (3.11) we get the following relations

[2]Note that the rest of the proof steps are quite different from the proof of the (unweighted) MMSE criterion, and involves considerable analysis in showing that the optimal precoder and decoder diagonalize the MIMO channel for any set of error weights

between $\mathbf{G}$ and $\mathbf{F}$ (respectively):

$$\mathbf{HF} = \mathbf{HFF^*H^*G^*} + \mathbf{R_{nn}G^*} \tag{3.14}$$

$$\mathbf{WGH} = \mathbf{F^*H^*G^*WGH} + \mu\mathbf{F^*} \tag{3.15}$$

To obtain the above equations (3.14) and (3.15), we have used the fact that $\frac{\partial tr(\mathbf{AXB})}{\partial \mathbf{X}} = \mathbf{BA}$ and $\frac{\partial tr(\mathbf{AX^*B})}{\partial \mathbf{X}} = 0$ from the theory of matrix derivatives (see, for example, [52]) for a complex-valued [3] $\mathbf{X}$.

3.4 Solution

We now solve the equations (3.14) and (3.15) to derive the optimum precoder and decoder. Let us define the eigenvalue decomposition (EVD):

$$\mathbf{H^*R_{nn}}^{-1}\mathbf{H} = (\mathbf{V}\ \ \tilde{\mathbf{V}}) \begin{pmatrix} \mathbf{\Lambda} & 0 \\ 0 & \tilde{\mathbf{\Lambda}} \end{pmatrix} (\mathbf{V}\ \ \tilde{\mathbf{V}})^* \tag{3.16}$$

where $\mathbf{V}$ is an $M_T \times B$ orthogonal matrix which forms a basis for the range space of $\mathbf{H^*R_{nn}}^{-1}\mathbf{H}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the B non-zero eigenvalues arranged in a decreasing order from top-left to bottom-right; $\tilde{\mathbf{\Lambda}}$ contains the zero

[3] When evaluating the derivatives in (3.10) and (3.11), $\mathbf{G}$ and $\mathbf{G}^*$ are treated as independent variables, and similarly for $\mathbf{F}$ and $\mathbf{F}^*$; the differentiations of (3.9) wrt $\mathbf{F}^*$ and $\mathbf{G}^*$ just gives (3.14) and (3.15), and hence brings no additional equations

eigenvalues; $\tilde{\mathbf{V}}$ is a $M_T \times (M_T - B)$ orthogonal matrix which constitutes a basis for the null-space of $\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}$. Note that for now we have assumed $rank(\mathbf{H}) = B$ as explained in the previous section.

Lemma 1 *The optimum* $\mathbf{F}$ *and* $\mathbf{G}$ *matrices can be assumed to have the following structure without any loss of generality* [4]*:*

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f} \tag{3.17}$$

$$\mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} \tag{3.18}$$

where $\mathbf{\Phi_f}$ *and* $\mathbf{\Phi_g}$ *are* $B \times B$ *matrices.*

Proof: See Appendix A.1 ∎

Lemma 2 *The matrices* $\mathbf{\Phi_f}$ *and* $\mathbf{\Phi_g}$ *in Lemma 1 are diagonal with non-negative elements on the diagonal. Hence, the optimum precoder and decoder diagonalize* $\mathbf{H}$ *for any set of error weights. Specifically, we have* $\mathbf{GHF} = \mathbf{\Phi_g}\mathbf{\Lambda}\mathbf{\Phi_f}$*.*

Proof: See Appendix A.2 ∎

Lemmas 1 and 2 can be better understood through the *equivalent system model* shown in Figure 3.3. The transmitted signal vector $\mathbf{s}$ is first scaled by the diagonal matrix $\mathbf{\Phi_f}$ and then rotated by the matrix $\mathbf{V}$, before being launched into the MIMO

[4]We thank Dr. Anna Scaglione for valuable comments on formulating and proving this lemma

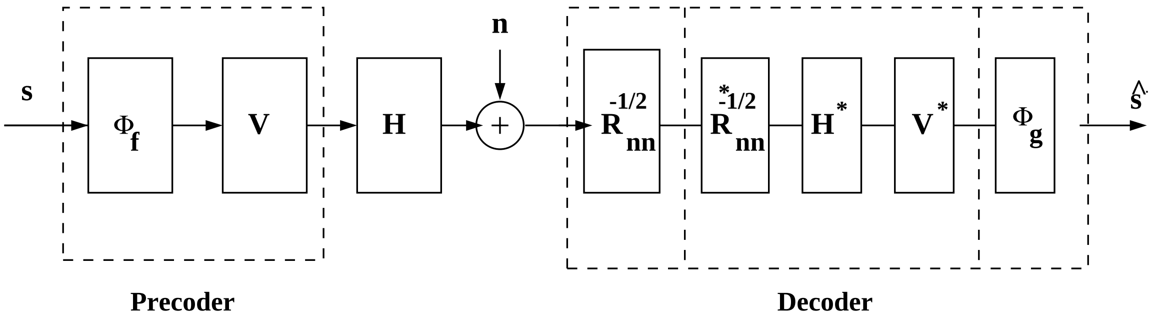

Figure 3.3: Equivalent system model

channel. A natural consequence of this rotation is that all power is poured only on the eigenmodes of the matrix $\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}$. The power allocation matrix $\boldsymbol{\Phi}_{\mathbf{f}}{}^2$, which is a function of the error weights, determines what fraction of the total power is sent in each eigenmode. These error weights can be chosen depending on the application, as will be illustrated later in the chapter. At the receiver, the first operation is noise whitening through the matrix $\mathbf{R}_{\mathbf{nn}}{}^{-1/2}$. This is followed by *matched filtering* of the noise-whitened signal, represented by the matrix product $\mathbf{V}^*\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^*{}^{-1/2}$. The whitened matched filtered signal is then scaled appropriately by the matrix $\boldsymbol{\Phi}_{\mathbf{g}}$ to get the received estimate, $\hat{\mathbf{s}}$.

The equivalent diagonalized channel, with the transmit power allocation, and receiver scaling on each eigenmode is given in Figure 3.4.

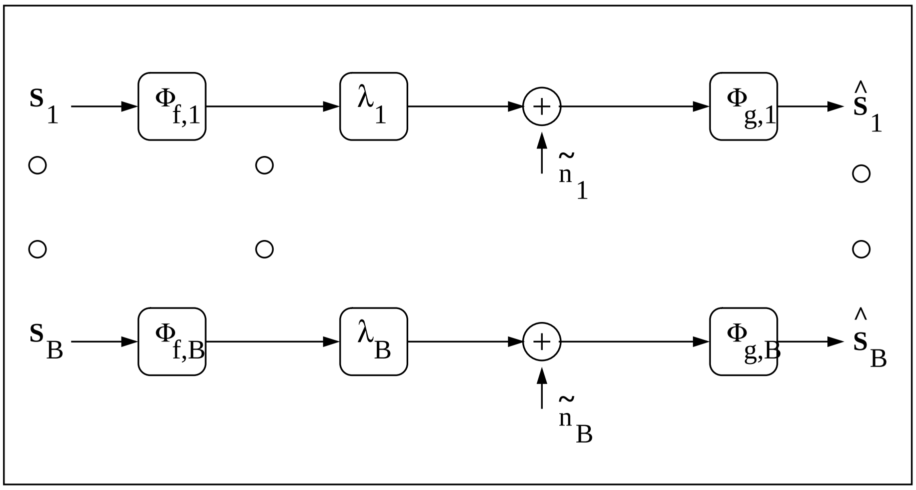

Figure 3.4: Equivalent diagonalized channel

Theorem 1 *The optimum* $\mathbf{\Phi_f}$ *and* $\mathbf{\Phi_g}$ *in Lemma 1 are given by:*

$$\mathbf{\Phi_f} = (\mu^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{1/2} - \mathbf{\Lambda}^{-1})_+^{1/2} \tag{3.19}$$

$$\mathbf{\Phi_g} = (\mu^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2} - \mu\mathbf{\Lambda}^{-1}\mathbf{W}^{-1})_+^{1/2}\mathbf{\Lambda}^{-1/2} \tag{3.20}$$

The symbol $(.)_+$ *denotes the fact that that the negative elements of the diagonal bracketed matrix are replaced by zero. The* μ *in (3.19) and (3.20) is chosen to satisfy the transmitter power constraint.*

Proof: See Appendix A.3 ■

We now show how to obtain a $\mu > 0$ that satisfies the power constraint and also is such that the bracketed matrices in both (3.19) and (3.20) are positive semi-definite. Let $\mathbf{\Lambda} = diag([\lambda_1, \lambda_2, ..., \lambda_B])$, $\mathbf{W} = diag([w_1, w_2, ..., w_B])$, and $\rho_k = \lambda_k w_k$ and let them be ordered in a decreasing manner: $\rho_1 \geq .. \geq \rho_k \geq .. \geq \rho_B$. We get the following expression for μ from the trace constraint $tr(\mathbf{\Phi_f}^2) = p_0$, when $k \leq B$ sub-channels are used for transmission.

$$\mu^{1/2} = \frac{\sum_{j=1}^{k}(\lambda_j^{-1/2} w_j^{1/2})}{p_0 + \sum_{j=1}^{k}(\lambda_j^{-1})} \tag{3.21}$$

The following iterative procedure initialized with $k = B$ can be used to optimally compute $\mathbf{\Phi_f}$. Let $\mathbf{\Phi_f} = diag([\phi_{f1}, \phi_{f2}, ..., \phi_{fB}])$.

1. Assume $\mu \leq \rho_k$, which from (3.19) implies $\phi_{fj} \geq 0$, for $j = 1, .., k$. Compute μ according to (3.21) under this assumption. If $\mu \leq \rho_k$ stop; else go to step 2.

2. Set $\phi_{f,k} = 0$ and $k = k - 1$. Go to step 1.

It can be seen that for each additional iteration, zero power is allocated to one of the eigenmodes. Since there are a total of B eigenmodes, the iteration will terminate in at most $B - 1$ steps. In such a case all power will be allocated to the only available (strongest) eigenmode.

Remark 1: The optimum precoder and decoder were derived for the case when $B = rank(\mathbf{H})$. The extension to the case when $B < rank(\mathbf{H})$ is more or less straightforward. The optimum precoder and decoder still forces transmission on the eigenmodes of the the MIMO channel as before. However, the power is likely to be allocated to only the the B eigenmodes with the highest gains. This is because any other choice of B eigenmodes will choose a weaker eigenmode over a stronger eigenmode and hurt system performance.

We now provide expressions for the BER, SNR and mean square error (MSE) across the sub-channels for the weighted MMSE design. We will need these expressions later on.

Assuming a square QAM constellation, the probability of symbol error (and hence the BER assuming Gray encoding) for the i-th sub-channel is given by (see for example, [77]): $P_{e,i} = N_{e,i}\, Q(\sqrt{\beta_{M_i}\,\Gamma_i\,\gamma_{c,i}(\Gamma_i)})$, where for a given constellation size M_i, $\beta_{M_i} = \frac{3}{(M_i-1)}$, $\gamma_{c,i}(\Gamma_i)$ is the coding gain of the channel code that is applied to the symbol stream prior to the linear precoder,[5] and Γ_i is the sub-channel SNR and is obtained from the (i,i)-th element in the SNR matrix defined as:

$$\Gamma = (\mathbf{F}^*\mathbf{H}^*\mathbf{G}^*)(\mathbf{G}\mathbf{R}_{\mathbf{nn}}\mathbf{G}^*)^{-1}\mathbf{G}\mathbf{H}\mathbf{F} \tag{3.22}$$

Note that $\gamma_{c,i}(\Gamma_i)$ is a function of the sub-channel SNR, Γ_i. In high SNR conditions, the coding gain is relatively constant across the range of SNRs over which the system is operated and we can assume $\gamma_{c,i}(\Gamma_i) \simeq \gamma_{c,i}$. In moderate to low SNRs, we typically have $\gamma^0_{c,i} \leq \gamma_{c,i}(\Gamma_i) \leq \gamma^1_{c,i}$, where $\gamma^0_{c,i}$ and $\gamma^1_{c,i}$ are a function of the code used and the range of SNRs over which the system is operated. In such a case, we can make a *first order* approximation and let $\gamma_{c,i}(\Gamma_i) \simeq \frac{\gamma^0_{c,i}+\gamma^1_{c,i}}{2}$. Hence, the probability of symbol

[5] If no channel coding has been used on the symbol stream prior to the precoder, $\gamma_{c,i}(\Gamma_i) = 1$

error can be approximated as:

$$P_{e,i} \simeq N_{e,i}\, Q(\sqrt{\beta_{M_i}\, \Gamma_i\, \gamma_{c,i}}) \tag{3.23}$$

Using the optimum $\mathbf{\Phi_f}$ and $\mathbf{\Phi_g}$ matrices from Lemma 1 and Theorem 1, the SNR matrix given in (3.22) can be simplified to

$$\Gamma = \mathbf{\Phi_f}^2 \mathbf{\Lambda} = (\mathbf{W}^{1/2} \mu^{-1/2} \mathbf{\Lambda}^{1/2} - \mathbf{I})_+ \tag{3.24}$$

and the $\mathbf{R_e}$ matrix in (3.7) can be simplified to,

$$\mathbf{R_e} = \mathbf{W}^{-1/2} \mathbf{\Lambda}^{-1/2} \mu^{1/2} \tag{3.25}$$

From equations (3.24) and (3.25), we get the following relationship between SNR and MSE, for *any set of error weights*.

$$\mathbf{\Gamma} = \mathbf{R_e}^{-1} - \mathbf{I} \tag{3.26}$$

An interesting observation from (3.26) is that equal SNRs on each subchannel (eigen-mode) implies equal MSEs, for any set of error weights.

3.5 Applications

In this section, we show how to obtain Linear Precoder and Decoder (LPD) designs for different applications by appropriately choosing the error weights $\mathbf{W}$.

3.5.1 Maximum Information Rate Design

Consider the system equation $\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn}$. The mutual information between $\hat{\mathbf{s}}$ and $\mathbf{s}$ is given by the following well-known expression (see for example [48]):

$$I(\hat{\mathbf{s}}; \mathbf{s}) = log_2(\, det[I + (\mathbf{GHF})^*(\mathbf{GR_{nn}G}^*)^{-1}(\mathbf{GHF})]\,) \tag{3.27}$$

We now restate the following Theorem derived in [47].

Theorem 2 *The optimum precoder and decoder that achieves the maximum information rate is:*

$$\begin{aligned} \mathbf{F} = \mathbf{V\Phi_f};\;\; \mathbf{G} = \mathbf{\Phi_g V^* H^* R_{nn}}^{-1} \\ \mathbf{\Phi_f} = (\frac{\mathbf{I}}{\mu^{1/2}} - \mathbf{\Lambda}^{-1})_+^{1/2} \end{aligned} \tag{3.28}$$

where $\mu > 0$ is computed according to the transmit power constraint, and $\mathbf{\Phi_g}$ can be any arbitrary full-rank diagonal matrix. $\mathbf{\Phi_f}$ is the well-known waterpouring *solution.*

From the above theorem, we get that $\mathbf{GHF} = \mathbf{\Phi_g \Lambda \Phi_f}$; and $\mathbf{GR_{nn}G}^* = \mathbf{\Phi_g \Lambda \Phi_g}^*$.

Using these results in (3.27), we get:

$$I(\hat{\mathbf{s}};\mathbf{s}) = \sum_{i=1}^{B} log_2(1+\phi_{f,i}^2\lambda_i) \tag{3.29}$$

which is maximized when $\mathbf{\Phi_f}$ is given by (3.28). The waterfilling solution puts more power on the stronger modes and less or no power on the weaker modes to achieve capacity. Note that $\mathbf{\Phi_g}$ can be any arbitrary full-rank diagonal matrix since it does not enter in the expression for the maximum information rate (3.29).

Lemma 3 *The choice* $\mathbf{W} = \mathbf{\Lambda}$ *in the expression for* $\mathbf{\Phi_f}$ *obtained from the weighted MMSE design (see (3.19)) results in the well-known waterpouring solution given in (3.28). Hence, the information rate maximizing design is just a special case of our generalized design.*

For the sake of completeness, we give below the expression for $\mathbf{\Phi_g}$ obtained when we substitute $\mathbf{W} = \mathbf{\Lambda}$ in (3.20):

$$\mathbf{\Phi_g} = (\mu^{1/2}\mathbf{\Lambda}^{-1} - \mu\mathbf{\Lambda}^{-2})_+^{1/2}\mathbf{\Lambda}^{-1/2} \tag{3.30}$$

The choice of error weights $\mathbf{W} = \mathbf{\Lambda}$ can also be obtained as follows. The mutual information in (3.29) can be rewritten as:

$$\begin{aligned} I(\hat{\mathbf{s}};\mathbf{s}) &= \sum_{i=1}^{B} log_2(1+\phi_{f,i}^2\lambda_i) \\ &= \sum_{i=1}^{B} log_2(w_i^{1/2}\mu^{-1/2}\lambda^{1/2}) \end{aligned} \tag{3.31}$$

where we have used the expression for ${\phi_{f,i}}^2\lambda_i$ from equation (3.24). Maximizing the above equation w.r.t w_i subject to the transmit power constraint, we get the solution $w_i = \lambda_i$.

It is interesting to note that the above choice of error weight *forces the cost (weighted error) for all eigenmodes over which non-zero power is allocated, to be equal.* This can be verified by substituting $\mathbf{W} = \mathbf{\Lambda}$ in equation (3.25) to get

$$\mathbf{W}\mathbf{R_e} = \mathbf{W}^{1/2}\mathbf{\Lambda}^{-1/2}\mu^{-1/2} = \mu^{-1/2} \tag{3.32}$$

Adaptive Transmission Systems-

The maximum information rate (max-IR) design finds application in *adaptive modulation systems* (see for example [55]), where more power and higher order modulation are used on subchannels with higher gains to improve data rates. The achievable rates across subchannels for asymptotically low probability of error are given by (3.29).

Table 3.1 shows how to compute the subchannel rates for a 5×5 spatial multiplexing system, for some given channel realization. We employ the maximum information

Subchannel Number(i)	λ_i	$\phi^2_{f,i}$	$\Gamma_i = \phi^2_{f,i}\lambda_i$ (in dB)	$\mathcal{R}_i = log(1+\Gamma_i)$ (bps/hz)	M_i
1	300	0.22	18.2	6.1	64
2	100	0.21	13.28	4.5	16
3	60	0.21	10.92	3.7	8
4	30	0.19	7.54	2.7	4
5	20	0.17	5.38	2.1	4

Table 3.1: Maximum information rate design

rate design to diagonalize the MIMO channel into 5 eigen subchannels (eigenmodes), and transmit power using the waterfilling solution. The signaling constellation across each subchannel is given by $M_i = 2^{floor(I_i)}$. By choosing the modulation scheme depending on subchannel SNRs, the maximum information rate design achieves a significant fraction of the channel capacity [55].

3.5.2 (Unweighted) MMSE Design

The jointly optimal precoder and decoder that minimize the *sum of symbol estimation errors* is well-known in the literature[56][48]. Such a design can be trivially obtained by choosing $\mathbf{W} = \mathbf{I}$ in Theorem 1.

We make the following observations.

- The MMSE power allocation policy is given as follows:

$$\boldsymbol{\Phi}_{\mathbf{f}}{}^2 = (\mu^{-1/2}\boldsymbol{\Lambda}^{-1/2} - \boldsymbol{\Lambda}^{-1})_+ \tag{3.33}$$

The MMSE policy is quite similar to the information rate maximizing *water-filling* solution in that it also drops the weaker eigenmodes. However, the main difference arises in the power reallocated among the remaining eigenmodes. The MMSE policy redistributes the power so that more power is allocated to the weaker modes and vice-versa. This is in contrast to the waterfilling policy that allocates more power to stronger eigenmodes and less power to weaker eigenmodes. This difference in power allocation policies leads to a different capacity and total mean square error between the two designs.

This difference is illustrated in Figures 3.5 and 3.6 that compares the capacity and total mean square error for both the power allocation policies for a MIMO system with $M_T = M_R = 3$ antennas. We assume the total transmit power across all antennas to be unity,white noise across the receive antennas and an iid fading channel. Our results are averaged over 10000 Monte-Carlo simulation runs. For each run, a new instantiation of $\mathbf{H}$, is used, and the $\mathbf{F}$ and $\mathbf{G}$ are recomputed. The SNR is defined as the total transmit power across all transmit antennas to the receive noise variance on each receive antenna. We see that due to the difference in power allocation policies, the MMSE power policy loses capacity while the waterfilling power policy results in a higher mean square error.

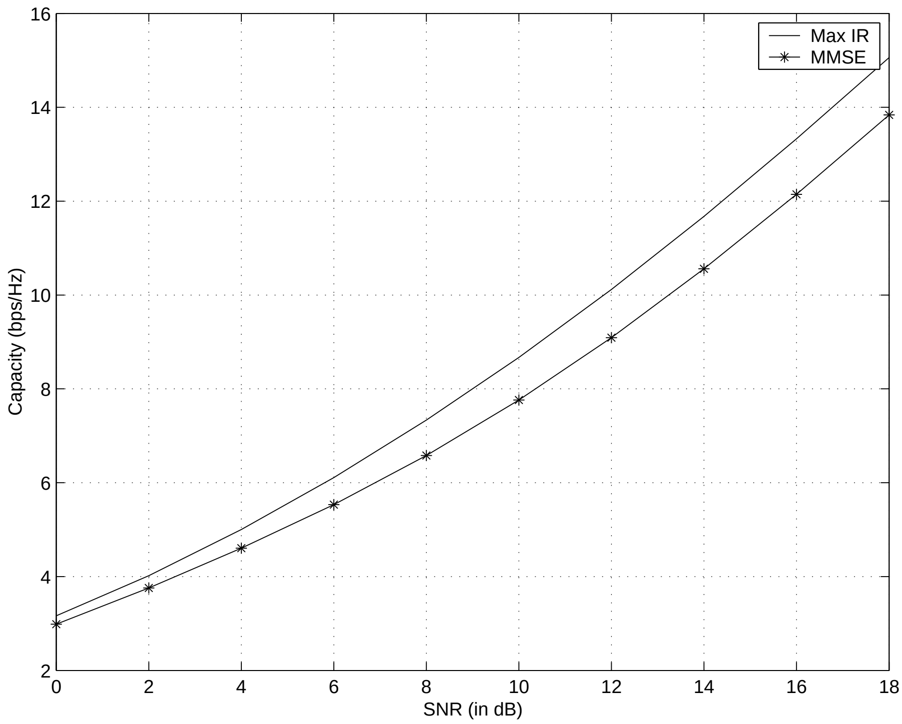

Figure 3.5: Capacity: MMSE vs. Information Rate Maximizing design

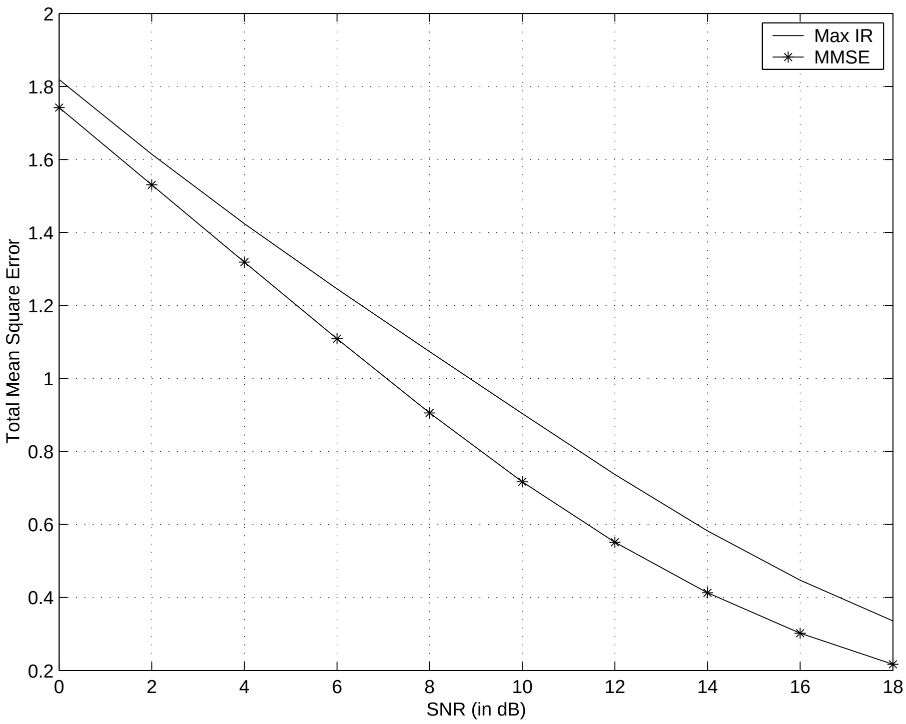

Figure 3.6: Total mean square error: MMSE vs. Information Rate Maximizing design

- At high gains (i.e., as $\mathbf{\Lambda}^{-1} \rightarrow 0$), we have:

$$\mathbf{\Phi_f} \simeq \mu^{-1/4}\mathbf{\Lambda}^{-1/4}; \qquad \mathbf{\Phi_g} \simeq \mu^{1/4}\mathbf{\Lambda}^{-3/4}; \qquad \mathbf{GHF} = \mathbf{\Phi_g}\mathbf{\Lambda}\mathbf{\Phi_f} \simeq \mathbf{I} \tag{3.34}$$

  providing the interpretation that the optimum precoder and decoder act as *spatial equalizers*, that attempt to convert the MIMO channel into an identity matrix.

- The MMSE design minimizes the sum of the symbol estimation errors across all subchannels, but does not guarantee that the MSEs and SNRs on each subchannel is minimized. In fact, the weaker subchannels can have a higher MSE than the stronger subchannel (as observed by substituting $\mathbf{W} = \mathbf{I}$ in (3.24) and (3.25)).

We showed in [5] that MMSE design improves system performance for fixed rate systems that transmit multiple data streams, each with an identical modulation scheme. Consider a downlink spatial multiplexing system, where the base-station (BTS) has M_T transmit antennas and the mobile have $M_R = 2$ receive antennas ($M_T \geq 2$). We send two independent streams of data ($B = 2$) , each 4-QAM modulated. We assume an iid matrix channel, and a total transmit power of unity. Figure 3.7 compares the performance of a system with no precoding (that employs only an

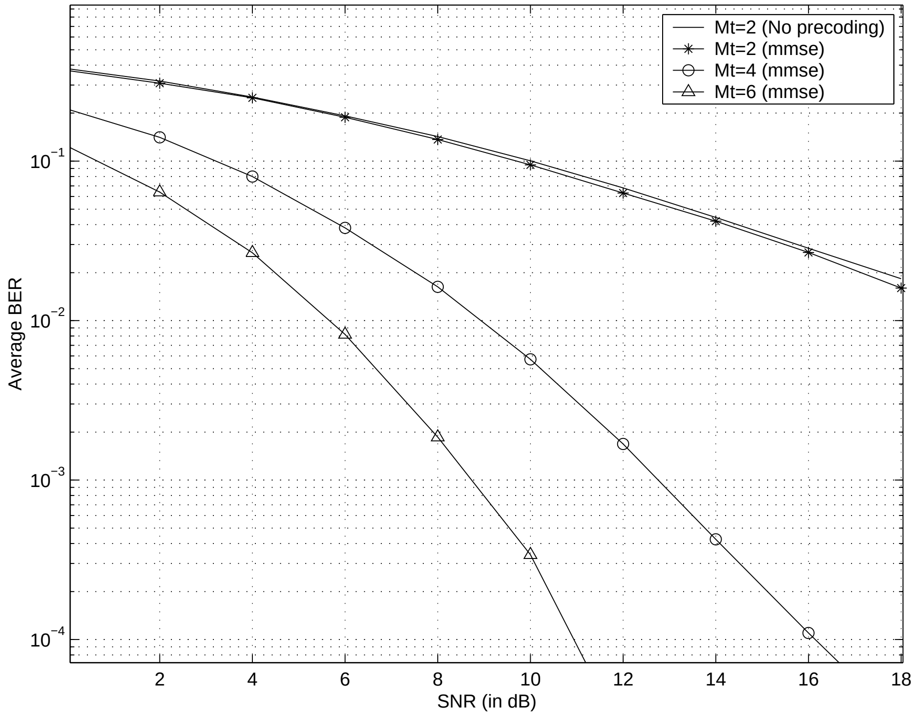

Figure 3.7: Performance improvement with an MMSE design

MMSE receiver), with that of a system with precoding. We see that precoding improves the system performance as the number of transmit antennas are increased. This can be explained by the redundancy added by the precoder that takes in $B = 2$ data streams and maps it over $M_T \geq 2$ transmit antennas, to get additional transmit diversity gain from additional transmit antennas. Note that our results are averaged over 200000 Monte-Carlo Simulation runs. Over each run, a new instantiation of $\mathbf{H}$ is chosen, and the optimum precoder and decoder are recomputed for that channel instantiation. The SNR is the total transmit power to the total noise variance across each receive antennas.

3.5.3 Equal Error Design

The MMSE design developed in the previous section minimizes the total error on symbol streams transmitted across all subchannels. However, it does not guarantee that the symbol stream on each subchannel have equal error. In fact, zero power can be sometimes allocated to symbol streams, if it is being transmitted over a weak eigenmode, leading to large errors on that symbol stream.

In some applications, such a discrepancy in symbol stream errors may not be desirable. For example, each symbol stream can belong to a different user, and it may be necessary to maintain equal errors on all streams. We will now show how to use the weighted error solution to design an optimal precoder-decoder scheme that can guarantee equal errors on all symbol streams.

From equation (3.23), providing equal symbol error probability implies that we need to design the precoder and decoder to guarantee equal SNRs for each symbol stream. This boils down to solving the following equation[6] with respect to $\mathbf{W}$:

$$\Gamma = \mathbf{W}^{1/2}\mu^{-1/2}\mathbf{\Lambda}^{1/2} - \mathbf{I} = \gamma\mathbf{I} \tag{3.35}$$

[6] We have assumed for simplicity that $\mathbf{\Phi_f} > \mathbf{0}$ and that $B = rank(\mathbf{H})$. The case when $B < rank(\mathbf{H})$ can be handled by using the argument in Remark 1.

where $\gamma > 0$ is a scalar. The solution $\mathbf{W}$ is readily computed from (3.35) as:

$$\mathbf{W}^{1/2} = (1+\gamma)\mathbf{\Lambda}^{-1/2}\mu^{1/2} \tag{3.36}$$

Note that μ is a function of $\mathbf{W}$, given from (3.21) with $k = B$ as:

$$\mu^{1/2} = \frac{tr(\mathbf{\Lambda}^{-1/2}\mathbf{W}^{1/2})}{tr(\mathbf{\Lambda}^{-1}) + p_0} \tag{3.37}$$

Substituting the expression for $\mathbf{W}$ from (3.36) in (3.37) and simplifying, we get $\gamma = p_0/tr(\mathbf{\Lambda}^{-1})$. Inserting this in (3.36) yields:

$$\mathbf{W}^{1/2} = \mu^{1/2}(\frac{p_0}{tr(\mathbf{\Lambda}^{-1})} + 1)\mathbf{\Lambda}^{-1/2} \tag{3.38}$$

Substituting (3.38) in the expressions for $\mathbf{\Phi_f}$ and $\mathbf{\Phi_g}$ in Theorem 1 and using the optimum precoder and decoder structure from Lemmas 1 and 2, we get the following design.

Theorem 3 *The optimum precoder and decoder that ensure equal errors for each* *data stream is given as:*

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f}; \;\; \mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}$$

$$\mathbf{\Phi_f} = \gamma^{1/2}\mathbf{\Lambda}^{-1/2} \tag{3.39}$$

$$\mathbf{\Phi_g} = \gamma^{1/2}\mathbf{\Lambda}^{-1/2}(\gamma + 1)^{-1} \tag{3.40}$$

$$\gamma = \frac{p_0}{tr(\mathbf{\Lambda}^{-1})} \tag{3.41}$$

We note that this solution was also obtained for the problem of maximizing output SNR, assuming a total transmit power constraint and assuming a zero forcing constraint on the transmitter ($\mathbf{GHF} = \beta\mathbf{I}$, for some $\beta > 0$) [48]. Some observations are now in order:

- From the power allocation policy, we see no subchannel is dropped, regardless of the channel realization. In other words, more power is used on weaker eigenmodes and less power on stronger eigenmodes to ensure that all streams have the same SNR. This is in contrast to the MMSE design, that dropped weaker eigenmodes to minimize the overall mean square error.

  We now illustrate the difference between the equal-error and MMSE designs, using Monte-Carlo simulations. Consider a fixed rate MIMO system with $M_T = M_R = 4$, over which we transmit $B = 3$ independent symbol streams each with 4 QAM modulation. Figure 3.8 shows the uncoded BER for each symbol stream.

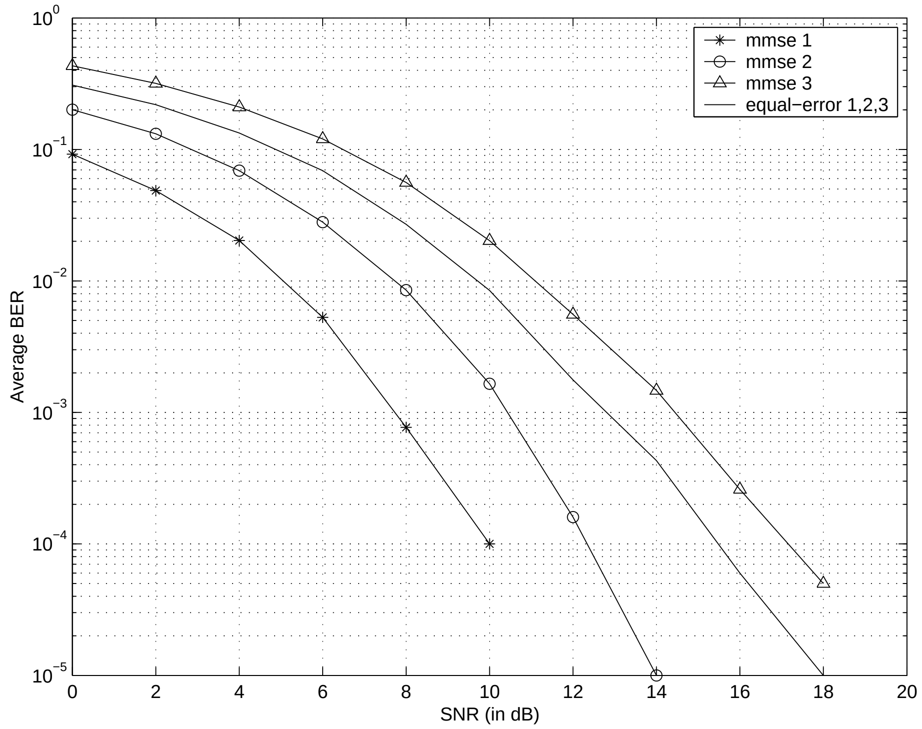

Figure 3.8: Comparison of sub-channel BERs: MMSE vs. Equal Error design

The equal-error design guarantees equal BER for all the symbol streams, unlike the MMSE design. In the latter case, some of the symbol streams have higher BER than the equal-error design and some have lower BER. Figure 3.9 shows the total average uncoded BER across all symbol streams for both the designs. The equal-error design performs worser at lower SNR since it tries to fight the channel and transmits all 3 symbol streams for all channel realizations. The MMSE design however, drops the weaker eigenmodes to improve the overall BER.

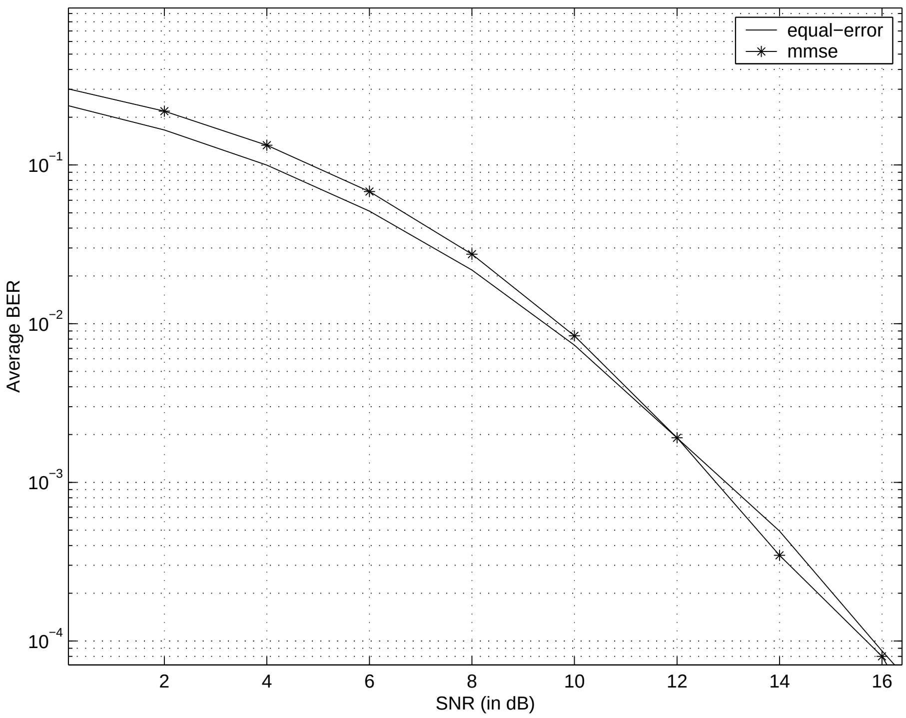

Figure 3.9: Comparison of total BER: MMSE vs. Equal Error design

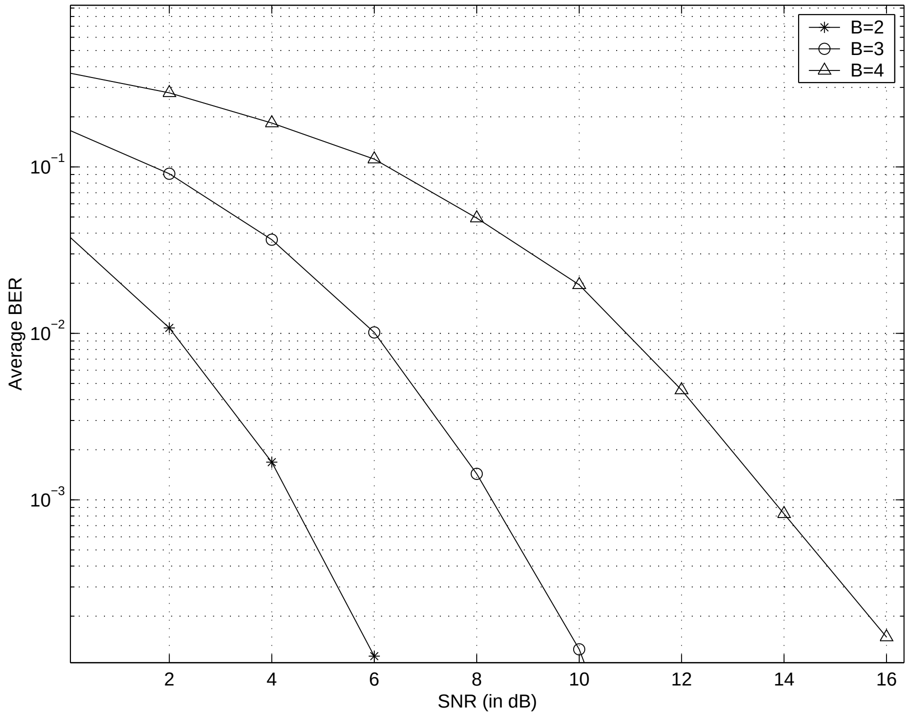

Figure 3.10: Performance of Equal Error design with varying B

- It can be verified that $\mathbf{GHF} = \frac{\gamma}{\gamma+1}\mathbf{I}$. Hence, the optimum design transforms the MIMO channel matrix into a scaled identity matrix. The precoder and decoder jointly invert the channel matrix to recover the transmitted signal.

- As B increases, more symbol streams are alloted to the weaker eigenmodes, which use up most of the power. This leads to a degradation in the bit error rate for all symbol streams. This is illustrated in Figure 3.10 that shows the performance degradation as the number of data streams (B) increase for a 5×5 MIMO system.

3.5.4 Max-SNR design

This design is typically motivated in a fading environment, where we want to deliver signal with the highest possible SNR at each instant.

Consider the system equation $\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn}$. Define the total output SNR as

$$J = Tr[(\mathbf{GHF})^*(\mathbf{G}\mathbf{R_{nn}}\mathbf{G}^*)^{-1}\mathbf{GHF}] \tag{3.42}$$

It is well-known that the optimum precoder that maximizes output SNR pours all power on the strongest eigenmode of the MIMO channel [78], i.e.,
$\mathbf{\Phi_f}^2 = diag([p_0, 0, 0..., 0])$.

Lemma 4 *The choice* $\mathbf{W} = diag([w_1, 0, .., 0])$*, where* $w_1 = \mu\lambda^{-1}(p_0\lambda_1 + 1)^2$ *in the weighted MMSE design (see Theorem 1) indeed results in all power being poured on the strongest eigenmode of the MIMO channel. Hence, the max-SNR design is just a special case of our generalized design.*

We note that this strategy allows only 1 input symbol stream to be transmitted. Figure 3.11 shows the performance gain possible in MIMO systems by using the max-SNR design. We have assumed an iid matrix channel, with total transmit power normalized to unity and $B = 1$ stream with 4-QAM modulation. The SNR is defined as the ratio of the total transmit power to the noise variance on each receive antenna. We have plotted the performance curves for a 1×1, 2×2, 3×3 and a rate 1 space-time code (Alamouti scheme). Firstly, we see that by increasing the number of antennas improves overall performance, for the same total transmit power. For example, for a 2×2 MIMO system employing the max-SNR design, we obtain 4th-order diversity gains. We also compare our scheme to the well-known Alamouti scheme [39], which is a rate 1 space-time code for a 2×2 system. Note that the Alamouti scheme assumes no channel knowledge at the transmitter, but perfect channel knowledge at the receiver. At an average BER of 10^{-3}, the max-SNR design obtains about 2.5 dB performance gain over Alamouti Scheme. The performance gain can be explained by the fact that the max-SNR design always transmits on the strongest eigenmode regardless of the channel realization. Our results have been obtained by averaging over 100000 Monte-Carlo simulation runs. For each simulation run, a new channel instantiation was used and the max-SNR design was recomputed.

3.5.5 QoS-based Design

In this section, we show how other useful designs can be obtained from the weighted MMSE design. We will focus on one such design, namely the QoS based design, which optimally transmits (and receives) multiple (independent) data streams, each with a certain target BER, coding gain and modulation constraints.

Assuming a square QAM constellation, the probability of symbol error (and hence

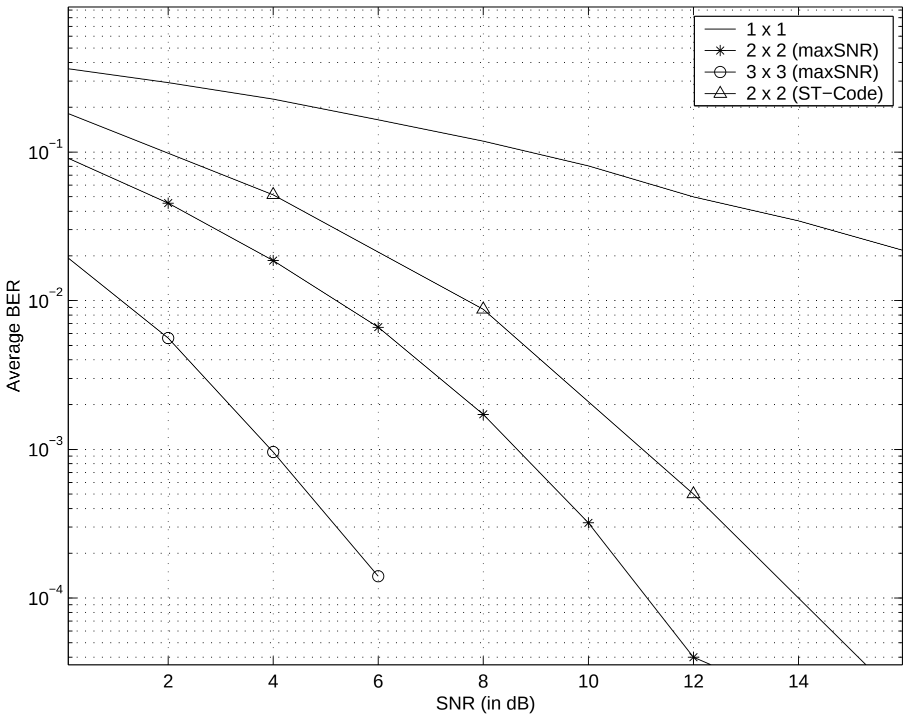

Figure 3.11: Precoding gain with Max-SNR design

the BER assuming Gray encoding) for the i-th subchannel is given as before by

$$P_{e,i} \simeq N_{e,i}\, Q(\sqrt{\beta_{M_i}\,\Gamma_i\,\gamma_{c,i}}) \tag{3.43}$$

and the SNR on the sub-channels is given as

$$\Gamma_i = \phi_{f,i}^2 \lambda_i = (w_i^{1/2} \mu^{-1/2} \lambda_i^{1/2} - 1)_+ \tag{3.44}$$

To satisfy the QoS constraint, we need $P_{e,i} \leq P_{e,i}^0$, where the latter is the target symbol error probability. Substituting (3.44) in (3.43), we get [7]:

$$w_i \geq \left[1 + \frac{M_i - 1}{3\gamma_{c,i}} \big[Q^{-1}(\frac{P_{e,i}^0}{N_{e,i}})\big]^2\right]^2 \frac{\mu}{\lambda_i} \tag{3.45}$$

Note that μ is a function of $\mathbf{W}$, given from (3.21) with $k = B$ as:

$$\mu^{1/2} = \frac{tr(\mathbf{\Lambda}^{-1/2}\mathbf{W}^{1/2})}{tr(\mathbf{\Lambda}^{-1}) + p_0} \tag{3.46}$$

This implies from (3.45) and (3.46) that we need

$$\sum_{i=1}^{B} \frac{M_i - 1}{3\gamma_{c,i}} \big[Q^{-1}(\frac{P_{e,i}^0}{N_{e,i}})\big]^2 \lambda_i^{-1} \leq p_0 \tag{3.47}$$

[7] We have assumed for simplicity that $\mathbf{\Phi_f} > \mathbf{0}$ and that $B = rank(\mathbf{H})$. The case when $B < rank(\mathbf{H})$ can be handled by using the argument in Remark 1.

to be satisfied. This inequality basically states there is enough power available at the transmitter to satisfy the target BER, coding and modulation constraints,which is true for any feasible communication link. Substituting (3.45) in Theorem 1 and using Lemmas 1 and 2, we get the following design.

Theorem 4 *The optimum precoder and decoder that provide any (feasible) set of target error probability, coding gain and modulation constraints, on each symbol stream is given as:*

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f};\;\; \mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}$$
$$\phi_{f,i}^2 \geq \lambda_i{}^{-1}\left[\frac{M_i - 1}{3\gamma_{c,i}}\left[Q^{-1}\left(\frac{P_{e,i}^0}{N_{e,i}}\right)\right]^2\right] \tag{3.48}$$

Note that $\mathbf{\Phi_g}$ can be any full-rank diagonal matrix, since it does not affect the above considered constraints.

We now consider a multimedia application, where we simultaneously need to transmit a voice and a data stream over a MIMO system. Typically, voice and data streams have different coding, modulation and BER requirements.

As an example, consider a 4×4 MIMO system over which we transmit 1 voice stream and 1 data stream, which are independent of each other. We assume the voice stream has the following system parameters: target BER of 10^{-3}, 4 QPSK modulation and a 4-state rate 1/2 convolutional code with 4 dB coding gain. Assume that the data stream has the following system parameters: target BER choice of $[10^{-3}, 10^{-5}, 10^{-7}]$, a modulation order choice of $[4, 16, 64]$ QAM, depending on the QoS requirement.

In Figure 3.12, we plot the ratio of powers allocated to each stream, for the above choice of parameters. We observe that as the transmission rate (for the data stream) is increased and/or the target BER (for the data stream) is decreased, more power is allocated to the data stream relative to the voice stream. For our simulations we have assumed uncorrelated noise at the receiver antennas. The MIMO channel ($\mathbf{H}$) is drawn from a unitary matrix distribution, and no delay spread is assumed for simplicity. The optimum precoder and decoder will pour power only on the strongest two eigenmodes of the MIMO channel, with the power allocation policy given by equation (3.48).

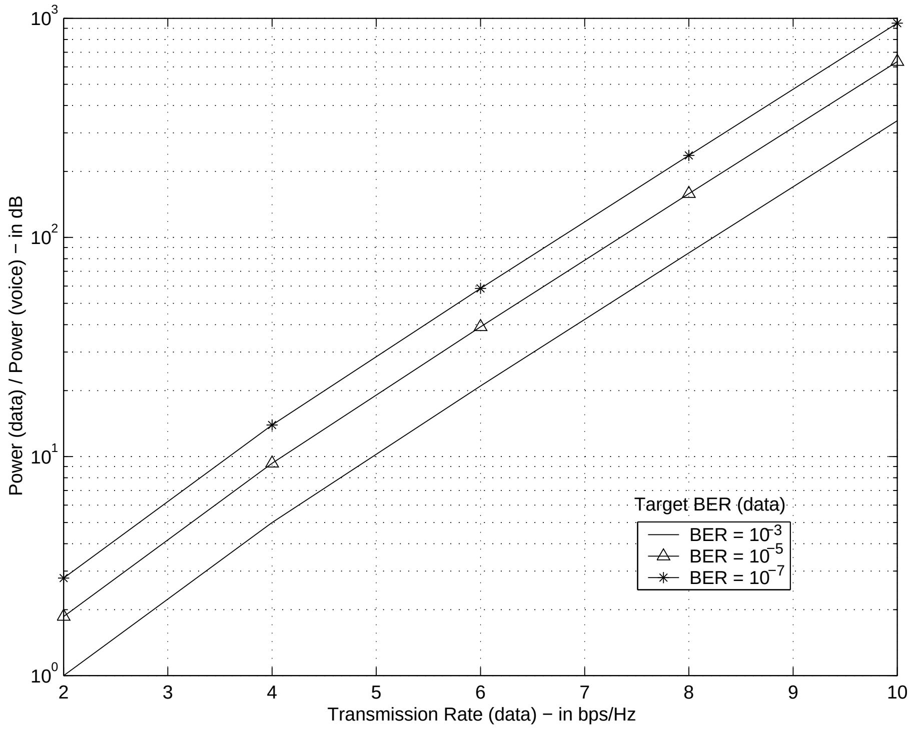

Figure 3.12: Simultaneous voice and data transmission

Chapter 4

LPD: New Constraints and Criteria

When the channel is fixed and known at the transmitter and receiver, linear precoding and decoding [1, 47, 48, 56] can be used to improve system performance. In Chapter 3, we presented a generalized linear precoder and decoder design based on the weighted MMSE criterion and a total transmit power constraint. We showed that for many applications, the optimal precoder and decoder diagonalize the MIMO channel into eigen subchannels. Only the power allocation across the eigenmodes vary depending upon the error weights.

In this chapter, we first consider the design of optimal space-time linear precoders and/or decoders for the following new criteria and constraints that were not considered in chapter 3.

- Criteria: Minimize the pairwise error probability (PEP) [15]

  Constraint: Total Transmit Power Constraint.

  (Note that the space-time codes of Tarokh et al [15] are designed based on *average* PEP, in which case the PEP is averaged over all channel realizations. We do not get into code construction, but rather optimize the precoder and decoder for a fixed and known channel realization.)

- Criteria: Minimize geometric mean square error [12, 13]

  Constraint: Total Transmit Power Constraint.

- Criteria: Maximize Information Rate [47]

  Constraint: Peak Power Constraint.

  In this problem, we find that the peak power constraint forces the optimum design to pour power on modes other the eigenmodes. As a result, eigenmode transmission is not optimum in this case.

In solving the above problems, we introduce several *novel mathematical tools* that can potentially be used to manipulate other MIMO optimization problems.

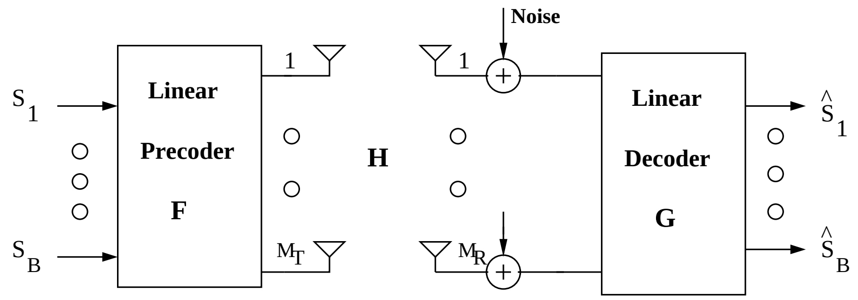

Figure 4.1: System model

4.1 Pairwise Error Probability (PEP) based design

For a MIMO channel with no delay spread, we have the following system equation (see Figure 4.1):

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn} \tag{4.1}$$

where $\mathbf{H}$ is an $(M_R \times M_T)$ channel matrix, $\hat{\mathbf{s}}$ is the $B \times 1$ received vector and $\mathbf{s}$ is the $B \times 1$ transmitted vector, where $B = rank(\mathbf{H}) \leq min(M_R, M_T)$ is the number of parallel streams of data that is to be transmitted; $\mathbf{n}$ is the $M_R \times 1$ noise vector (at the given symbol time); $\mathbf{G}$ is the $B \times M_R$ decoder matrix, and $\mathbf{F}$ is the $M_T \times B$ precoder matrix.

In this section, we minimize the target probability of error. Designs involving the probability of error depend on the modulation (symbol alphabet) and also on the detection rule. It is well-known that the optimal decision rule is the maximum likelihood detector, provided the noise is Gaussian and the symbols are i.i.d. If we let $\mathbf{s}(\mathcal{H}_k)$ denote the transmit symbol vector corresponding to the hypothesis $\mathcal{H}_k$, and let $\mathcal{D}$ denote the decision on the transmitted symbol vector, then the ML decision rule is [10]:

$$\mathcal{D} = arg \min_{\mathcal{H}_k} [\hat{\mathbf{s}} - \mathbf{GHFs}(\mathcal{H}_k)]^* (\mathbf{GR_{nn}G}^*)^{-1} [\hat{\mathbf{s}} - \mathbf{GHFs}(\mathcal{H}_k)] \tag{4.2}$$

Define the pairwise error probability (PEP) to be the error probability of choosing in favor of $\mathbf{s}(\mathcal{H}_l)$ instead of $\mathbf{s}(\mathcal{H}_k)$. Assuming ideal channel state information at the transmitter and receiver, the PEP is well-approximated by (see for example [15]):

$$P(\mathbf{s}(\mathcal{H}_k) \to \mathbf{s}(\mathcal{H}_l)) \leq e^{-d_{min}^2/2} \tag{4.3}$$

where

$$d_{min}^2 = \min_{m,n} [\mathbf{s}(\mathcal{H}_m) - \mathbf{s}(\mathcal{H}_n)]^* \mathbf{SNR}(\mathbf{F}, \mathbf{G}) [\mathbf{s}(\mathcal{H}_m) - \mathbf{s}(\mathcal{H}_n)] \tag{4.4}$$

and

$$\mathbf{SNR} = \mathbf{F}^* \mathbf{H}^* \mathbf{G}^* (\mathbf{GR_{nn}G}^*)^{-1} \mathbf{GHF} \tag{4.5}$$

The minimum distance (d_{min}) in (4.4) can be lower bounded as [14]:

$$d_{min}^2 \geq \lambda_{min}[\mathbf{SNR}(\mathbf{F},\mathbf{G})] \min_{m,n} ||\mathbf{s}(\mathcal{H}_m) - \mathbf{s}(\mathcal{H}_n)||^2 \tag{4.6}$$

Substituting (4.6) in (4.3), we get the PEP to be:

$$P(\mathbf{s}(\mathcal{H}_k) \rightarrow \mathbf{s}(\mathcal{H}_l)) \leq e^{-[\lambda_{min}(\mathbf{SNR}(\mathbf{F},\mathbf{G}))\ \min_{m,n} ||\mathbf{s}(\mathcal{H}_m) - \mathbf{s}(\mathcal{H}_n)||^2]} \tag{4.7}$$

The upper bound on the PEP can hence be minimized if we maximize $\lambda_{min}[\mathbf{SNR}(\mathbf{F},\mathbf{G})]$. This leads us to the following optimization problem.

$$\begin{aligned} &\max_{\mathbf{F},\mathbf{G}} \mathcal{J} = \lambda_{min}[\mathbf{SNR}(\mathbf{F},\mathbf{G})] \\ &subject\ to: \quad Tr(\mathbf{F}\mathbf{F}^*) = p_0 \end{aligned} \tag{4.8}$$

where $Tr(\mathbf{F}\mathbf{F}^*) = p_0$ is the total power constraint at the transmitter.

Let us define the eigen-value decomposition (EVD):

$$\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H} = (\mathbf{V}\ \ \tilde{\mathbf{V}}) \begin{pmatrix} \mathbf{\Lambda} & 0 \\ 0 & \tilde{\mathbf{\Lambda}} \end{pmatrix} (\mathbf{V}\ \ \tilde{\mathbf{V}})^* \tag{4.9}$$

where $\mathbf{V}$ is an $M_T \times B$ orthogonal matrix which forms a basis for the range space of $\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the B non-zero eigen-values arranged in a decreasing order from top-left to bottom-right; $\tilde{\mathbf{\Lambda}}$ contains the zero eigen-values; $\tilde{\mathbf{V}}$ is a $M_T \times (M_T - B)$ orthogonal matrix which constitutes a basis for the null-space of $\mathbf{H}^*{\mathbf{R_{nn}}}^{-1}\mathbf{H}$.

Lemma 5 *The solution to the optimization problem posed in (4.8) is given as:*

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f}$$

$$\mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}^{-1}}$$

$$|\phi_{f,i}|^2 = \frac{p_0}{\sum_{k=1}^{B} \lambda_k^{-1}} \lambda_k^{-1} \tag{4.10}$$

and where $\mathbf{\Phi_g}$ *can be any full-rank* $B \times B$ *matrix and* $\mathbf{\Phi_f}$ *is a* $B \times B$ *diagonal matrix with diagonal entries given by* $\phi_{f,i}$.

Proof: See Appendix A.4 ■

It can be verified that $\mathbf{GHF} = \mathbf{\Phi_g}\mathbf{\Lambda}\mathbf{\Phi_f}$ which is diagonal. Hence, the optimum design diagonalizes the MIMO channel into eigen subchannels as suggested by the generalized design in Chapter 3. The power allocation is such that more power is allocated to weaker subchannels and less power to stronger subchannels, to equalize the subchannel SNRs. The solution can be explained as follows. The current design minimizes the worse-case PEP (and hence SNR) across all eigenmodes. The minimum is attained if and only if the SNRs on all eigenmodes are forced equal. Indeed, substituting the expressions for $\mathbf{G}$ and $\mathbf{F}$ from Lemma 5 in equation (4.5), it can be verified that $\mathbf{SNR} = \frac{p_0}{Tr(\Lambda)}\mathbf{I}$, implying that the SNRs on all subchannels are forced to be equal. Note that the present design is identical to the *equal-error* design can be obtained by choosing $\mathbf{W} = \mathbf{\Lambda}^{-1}$ in the weighted MMSE design.

4.2 Mean square error based design

In Chapter 3, we designed the precoder and decoder to minimize any weighted sum of mean square errors across the transmitted signal vector elements. We showed that the choice of error weight $\mathbf{W} = \mathbf{\Lambda}$ resulted in the information rate maximizing design. In this section, we show that minimizing the the geometric mean square error defined as the determinant of the error covariance matrix (see for example [12, 13]) also leads to the information rate maximizing design.

We consider again the system equation (4.1). Following the discussion in Chapter 3, the output symbol vector error covariance matrix is given as:

$$\mathbf{R_e}(\mathbf{G}, \mathbf{F}) = (\mathbf{GHF} - \mathbf{I})(\mathbf{GHF} - \mathbf{I})^* + \mathbf{G}\mathbf{R_{nn}}\mathbf{G}^* \tag{4.11}$$

Minimizing the determinant of the error covariance matrix, leads to the following optimization problem:

$$\begin{aligned} \min_{\mathbf{F},\mathbf{G}} J = \; & det(\mathbf{R_e}\,(\mathbf{F},\mathbf{G})) \\ subject\ to: \quad & Tr(\mathbf{F}\mathbf{F}^*) = p_0 \end{aligned} \tag{4.12}$$

Lemma 6 *The solution to the optimization problem posed in (4.12) is given as:*

$$\begin{aligned} \mathbf{F} &= \mathbf{V}\mathbf{\Phi_f} \\ \mathbf{G} &= \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} \\ \mathbf{\Phi_f}^2 &= \left(\mu\mathbf{I} - \mathbf{\Lambda}^{-1}\right)_+ \end{aligned} \tag{4.13}$$

where $\mathbf{\Phi_f}$ and $\mathbf{\Phi_g}$ are $B \times B$ diagonal matrices, with $\mathbf{\Phi_g}$ invertible; $(.)_+$ means that only non-negative values are allowed and μ is computed from the transmit power constraint.

Proof: See Appendix A.5 ■

Note that the optimum power allocation policy is the well-known *waterfilling* policy. Hence, the minimization of the determinant of the $\mathbf{R_e}$ yields the same design as the *maximum information rate* design described in chapter 3.[1] This result makes sense since the information rate can be written as the logarithm of determinant of the error covariance matrix, as shown below.

[1]This result was also noted in [12] albeit using a different proof and in a different context.

The mutual information between $\hat{\mathbf{s}}$ and $\mathbf{s}$ is given as [47]:

$$I(\hat{\mathbf{s}};\mathbf{s}) = log_2(\, det[I + (\mathbf{GHF})^*(\mathbf{G}\mathbf{R_{nn}}\mathbf{G}^*)^{-1}(\mathbf{GHF})]\,) \tag{4.14}$$

Substituting the expression for $\mathbf{G}$ from Lemma 6 in (4.14), it can be verified that the mutual information does not depend on $\mathbf{G}$, i.e., $I(\hat{\mathbf{s}};\mathbf{s}) = log_2(\, det[I+(\mathbf{HF})^*\mathbf{R_{nn}^{-*}}\mathbf{HF}]\,) = log_2(\, det(\mathbf{R_e}^{-1})\,)$ where the last equation was obtained from (A.34) in Appendix A.5. Hence, minimizing the determinant of $\mathbf{R_e}$ is equivalent to maximizing mutual information between $\mathbf{s}$ and $\hat{\mathbf{s}}$.

It is interesting to note that the minimizing the trace and determinant of the mean square error results in different power allocation policies, namely the *inverse waterfilling* and *waterfilling* policies, respectively. As shown in Chapter 3, these power policies are similar in that they drop the weak eigenmodes. However, the inverse waterfilling policy redistributes the power so that more power is allocated to the weaker modes and vice-versa. This is in contrast to the waterfilling policy that allocates more power to stronger eigenmodes and less power to weaker eigenmodes.

4.3 Designs with Peak Power Constraint

So far, we have considered designs with a total transmit power constraint across the antennas, which may result in some antennas transmitting more power than

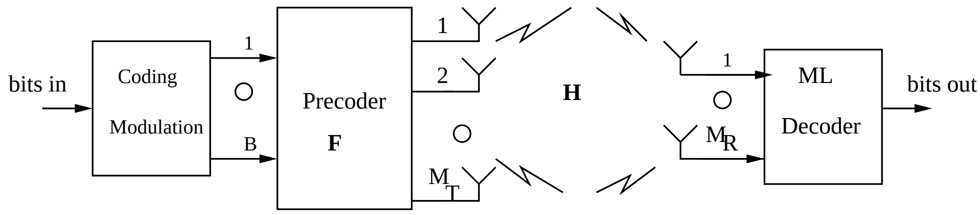

Figure 4.2: System model

the others. This can saturate the RF circuitry and cause non-linear distortions. Introducing an expected peak power constraint on each transmit antenna minimizes these distortions.

Consider the following system equation (see Figure 4.2)

$$\mathbf{x} = \mathbf{HFs} + \mathbf{n} \tag{4.15}$$

where $\mathbf{H}$ is the $M_R \times M_T$ channel matrix, $\mathbf{F}$ is the $M_T \times B$ precoding matrix, $\mathbf{s}$ is the $B \times 1$ transmit signal and $\mathbf{x}$ is the $M_R \times 1$ received signal. We assume white noise, i.e., $\mathbf{R_{nn}} = \sigma^2\mathbf{I}$ for simplicity of analysis, and a maximum likelihood receiver.

Denote the signal vector that is launched from the transmit antennas is given as $\mathbf{y} = \mathbf{Fs}$. The average power constraint on antenna i requires that $E(y_i^2) = E(\,|\mathbf{Fs}|_i^2\,) \leq p$, for $i = 1, 2, ...M_T$. Assuming $E(\mathbf{ss}^*) = \mathbf{I}$, we get the power constraint as: $(\mathbf{FF}^*)_{ii} \leq p$. Note that the average transmit signal power is normalized to unity, i.e., $E(s_j^2) = 1$, for $j = 1, 2, ..B$, and hence the peak power constraint is really an expected peak power constraint.

Our goal is to find the optimum precoder ($\mathbf{F}$) that maximizes information rate, subject to the expected peak power constraint. The analysis for other criteria such as MMSE criteria subject to the expected peak power constraint is on similar lines and will not be dealt with here.

The mutual information between $\mathbf{x}$ and $\mathbf{s}$, assuming a Gaussian $\mathbf{n}$, is maximized according to:

$$\begin{aligned} &\max_{\mathbf{F}} I = log_2\, det\big(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{HFF}^*\mathbf{H}^*\big) \\ subject\,to: \quad &(\mathbf{FF}^*)_{ii} \leq p \quad i = 1, 2, .., M_T \end{aligned} \tag{4.16}$$

The cost function (I) to be maximized is concave in $\mathbf{FF}^*$ and the M_T power constraints are affine in $\mathbf{FF}^*$. Hence, this is a convex optimization problem for which a unique solution exists.

For LPD optimization problems subject to the total power constraint, the optimum transmission was shown to be on the eigenmodes of $\mathbf{HH}^*$. This decoupled the channel into parallel independent subchannels and we were able to provide an analytical solution. However, with the peak power constraint, we were unable to prove that eigenmode transmission is optimum. As a result, deriving an analytically closed form solution was deemed prohibitive.

We solve the optimization problem (4.16) using numerical techniques. Specifically, we used a semi-definite programming tool box, SDPSOL, [79, 80] that uses an interior-point algorithm [81] to find a numerically unique solution.

We report the following observations from the numerical solution:

- The optimum precoder satisfies all the transmit power *strictly* $(\mathbf{FF}^*)_{ii} = p$, for $i = 1, 2..., M_T$. In other words, all constraints are forced active.

- The optimum transmission was found to be not necessarily on the eigenmodes of $\mathbf{HH}^*$.

- The SDPSOL tool box yielded a unique solution for $\mathbf{FF}^*$. Taking the EVD, we have $\mathbf{FF}^* = \mathbf{\Theta\Phi_f^2\Theta}^*$, where $\mathbf{\Theta}$ is any $M_T \times B$ orthogonal matrix, that is not necessarily equal to $\mathbf{V}$ and $\mathbf{\Phi_f}^2$ is a diagonal matrix of powers allocated to the modes of $\mathbf{\Theta}$. This implies that the optimum $\mathbf{F} = \mathbf{\Theta\Phi_f\Theta_2}$ where $\mathbf{\Theta_2}$ is any $B \times B$ orthogonal matrix. It can be verified that we can choose $\mathbf{\Theta_2} = \mathbf{I}$ without any loss of generality and not affect $\mathbf{FF}^*$. Hence, for the purposes for system implementation, the optimum precoder can be chosen without any loss of generality as:

$$\mathbf{F} = \mathbf{\Theta\Phi_f} \tag{4.17}$$

We were unable to find an analytical solution for the optimum $\mathbf{\Theta}$ and $\mathbf{\Phi_f}$. In practice, they can be computed using efficient numerical techniques such as the one described in [81].

The main point to note that is the eigenmode transmission and channel diagonalization need not be the optimum strategy with the peak power constraint.

We now compare the maximum information rate design subject to the total and peak power constraints, respectively. Figures 4.3 and 4.4 plot the peak to average ratio (PAR), and capacity, obtained from both the designs, for a 3×3 spatial multiplexing system. PAR is defined as the ratio of the maximum transmit power across all transmit antennas to the average power (averaged across all transmit antennas). We assume a flat-fading iid channel matrix,white noise across receive antennas and the total power across all transmit antennas normalized to unity. Our results are obtained by averaging over 1000 channel realizations. For each channel realization, the optimum precoder, as well as the resulting capacity and PAR were recomputed.

From Figure 4.3 We note that the LPD design with the peak power constraint ensures a PAR of unity, unlike the design with total power constraint. In the latter case, the PAR decreases with increasing SNR since the waterfilling solution pours near equal power on all eigenmodes at high SNRs. This maps to a more uniform power distribution across the transmit antennas. In low SNRs, eigenmodes can carry different powers leading to a large PAR across the transmit antennas. From Figure 4.4, we note that the LPD design with peak power constraint leads to a small loss in capacity, when compared to the design with total power constraint. This can be

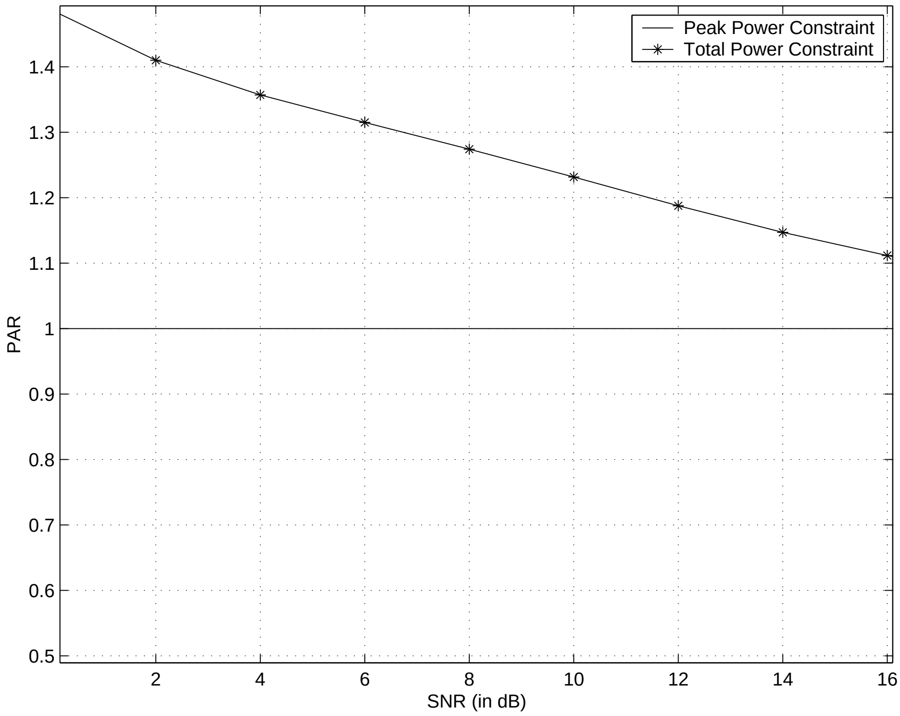

Figure 4.3: PAR: Peak power vs. Total power constraint

explained by the fact that the optimum transmission occurs on slightly perturbed versions of the channel eigenmodes.

An alternate strategy to (loosely) bound the peak power was considered in [6]. In the latter case, the maximum eigenvalue of the precoder matrix was constrained and an analytical closed form solution was obtained. The optimum precoder poured equal power on all eigenmodes of the channel regardless of the optimization criteria and channel realization. However, the solution did not guarantee that the antennas were transmitting at the peak power.

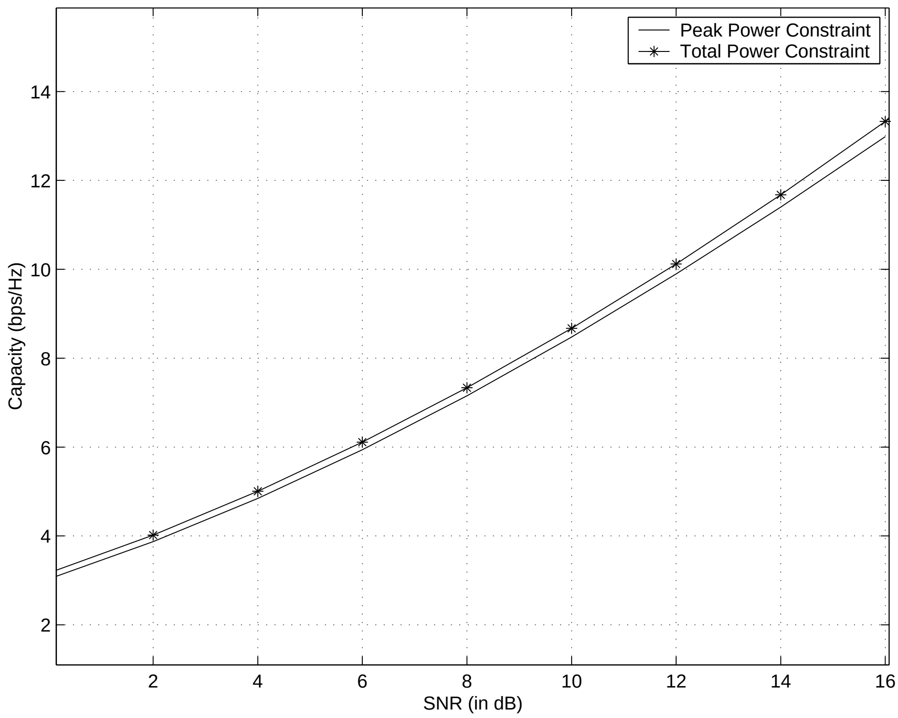

Figure 4.4: Capacity: Peak power vs. Total power constraint

Chapter 5

Linear Precoding for Delay Spread Channels

In this Chapter, we show how to extend the linear precoder and decoder framework to delay spread channels and multicarrier systems employing OFDM modulation. In the former case, we employ block transmission with guard symbols inserted between data blocks to prevent inter-block interference. In OFDM modulation, redundancy is added in the form of a cyclic prefix to handle delay spread.

We then present a novel Finite Impulse Response (FIR) precoder structure to pre-equalize the MIMO channel with delay spread assuming a zero-forcing constraint and subject to a transmit power constraint. FIR precoders are different from block precoders in that they operate on a continuous stream of data and do not require guard symbols to be transmitted.

5.1 Linear Block Precoding

In this section, we present space-time and space-frequency block precoders and decoders. The former operates in the time domain and can be used for single carrier systems. The latter operates in the frequency domain and can be used for multicarrier systems employing OFDM modulation.

5.1.1 Linear Space-Time Block Precoding and Decoding

For a MIMO channel with finite delay-spread that spans L symbol periods, we can write the system equation for a finite number of samples in *block form*. Such a design has also been considered in [47, 48] for the information rate maximizing criteria and the MMSE criteria. We now describe this technique assuming a generalized precoder and decoder framework. A natural result is that such a block precoder and decoder can in fact be optimized for any of the criteria or constraints that have been discussed in Chapters 3 and 4.

Consider a data *block* with Q input symbol vectors (in time succession), denoted as $\mathcal{S}_{ij}$, where i is the symbol time index within a block and j is the block index (see Figure 5.1). Following the discussion in Chapter 2, we stack Q input symbol vectors, $\mathcal{S}_{1j}, \mathcal{S}_{2j}, ..., \mathcal{S}_{Qj}$ in a $(QB \times 1)$ vector $\mathcal{S}_j = vec([\mathcal{S}_{1j} \dots \mathcal{S}_{jQ}])$. This vector is processed by a $QM_T \times QB$ precoder matrix $\mathcal{F}$ that outputs QM_T symbols. These symbols are passed through a parallel to serial (P/S) converter to get Q vectors of size $M_T \times 1$ each. These vectors are then launched into the MIMO channel in time succession, at the end of which $L-1$ zero guard symbol vectors are transmitted to prevent inter-block interference.

The system equation for the jth data block is given as:

$$\hat{\mathcal{S}}_j = \mathcal{G}\mathcal{H}\mathcal{F}\mathcal{S}_j + \mathcal{G}\mathcal{N}_j \tag{5.1}$$

where $\mathcal{G}$ is a $QB \times (Q+L-1)M_R$ space-time decoder block matrix that outputs an estimate $\hat{\mathcal{S}}_j = vec([\hat{\mathcal{S}}_{1j} \ldots \hat{\mathcal{S}}_{jQ}])$ of the transmitted symbol vectors, and $\mathcal{H}$ is given by:

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \mathcal{H}_L & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \mathcal{H}_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathcal{H}_L \end{pmatrix}_{(Q+L-1)M_R \times QM_T} \tag{5.2}$$

where $\mathcal{H}_j$, $1 \leq j \leq L$, denotes the j-th tap of the MIMO channel impulse response matrix and where we have assumed that the channel is stationary over the input block

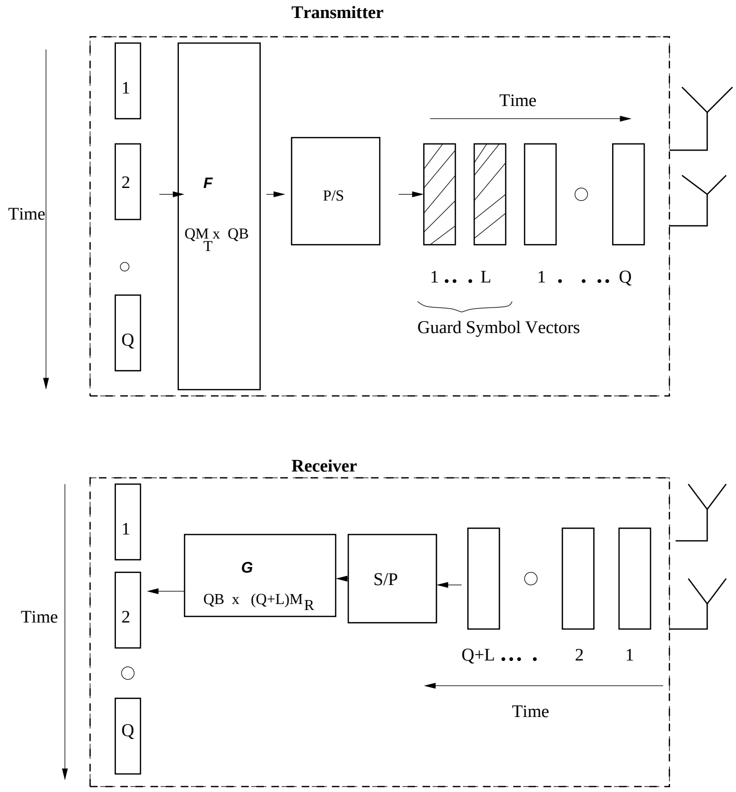

Figure 5.1: Block precoding and decoding for delay spread channels

length.

Note that $\mathcal{F}$ and $\mathcal{G}$ need not be Toeplitz matrices i.e., we are not imposing any structure on $\mathcal{F}$ and $\mathcal{G}$. Also, $\mathcal{F}$ and $\mathcal{G}$ operate across space and time and can be thought of as a *spatio-temporal channel dependent block coder and decoder*, respectively that operates in the complex field.

At the receiver, the $Q + L - 1$ snapshot vectors of size $M_R \times 1$ each are passed through a serial-to-parallel (S/P) converter to get a $(Q + L - 1)M_R \times 1$ vector. The latter is then processed by the decoder $\mathcal{G}$ to yield the output $\hat{\mathcal{S}}_j$.

Making the change of variables: $\hat{\mathcal{S}}_j \rightarrow \hat{\mathbf{s}}$, $\mathcal{S}_j \rightarrow \mathbf{s}$, $\mathcal{N}_j \rightarrow \mathbf{n}$, $\mathcal{F} \rightarrow \mathbf{F}$,$\mathcal{G} \rightarrow \mathbf{G}$ and $\mathcal{H} \rightarrow \mathbf{H}$, we get

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn} \tag{5.3}$$

which has the same form as the system equation for a flat-fading channel.

In the above design, the optimal choice of Q is dependent on a number of system requirements. A large value of Q is more data rate efficient since the redundancy-factor $(L-1)/Q$ is lower. The drawback of high Q is the increased implementation complexity, since the precoder and decoder now have to process a lot of data symbols.

Figures 5.2 and 5.3 illustrate the performance of an MMSE linear precoder-decoder system assuming $M_T = M_R = 3$ over which we want to transmit $B = 2$ parallel streams of data, each with 4 QAM modulation. We assume the total power constraint, where the total power across a block of $M_T \times Q$ symbols is now constrained to Q. The noise across the $M_R \times Q$ received symbols is white. We assume an L tap MIMO channel, where each tap is delayed by a symbol period and is an iid matrix with elements having a variance of $1/L$.

From Figure 5.2, we see that for $Q = 5$, increasing the number of channel taps from $L = 1$ to $L = 4$ improves performance. This can be explained by the additional path diversity gains obtained by the precoder and decoder as L increases. From Figure 5.3, we see that for a $L = 2$ tap channel, increasing Q (for $Q > L$) from $Q = 5$

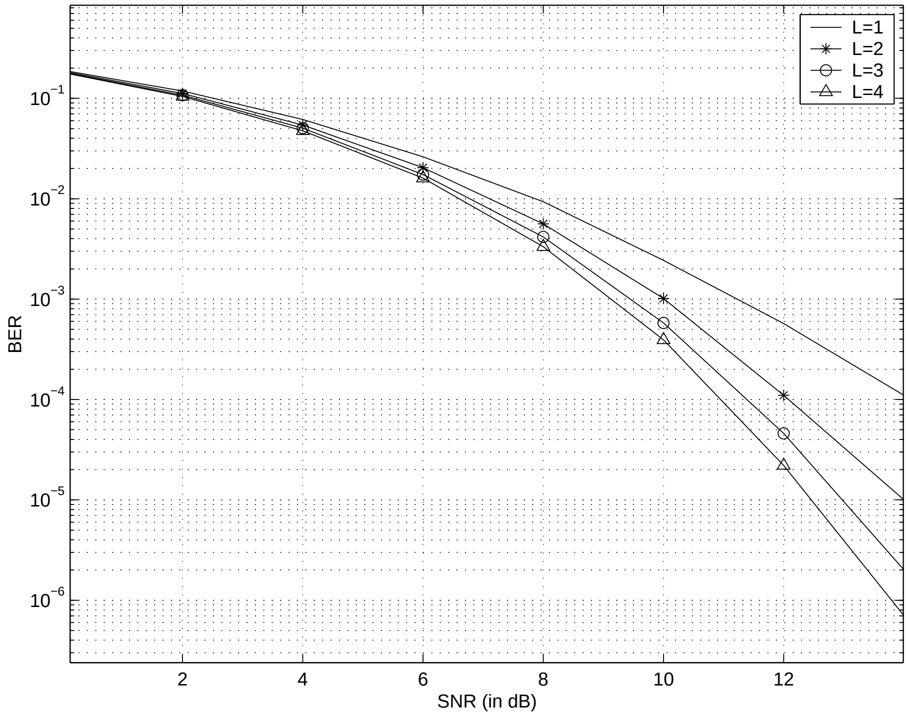

Figure 5.2: Path diversity gain with Space-time block precoder and decoder

to $Q = 15$ does not affect system performance. Hence, the only benefit of increasing Q is improving rate efficiency, at the expense of implementation complexity.

5.1.2 Linear OFDM Precoding and Decoding

Recently, OFDM (Orthogonal Frequency Division Multiplexing) or DMMT (Discrete Matrix Multi Tone) modulation has received considerable attention for MIMO transmission systems to combat delay spread [49, 70, 71].

In such a scheme, the total frequency bandwidth is divided into Q frequency bins. The jth transmit signal block consists of Q symbol vectors, $\mathcal{S}_{ij}$, $i = 1, ...Q$, each

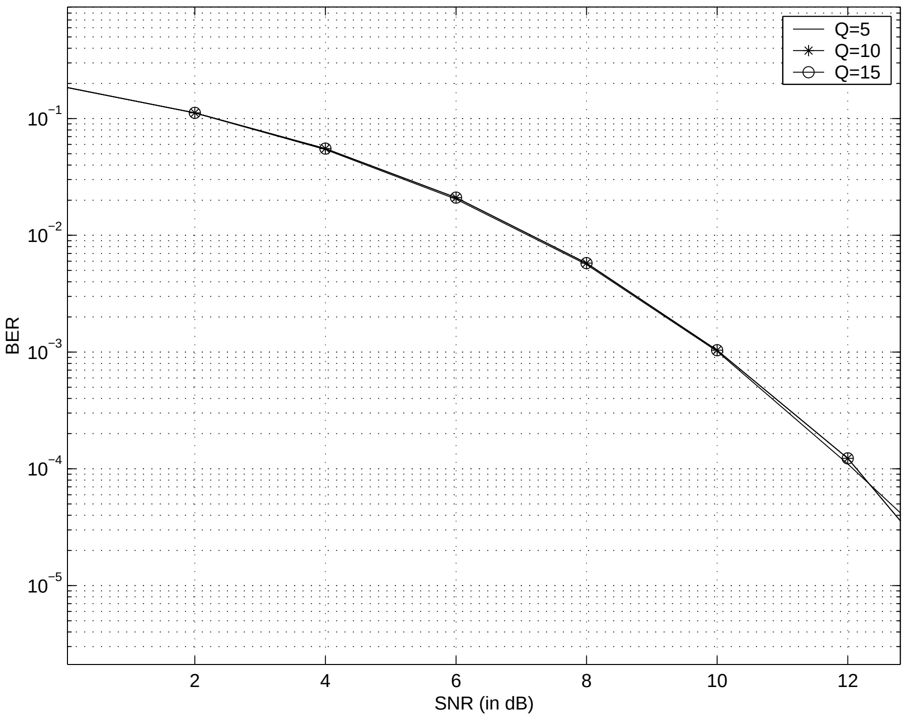

Figure 5.3: Performance of space-time block precoder and decoder vs. Block size(Q)

transmitted over the frequency bin i. The $B \times 1$ symbol vector $\mathcal{S}_{ij}$ at frequency bin i is processed by an $M_T \times B$ size precoder matrix $\tilde{\mathcal{F}}_i$. Following the discussion on OFDM modulation in Chapter 2, it can be shown that the frequency bins are indeed orthogonal to each other, i.e., the symbol stream transmitted in disjoint frequency bins do not interfere with each other. Hence, the precoder output $\tilde{\mathcal{F}}_i \mathcal{S}_{ij}$ at frequency bin i only sees the $M_R \times M_T$ MIMO channel transfer function $\tilde{\mathbf{H}}_i$. The $M_R \times 1$ received vector at frequency bin i is then processed by an $B \times M_R$ matrix decoder $\mathcal{G}_i$ to get $\hat{\mathcal{S}}_{ij}$.

One can thus write down the following system equation for the receiver output for the jth block in the frequency domain.

$$\hat{\mathcal{S}}_j = \tilde{\mathcal{G}} \tilde{\mathbf{H}} \tilde{\mathcal{F}} \mathcal{S}_j + \tilde{\mathcal{G}} \tilde{N}_j$$

where $\tilde{\mathbf{H}} = diag(\tilde{\mathbf{H}}_1, ..., \tilde{\mathbf{H}}_Q)$ is the $QM_R \times QM_T$ block diagonal matrix made from the MIMO transfer function values at the frequency bins $(1, 2, ..., Q)$; $\tilde{\mathcal{F}} = diag(\tilde{\mathcal{F}}_1, .., \tilde{\mathcal{F}}_Q)$ is the block diagonal $QM_T \times QB$ precoder and similarly for $\tilde{\mathcal{G}}$.

Making the change of variables: $\hat{\mathcal{S}}_j \rightarrow \hat{\mathbf{s}}$, $\mathcal{S}_j \rightarrow \mathbf{s}$, $\mathcal{N}_j \rightarrow \mathbf{n}$, $\tilde{\mathcal{F}} \rightarrow \mathbf{F}$,$\tilde{\mathcal{G}} \rightarrow \mathbf{G}$ and $\tilde{\mathbf{H}} \rightarrow \mathbf{H}$, we get

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn} \tag{5.4}$$

which has the same form as the system equation for a flat-fading channel. Hence, the LPD designs developed in Chapters 3 and 4 can be naturally extended for the OFDM case.

Note that the OFDM transmission incurs a loss in bandwidth due to the addition of the cyclic prefix. However, one advantage of using the OFDM scheme is the low implementation complexity due to the block diagonal structure of the precoder, decoder and the channel matrix. As a result, $\mathcal{F}_i$ and $\mathcal{G}_i$ can be computed by using only the eigenvalue decomposition (EVD) of $\tilde{\mathbf{H}}_i$, the latter being only a size $M_R \times M_T$ matrix. Since there are Q frequency bins, we need to compute Q such EVDs. In contrast, the time-domain linear block precoder and decoder developed in the earlier section need to be computed by using an EVD of $\mathcal{H}$ which is a size $(Q+L-1)M_R \times QM_T$ matrix and hence computationally more intense.

The optimum choice of Q is a system design parameter. The number of frequency bins (Q) should be large enough to accurately sample the channel response in the frequency domain. A small Q leads to undersampling the frequency channel response, leading to deterioration in system performance. On the other hand, a large Q increases system complexity due to the size of the FFT and IFFT. Furthermore, a large Q causes each frequency bin to span a very small frequency bandwidth. This makes the carriers on the frequency bins more susceptible to local oscillator (LO) drifts, RF phase noise, Doppler spread and intercarrier distortion.

We now illustrate an MMSE space-frequency precoder and decoder for a $M_T = M_R = 3$ spatial multiplexing system using $Q = 64$ frequency bins, over which we transmit $B = 128$ parallel symbol streams, each with 4 QAM modulation. We again assume an L tap MIMO channel, where each tap is delayed by a symbol period and is an iid matrix with elements having a variance of $1/L$. The total transmit power across all tones and antennas is normalized to Q. We assume white noise across all receiver antennas and tones.

It is seen from Figure 5.4 that increasing L increases system performance, since the MMSE design is able to extract frequency diversity available in the channel. The optimum precoder and decoder diagonalizes the MIMO channel at each frequency bin and creates a total of $min(M_R, M_T) \times B$ subchannels. It then transmits the B data streams on the strongest space-frequency subchannels, according to the inverse waterfilling solution.

5.2 FIR Precoding

In this section, we will present a novel Finite Impulse Response (FIR) precoder structure to pre-equalize the MIMO channel with delay spread assuming a zero-forcing constraint and subject to a transmit power constraint. FIR precoding is different from block precoding considered above in that it uses an FIR filter structure that operates continuously on the input symbol vector stream and requires no guard symbol

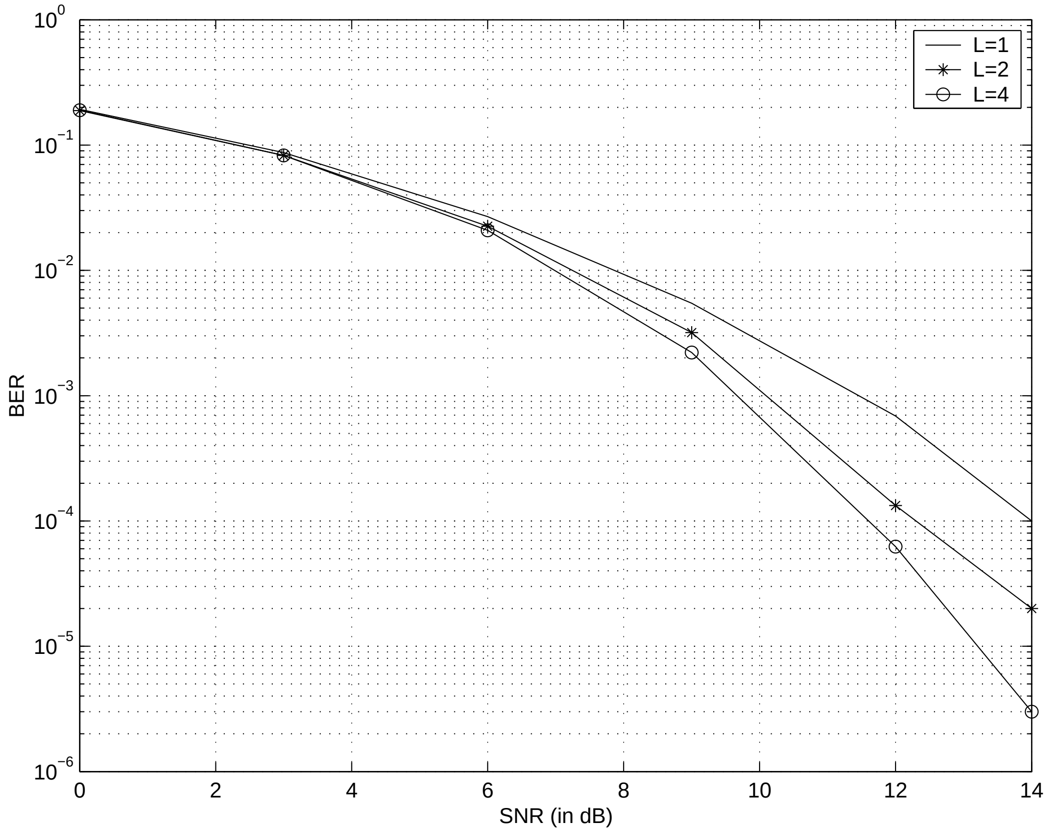

Figure 5.4: Frequency diversity gain with space-frequency MMSE precoder and decoder

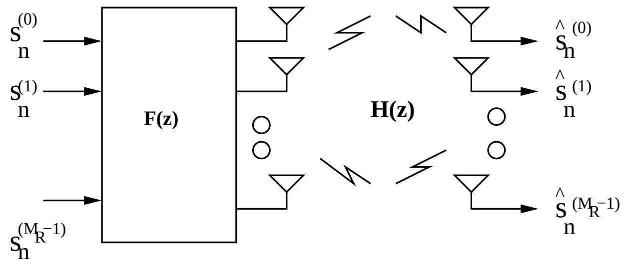

Figure 5.5: FIR precoding for delay spread channels

transmission.

To date, FIR precoding assuming perfect channel knowledge has not been considered in the literature. The solutions for optimum FIR precoders that optimize criteria such as information rate maximizing criteria, MMSE, SNR maximizing, and zero-forcing criteria are open research problems. It appears to be difficult to obtain closed form solutions to the precoders, unless we assume some a priori structure on the channel or the precoder.

In this section we exploit one such structure on MIMO wireless channels, which is typically found in the downlink transmission, to simplify our optimization problem.

5.2.1 System and Channel Model

Consider a downlink wireless MIMO system (see Fig. 5.5) with $M_T \geq M_R$ receive antennas. We consider transmission of M_R symbol streams $s_n^{(i)}$ $(i = 0, ..., M_R - 1)$ which are pre-processed by the $M_T \times M_R$ space-time FIR precoder (filter) [1] $\mathbf{F}(n) =$

[1]We assume that the precoding filter is anti-causal in order to simplify the presentation in the subsequent developments

$\sum_{l=-L'+1}^{0} \mathbf{F}_l\, \delta(n-l)$, where $n = ...-2,-1,0,1,2,....$ is the discrete time index and the $M_T \times M_R$ matrix $\mathbf{F}_l\,(l = -L'+1, -L'+2, ...1)$ denotes the l-th tap of the precoding filter. Taking the z-transform, we have $\mathbf{F}(z) = \sum_{l=-L'+1}^{0} \mathbf{F}_l\, z^{-l}$ (see Fig 5.5).

The space-time filtered signal is then launched into the $M_R \times M_T$ MIMO channel

$$\mathbf{H}(z) = \sum_{l=0}^{L-1} \mathbf{H}_l\, z^{-l}.$$

The $M_R \times 1$ received vector signal can now be written as

$$\hat{\mathbf{s}}_n = \sum_{k=-\infty}^{\infty} \mathbf{H}_{n-k} \sum_{p=-\infty}^{\infty} \mathbf{F}_{k-p} \mathbf{s}_p + \mathbf{v}_n \tag{5.5}$$

where $\mathbf{v}_n$ is the noise vector sequence, $\mathbf{s}_n = vec[s_n^{(0)}\ \ s_n^{(1)}\ \ ..s_n^{(M_R-1)}]$ is the $M_R \times 1$ transmitted vector sequence and $\hat{\mathbf{s}}$ is defined similarly.

In the z-transform domain, (5.5) can be rewritten as

$$\hat{\mathbf{s}}(z) = \mathbf{H}(z)\mathbf{F}(z)\mathbf{s}(z) + \mathbf{v}(z) \tag{5.6}$$

with $\mathbf{H}(z) = \sum_{l=0}^{L-1} \mathbf{H}_l\, z^{-l}$, $\mathbf{F}(z) = \sum_{l=-L'+1}^{0} \mathbf{F}_l\, z^{-l}$, and $\mathbf{s}(z), \hat{\mathbf{s}}(z)$ and $\mathbf{v}(z)$ defined similarly.

From Chapter 2, the transfer function of the MIMO wireless channel is:

$$\mathbf{H}(z) = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{R,l}\, \mathbf{a}_{T,l}^{*}\, z^{-l} = \sum_{l=1}^{L} \mathbf{H}_l\, z^{-l}, \tag{5.7}$$

where α_l denotes the complex path gain of the l-th path, $\mathbf{a}_{R,l}$ is the array response vector at the receiver due to the l-th multipath that makes an angle $\theta_{R,l}$ with the receiver array and similarly for $\mathbf{a}_{T,l}$ at the transmitter; and $\mathbf{H}_l = \alpha_l \mathbf{a}_{R,l}\, \mathbf{a}_{T,l}^{*}$ is the l-th tap of the MIMO channel. It can be verified that for a sufficiently large number of transmit antennas, assuming that the angles of departure $\theta_{T,l}$ at the transmitter are fairly well separated, we have the following:

$$\mathbf{a}_{T,l}^{*}\mathbf{a}_{T,l'} \ll \mathbf{a}_{T,l}^{*}\mathbf{a}_{T,l} \quad for \quad l \neq l'. \tag{5.8}$$

A natural consequence is that

$$\mathbf{H}_l\mathbf{H}_{l'}^{*} \ll \mathbf{H}_l\mathbf{H}_l^{*} \quad for \quad l \neq l'. \tag{5.9}$$

5.2.2 Problem Formulation and Solution

Assuming that the transmitter has full channel knowledge, we derive an analytic expression for the taps of the zero-forcing precoder (pre-equalizer) satisfying

$$\mathbf{H}(z)\mathbf{F}(z) = \gamma\,\mathbf{I}_{M_R}, \tag{5.10}$$

where $\gamma > 0$ is a (real) constant chosen such that the power constraint $\mathrm{Tr}\left\{\sum_{l=-L'+1}^{0} \mathbf{F}_l \mathbf{F}_l^H\right\} = 1$ is satisfied. Now, for $M_T = M_R$, assuming that $\mathbf{H}(z)$ is invertible the unique solution of (5.10) is given by $\mathbf{F}(z) = \gamma\,\mathbf{H}^{-1}(z) = \gamma\,\frac{adj\mathbf{H}(z)}{det\mathbf{H}(z)}$ where *adj* denotes the adjoint matrix and *det* stands for the determinant of a matrix. If $M_T > M_R$ there is in principle an infinite number of solutions. We shall, however, restrict our attention to the pseudo-inverse, which is given by

$$\mathbf{F}(z) = \gamma\,\tilde{\mathbf{H}}(z)\,[\mathbf{H}(z)\tilde{\mathbf{H}}(z)]^{-1} = \gamma\,\tilde{\mathbf{H}}(z)\,\frac{adj\,[\mathbf{H}(z)\tilde{\mathbf{H}}(z)]}{det\,[\mathbf{H}(z)\tilde{\mathbf{H}}(z)]}, \tag{5.11}$$

where $\tilde{\mathbf{H}}(z) = \mathbf{H}^*\left(\frac{1}{z^*}\right)$.

Now, from (5.7) it follows that

$$\mathbf{H}(z)\tilde{\mathbf{H}}(z) = \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} \mathbf{H}_l \mathbf{H}_{l'}^* z^{-l+l'}$$

Invoking (5.9), we can ignore the terms $\mathbf{H}_l\mathbf{H}_{l'}^*$ for $l \neq l'$ in the above equation and therefore have

$$\mathbf{H}(z)\tilde{\mathbf{H}}(z) \approx \sum_{l=0}^{L-1} \mathbf{H}_l \mathbf{H}_l^* = \mathbf{H}_c.$$

Assuming that $L \geq M_R$, the matrix $\mathbf{H}_c$ will generally be full rank and hence we can write

$$\mathbf{F}(z) \simeq \gamma\, \tilde{\mathbf{H}}(z) \mathbf{H}_c^{-1} \tag{5.12}$$

or equivalently

$$\mathbf{F}_{-l} \simeq \gamma\, \mathbf{H}_l^* \mathbf{H}_c^{-1}.$$

We have thus found an analytic expression of the zero-forcing precoding filter which happens to (approx) equal the matched filter multiplied by $\mathbf{H}_c^{-1}$ from the right.

Finally, we note that $\mathbf{H}_c$ can be further simplified to yield

$$\mathbf{H}_c = \sum_{l=0}^{L-1} |\alpha_l|^2 \mathbf{a}_{R,l} \mathbf{a}_{R,l}^*, \tag{5.13}$$

where we have assumed that the transmit array signature vectors have unit norm.

In many applications, it is sufficient to equalize the MIMO channel up to a mixing (flat-fading) matrix, where we are again assuming a zero-forcing criteria. This can be achieved by setting

$$\mathbf{F}(z) \simeq \gamma \tilde{\mathbf{H}}(z). \tag{5.14}$$

Substituting $\mathbf{F}(z)$ from (5.14) into (5.5) we can write in the time domain

$$\hat{\mathbf{s}}_n = \gamma \mathbf{H}_c \mathbf{s}_n + \gamma \sum_{l,l'=0,l\neq l'}^{L-1} \mathbf{H}_l \mathbf{H}_{l'}^* \mathbf{s}_{n-l+l'} + \mathbf{v}_n, \tag{5.15}$$

where the first term is the desired signal and the second term corresponds to the residual ISI in the system.

Using (5.13), we can see that the flat-fading channel matrix $\mathbf{H}_c$ has contributions from all L multipaths added coherently, which implies that the precoder in (5.14) not only mitigates ISI, but also converts the available path diversity into spatial diversity.

The following measure quantifies the residual ISI:

$$\mathcal{D}_{ms} = \frac{\mathrm{Tr}\{\sum_{l,l'=0,l\neq l'}^{L-1} \mathbf{H}_l \mathbf{H}_{l'}^* \mathbf{H}_l \mathbf{H}_{l'}^*\}}{\mathrm{Tr}\{\mathbf{H}_c \mathbf{H}_c^*\}} \tag{5.16}$$

which is the energy in the ISI-terms in (5.15) divided by the energy in the ISI-free component $\mathbf{H}_c$.

5.2.3 Simulations

Simulation Example 1. We study the behavior of $\mathcal{D}_{ms}$ as a function of the angle spread and M_T. We assume a 2 path propagation model with a direct line of sight path included, i.e., $L = 2$ with $\theta_{T,1} = 0$. The two paths were separated by one symbol period and $|\alpha_1|^2 = |\alpha_2|^2$=1. Figure 5.6 shows $\mathcal{D}_{ms}$ as a function of $\theta_{T,2}$, where $\theta_{R,1} = 0$ and $\theta_{R,2}$ was drawn from a uniform distribution in the interval $\left[0, \frac{\pi}{2}\right]$, and $\mathcal{D}_{ms}$ was computed by averaging over $5,000$ independent Monte Carlo runs for each value of $\theta_{T,2}$ (which was varied in steps of 2 degrees). We can see that for a fixed number of transmit antennas, $\mathcal{D}_{ms}$ decreases for increasing angle separation at the transmitter. We furthermore observe that for fixed angle spread at the transmitter, $\mathcal{D}_{ms}$ decreases as M_T increases. We thus conclude that the assumption (5.8) will be well-justified if M_T and/or the angle separation at the transmitter is large.

Simulation Example 2. In this simulation example we corroborate our statement that FIR precoding mitigates ISI and converts multipath diversity into spatial diversity. For $M_T = 6, M_R = 2$ and varying L, we study the performance of the precoded system in terms of bit error rate. The transmit array signature vectors were chosen to be the columns of an $M_T \times L$ unitary matrix. For $L = 2, 3, ..., 6$ we therefore have perfect orthogonality of the $\mathbf{a}_{T,l}$. The SNR defined as the total transmit power (over all transmit antennas) divided by the noise power at one receive antenna was set to 10dB. We performed $50,000$ Monte Carlo runs for each value of L. In the receiver we employed a uniform linear array and drew the angles of arrival from independent uniform distributions in the interval $\left[0, \frac{\pi}{2}\right]$. We employed $4-$QAM modulation and performed ML decoding on the 2×2 matrix $\mathbf{H}_c$. From Figure 5.7 it follows that the performance of the precoded system improves for increasing L, because of increased ISI mitigation. For $L > M_T$ the transmit array signature separation is poor and hence the assumption (5.9) is not valid. In other words, the precoder increasingly fails to mitigate ISI which results in a performance degradation since the receiver performs ML decoding on $\mathbf{H}_c$ only and does not take into account the resulting ISI terms.

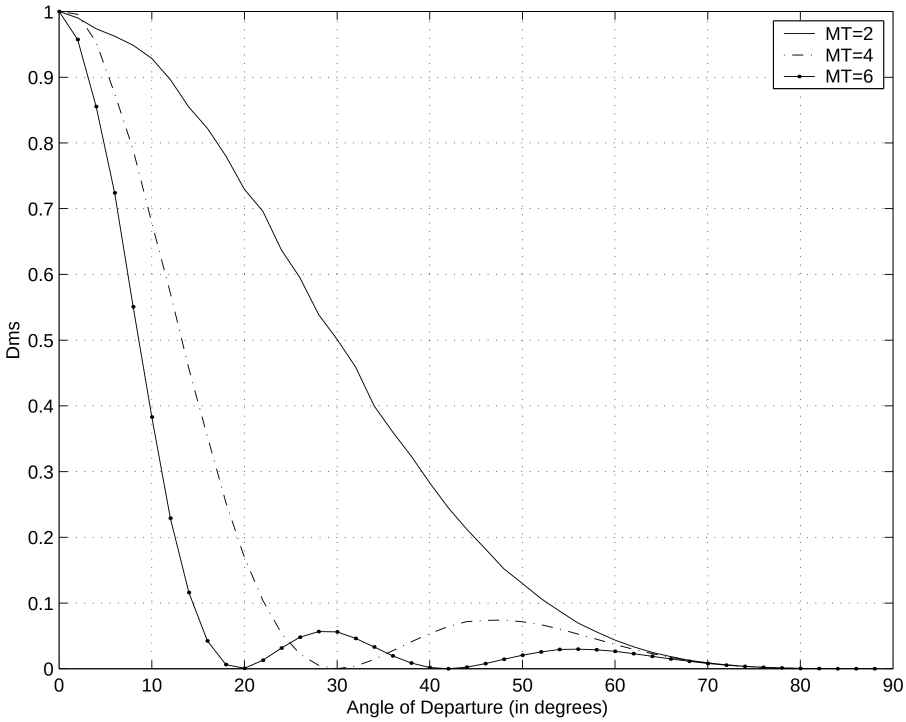

Figure 5.6: Behavior of $\mathcal{D}_{ms}$ as a function of angle separation at the transmitter for different values of M_T.

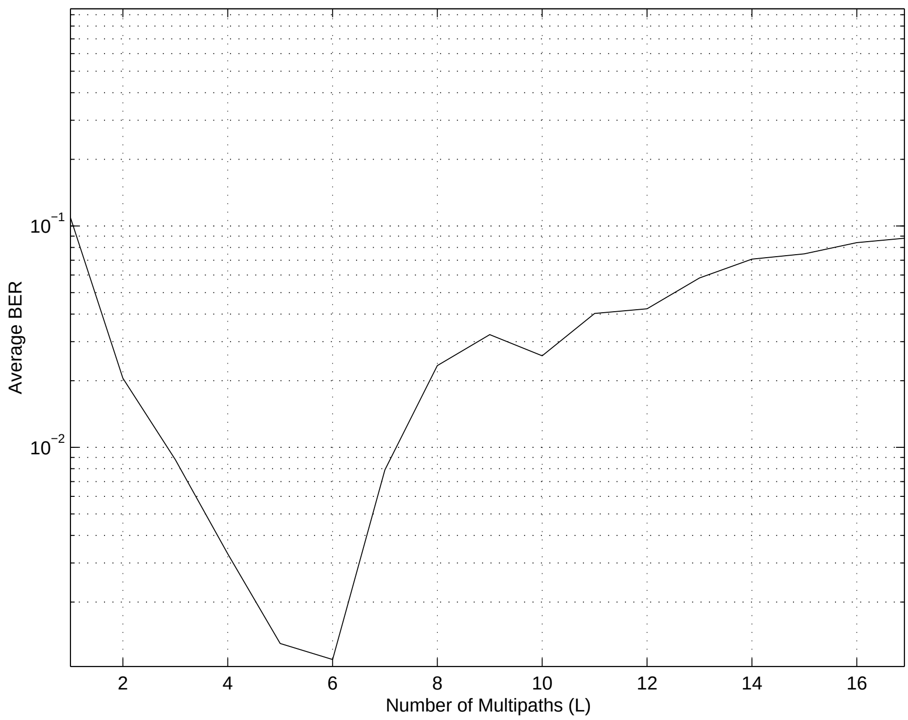

Figure 5.7: Average bit error rate as a function of the number of multipaths for SNR = 10dB, $M_T = 6$, and $M_R = 2$

Chapter 6

Precoding with Partial Channel Knowledge

Note: This is a revised version of the original thesis chapter. This revision reflects the Errata to the article, H. Sampath and A. Paulraj. "Linear Precoding for Space-Time Coded Systems with Known Fading Correlations." IEEE Communications Letters, vol. 6., No. 6, June 2002.

Space-time coded systems [15, 16, 18, 19, 20] encode the input symbol stream across multiple antennas and across time to get transmit diversity and coding gain. Typically, space-time codes are designed for an iid flat-fading matrix channel and assume no channel knowledge at the transmitter. The code design criteria assumes that the transmit and receive antennas are uncorrelated and each element of the

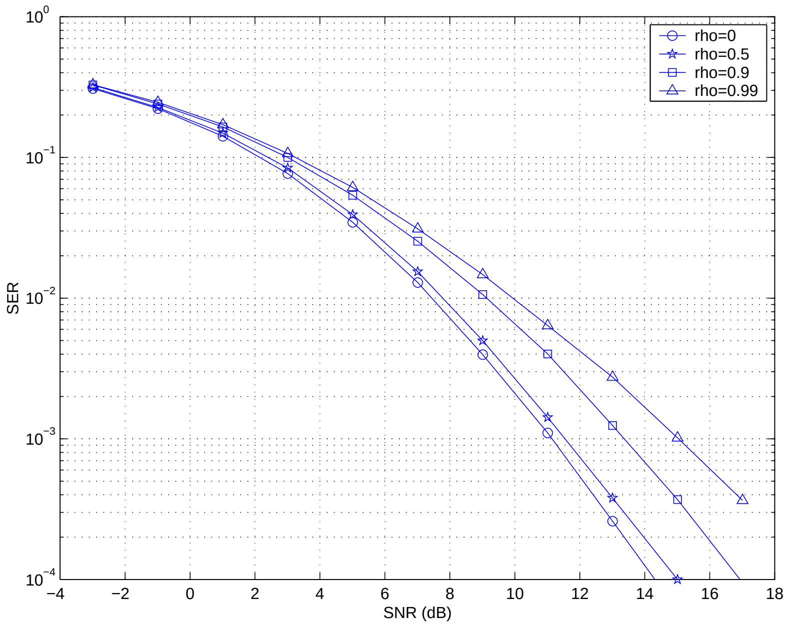

Figure 6.1: Performance of Alamouti space-time code as a function of antenna correlation

MIMO channel matrix fades independently.

However, such an assumption may not be true in practice. For example, in the downlink, the base-station is unobstructed and has no local scatterers. This induces correlation across transmit antennas as a result of which the MIMO channel matrix entries do not fade independently. In such a case, the performance of space-time codes deteriorates as shown in Figure 6.1, where we have illustrated the performance of the well-known Alamouti code [39] for an $M_T = M_R = 2$ system, for antenna correlation coefficients of $\rho = 0, 0.5, 0.9, 0.99$. This can be explained by the loss of space-diversity in the MIMO channel with the increase in antenna correlation.

Antenna correlation is typically stationary over time and hence can be estimated and fed back to the transmitter. A natural question to ask is how to design optimal precoders that use antenna correlation to improve system performance. There can be many designs possible depending on the application. A design that is optimum for a certain application may not be optimum for another application. Recently, linear precoding for space-time coded systems has been considered for the case when the transmitter has knowledge of only the *angle of departures* of the channel multipaths[36, 37].

In this chapter, we will use the *average PEP criteria* to design an optimum linear precoder for a space-time coded system, assuming knowledge of only the transmit antenna correlation matrix. We assume a flat fading channel and a maximum likelihood receiver. We show that the optimum precoder pours power on the eigenmodes of the transmit antenna correlation matrix. The power allocation policy on the eigenmodes is given by a form of the well-known *waterpouring* policy. The power allocation policy is dependant on the eigenvalues of the transmit correlation matrix, noise variance and the symbol constellation. Simulation results are presented to illustrate the performance improvement possible with linear precoder in a space-time coded modem. (The codes designed by Tarokh et al [15] do not incorporate transmit antenna correlations.)

Finally, we note that the problem of finding the optimal linear precoder that maximizes the expected capacity of a MIMO channel assuming known fading correlations has recently been solved by Shiu and Kahn [73, 74]. The optimum precoder was shown to also pour power on the eigenmodes of the transmit correlation matrix according to the well-known *waterpouring* policy. The difference (w.r.t the PEP-based policy) is that the capacity achieving power allocation policy is not dependant on the symbol constellation size.

6.1 System and Channel Model

Consider the system model for a flat fading channel shown in Figure 6.2. The transmitter has M_T antennas and receiver has M_R antennas. At time instant n, the space-time encoder (see [15]) takes in a set of input bits and creates a $B \times 1$ output code symbol vector $\mathbf{c}(n) = vec[c_1(n), c_2(n), ..., c_B(n)]$, where $B \leq M_T$. A time sequence of N code symbol vectors $\mathbf{c}(Nt), \mathbf{c}(Nt-1), ..., \mathbf{c}(Nt-N+1)$ form a $B \times N$ space-time codeword, $\mathbf{s}(t) = [\mathbf{c}(Nt)\ \mathbf{c}(Nt-1)\ \mathbf{c}(Nt-N+1)]$. The encoder chooses the output codewords so that the coding gain and diversity gain are maximized [15].

The $B \times N$ space-time codeword is then processed by the $M_T \times B$ precoder matrix ($\mathbf{F}$) and launched into the MIMO channel using M_T independent transmit antennas. The precoder ($\mathbf{F}$) can add redundancy of $M_T - B$ and provide additional diversity gains.

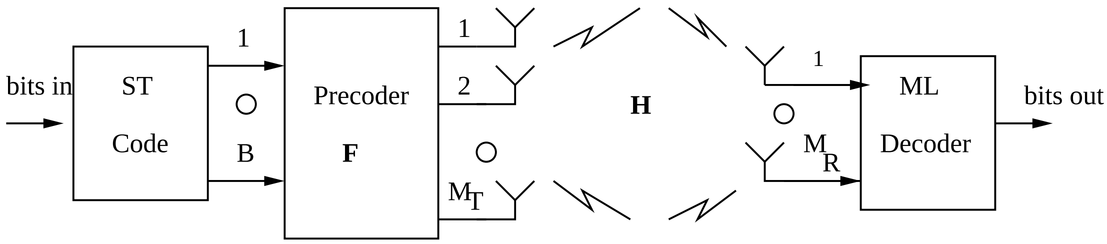

Figure 6.2: System model: Combined precoding and space-time coding

The system equation is thus given as (see Figure 6.2):

$$\mathbf{x}(t) = \mathbf{H}(t)\mathbf{F}\mathbf{s}(t) + \mathbf{n}(t) \tag{6.1}$$

where $\mathbf{s}(t)$ is the $B \times N$ codeword matrix, $\mathbf{F}$ is the $M_T \times B$ precoder matrix, $\mathbf{H}(t)$ is the $M_R \times M_T$ channel matrix, $\mathbf{x}(t)$ is the $M_R \times N$ received signal matrix, and $\mathbf{n}(t)$ is the $M_R \times N$ noise matrix. We assume that the noise is white across space and time for simplicity of analysis, i.e., $\mathbf{R_{nn}}(t) = \sigma^2\mathbf{I}$.

Assuming $\mathbf{H}(t)$ and $\mathbf{F}$ are known at the receiver, the latter computes the decision metric

$$||\mathbf{x}(t) - \mathbf{H}(t)\mathbf{F}\mathbf{s}(t)||^2 \tag{6.2}$$

over all possible codewords $\mathbf{s}(t)$ and decides in favor of the code word with the minimum metric.

Typically, in a wireless system, the transmit antennas are correlated in the downlink. This is because the base-station antennas are normally high enough and have no local scatterers. This means that the columns of $\mathbf{H}$ have correlated fading. However, the mobile antennas are typically uncorrelated due to local scatterers. This means that the rows of $\mathbf{H}$ are independently fading with respect to each other.

The MIMO channel model which incorporates the antenna correlations at the transmitter is given from Chapter 2 as:

$$\mathbf{H}(t) = \mathbf{H}_{\mathbf{w}}(t)\,\mathbf{R}_{a,T}^{1/2} \tag{6.3}$$

where $\mathbf{R}_{a,T}$ is the transmit antenna correlation matrix, and $\mathbf{H}_{\mathbf{w}}(t)$ is an iid matrix with elements having mean 0 and variance 1. Substituting (6.3) in (6.1) we get the system equation as

$$\mathbf{x}(t) = \mathbf{H}_{\mathbf{w}}(t)\mathbf{R}_{a,T}^{1/2}\mathbf{F}\mathbf{s}(t) + \mathbf{n}(t) \tag{6.4}$$

6.2 Average PEP minimization

Assuming an ML detector, the pairwise error probability (PEP) is defined to be the error probability of the receiver choosing in favor of the nearest $B \times N$ codeword $\mathbf{s}^l(t)$ instead of the transmitted codeword $\mathbf{s}^k(t)$. Define $\tilde{\mathbf{E}}(k,l,t) := [\mathbf{s}^l(t) - \mathbf{s}^k(t)]$ to be the code error matrix. The performance of the space-time code is dominated by the so-called minimum distance error matrix, $\mathbf{E} = arg\,min_{\tilde{\mathbf{E}}(k,l,t)}\ det\,[\tilde{\mathbf{E}}(\mathbf{k},\mathbf{l},\mathbf{t})\tilde{\mathbf{E}}(\mathbf{k},\mathbf{l},\mathbf{t})^*]$.

The rank of the minimum distance error matrix gives a measure of diversity gain and the determinant of the error matrix provides a measure of the coding gain. Hence, a good space-time code has a full-rank error matrix with a large determinant $det\,[\mathbf{E}\mathbf{E}^*]$ [15]. We assume for the rest of the discussion that the error performance of the space-time code is dominated by the properties of $\mathbf{E}$ (and influenced by some other error matrix, say, $\hat{\mathbf{E}}$ with negligible probability) so that it is sufficient to optimize the linear precoder based on $\mathbf{E}$ alone.

When the channel is known at the receiver, the PEP is well-approximated by (see for example [15]):

$$P(\mathbf{s}^k(t) \rightarrow \mathbf{s}^l(t)) \leq e^{-d^2_{min}/2} \tag{6.5}$$

where we have

$$d^2_{min} = \frac{1}{2\sigma^2}||\mathbf{H_w}(t)\mathbf{R}^{1/2}_{a,T}\mathbf{F}\mathbf{E}||^2 \tag{6.6}$$

Also, let $\mathcal{D} = \frac{1}{4\sigma^2}\mathbf{R}^{1/2}_{a,T}\mathbf{F}\mathbf{E}\mathbf{E}^*\mathbf{F}^*\mathbf{R}^{*1/2}_{a,T}$. Define the EVD : $\mathcal{D} = \mathbf{V_d}\mathbf{\Lambda_d}\mathbf{V}^*_\mathbf{d}$, where $\mathbf{V_d}$ is a $M_T \times M_T$ orthogonal eigenmatrix and $\mathbf{\Lambda_d}$ is a $M_T \times M_T$ diagonal matrix with diagonal elements, $\lambda_{d,i}$, for $i = 1, 2, ...M_T$.

An upper bound on the *average* PEP can be obtained by taking the expectation of (6.5) with respect to $\mathbf{H_w}(t)$. The average PEP was derived in [15] as:

$$\bar{P}(\mathbf{s}^k \rightarrow \mathbf{s}^l) \leq \left[\prod_{i=1}^{M_T}\left(\frac{1}{1+\lambda_{d,i}}\right)\right]^{M_R} \tag{6.7}$$

We now note that $\prod_{i=1}^{M_T}(1+\lambda_{d,i}) = det(\mathbf{I}+\mathcal{D})$. Hence, (6.7) can be re-written as:

$$\begin{aligned}\bar{P}(\mathbf{s}^k \rightarrow \mathbf{s}^l) &\leq \left[det(\mathbf{I}+\mathcal{D})\right]^{-M_R}\\ &= \left[det(\mathbf{I}+\frac{\mathbf{R}_{a,T}^{1/2}\mathbf{F}\mathbf{E}\mathbf{E}^*\mathbf{F}^*\mathbf{R}_{a,T}^{*1/2}}{4\sigma^2})\right]^{-M_R}\end{aligned} \tag{6.8}$$

The optimum $\mathbf{F}$ to minimize the average PEP can be obtained by solving the following optimization problem:

$$\begin{aligned}\max_{\mathbf{F}}\ J = &\ det(\mathbf{I}+\frac{1}{4\sigma^2}\mathbf{R}_{a,T}^{1/2}\mathbf{F}\mathbf{E}\mathbf{E}^*\mathbf{F}^*\mathbf{R}_{a,T}^{1/2})\\ &subject\ to: \quad Tr(\mathbf{F}\mathbf{F}^*) = p_0\end{aligned} \tag{6.9}$$

where p_0 is the total power constraint on the linear precoder. Let us define the EVD: $\mathbf{E}\mathbf{E}^* = \mathbf{V_e}\mathbf{\Lambda_e}\mathbf{V_e}^*$ where $\mathbf{V_e}$ is the $B \times B$ orthogonal eigen matrix and $\mathbf{\Lambda_e}$ is the $B \times B$ diagonal matrix of eigenvalues, $\lambda_{e,i}$, for $i = 1, 2.., B$. Let us also define the singular value decomposition (SVD)

$$\mathbf{R}_{a,T}^{1/2} = (\mathbf{U_r} \ \ \tilde{\mathbf{U}}_\mathbf{r}) \begin{pmatrix} \mathbf{\Lambda_r} & 0 \\ 0 & \tilde{\mathbf{\Lambda}}_\mathbf{r} \end{pmatrix} (\mathbf{V_r} \ \ \tilde{\mathbf{V}}_\mathbf{r})^* \tag{6.10}$$

where $\mathbf{U_r}$ and $\mathbf{V_r}$ are $M_T \times B$ orthogonal eigenmatrices which form a basis for the range space of $\mathbf{R}_{a,T}^{1/2}$, $\mathbf{\Lambda_r}$ is a diagonal matrix containing the B non-zero singular-values arranged in a decreasing order from top-left to bottom-right; $\tilde{\mathbf{\Lambda}}_\mathbf{r}$ contains the zero singular-values; $\tilde{\mathbf{V}}_\mathbf{r}$ and $\tilde{\mathbf{U}}_\mathbf{r}$ are $M_T \times (M_T - B)$ orthogonal eigenmatrices which constitute a basis for the null-space of $\mathbf{R}_{a,T}^{1/2}$. We will assume that $rank(\mathbf{H}) = B$ in our analysis.

Lemma 7 *The solution of the optimization problem (6.9) is given as:*

$$\mathbf{F} = \mathbf{V_r \Phi_f V_e^*}$$

$$\mathbf{\Phi_f}^2 = (\gamma\mathbf{I} - 4\sigma^2 \mathbf{\Lambda_r^{-2} \Lambda_e^{-1}})_+ \tag{6.11}$$

where the $(.)_+$ *sign means that non-negative values are not allowed, and* $\gamma > 0$ *is a constant that is computed from the trace constraint.*

Proof: See Appendix (A.6) ■

A few observations are now in order:

- The optimum precoder structure is a function of the minimum distance error matrix ($\mathbf{E}$) and the transmit antenna correlation matrix ($\mathbf{R}_{a,T}$), and hence its properties will change when either of them changes. The rotation matrices $\mathbf{V_e}$ and $\mathbf{V_r}$ in the linear precoder $\mathbf{F}$ ensure that the eigenmodes of the $\mathbf{E}$ are aligned in the same direction as the largest eigenmodes of $\mathbf{R}_{a,T}$. Intuitively, this maximizes the minimum distance and hence minimizes the average PEP. For a non-orthogonal space-time code, the computation of $\mathbf{E}$ and the sensitivity of the performance of space-time code w.r.t $\mathbf{E}$ is difficult to evaluate.

For an orthogonal space-time code the solution simplifies since we have that $\mathbf{E}\mathbf{E}^* = \beta\mathbf{I}$ and hence $\mathbf{\Lambda_e} = \beta\mathbf{I}$ and $\mathbf{V}_e = \mathbf{I}$. Assuming QAM transmit symbol constellation and unit average constellation energy, we have that $\beta = \frac{6}{M-1}$, where M is the constellation size. In such a case, the optimum precoder simplifies to the following:

$$\mathbf{F} = \mathbf{V_r}\mathbf{\Phi_f}$$
$$\mathbf{\Phi_f}^2 = (\gamma\mathbf{I} - 2\sigma^2\frac{M-1}{3}\mathbf{\Lambda_r^{-2}})_+ \tag{6.12}$$

Note that the power allocation policy on the eigenmodes of $\mathbf{R}_{a,T}$ is a form of the well-known *waterpouring* policy, and is dependant on the constellation size, noise variance and eigenvalues of $\mathbf{R}_{a,T}$. The policy dictates that more power is transmitted on eigenmodes with higher gains and less power is transmitted on eigenmodes with weaker gains.

A power allocation policy similar in spirit to the above waterfilling policy also achieves the expected capacity of the MIMO channel, when only the transmit antenna correlations are known[73]. The similarity of the two designs can be explained as follows. Mathematically, the cost functions that are optimized in either designs are similar in that both are derivatives of the form $det(I + X)$. The difference between the 2 designs is that the capacity achieving waterfilling policy is independent of the constellation size, while the PEP-based design is dependent on the constellation size.

6.3 Simulations

We will now illustrate the performance improvement of a space-time block coded system using the precoder developed above.

For our simulations, we assume that the precoder matrix is power-preserving i.e, $Tr(\mathbf{\Phi_f^2}) = B$. Furthermore, we assume for our simulations that

$$\mathbf{R}_{a,T} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \vdots \\ \cdots & \cdots & \rho & 1 \end{pmatrix}_{M_T \times M_T} \tag{6.13}$$

where ρ is the correlation coefficient between transmit antennas. The elements of $\mathbf{H}(t)$ are iid Gaussian with mean 0 and variance 1. The noise is assumed white across the receive antennas. Our simulations are performed over 100000 to 1000000 Monte-Carlo runs, depending on the symbol error rate (SER). For each simulation run, a different instantiation of $\mathbf{H}(t)$, noise samples, and transmit symbols are used. The transmit symbols are uniformly distributed from a 4-QAM constellation, with unity average energy. We further assume that $p_0 = B$, regardless of M_T, i.e., the linear precoder is power preserving. In the performance plots (SER vs. SNR plots) to be shown below, the SNR is defined as the ratio of the total transmit power across all antennas and the received noise variance per receive antenna (assuming unity constellation energy and power preserving linear precoder).

We consider a MIMO system with $M_R = 2$ receive antennas and $M_T \geq 2$ transmit antennas. We use the rate 1 space-time Alamouti block code with $B = 2$ output streams, developed in [39]. The output of this space-time code is then processed by a $M_T \times 2$ precoder that optimally maps the 2 symbol streams on to M_T transmit antennas.

Figures 6.3, 6.4, 6.5 and 6.6 corroborate the performance improvement of using our precoder with such a space-time code, for transmit antenna correlation coefficients of $\rho = 0, 0.5, 0.9, 0.99$. It can be seen that as the antenna correlation increases from 0 to 0.99, precoding gains can be significant. This can be explained as follows.

As correlation increase, $\mathbf{R}_{a,T}$ has only a few dominant eigenmodes. The precoder transmits more power on the stronger eigenmodes and hence ensures better receiver SNR. In contrast, a system without a precoder will transmit power equally on all eigenmodes (including the null modes) and will result in a poor receiver SNR. We also note that when the antenna correlation is near 0 and the SNR is high enough, all eigenmodes have high gains. In such a case, a system with or without a precoder pours power uniformly on all eigenmodes. Hence, precoding gains are negligible in such cases.

When $M_T > B$, the precoder maps the B coded symbol streams onto $M_T > B$ transmit antennas to give further diversity gains, while maintaining the same receiver complexity. (Typically, in the absence of a precoder, one has to design a new space-time code for $M_T > B$ coded output streams, leading to a higher receiver decoding complexity). This is corroborated in Figures 6.3, 6.4,6.5 and 6.6, where increasing the number of transmit antennas buys us performance, despite the fact that the total transmit power is preserved regardless of the number of transmit antennas.

Figure 6.7 illustrates the performance improvement of the precoder on a rate 3/4 space-time block code developed for a $M_T = 3$ antennas in [19, 20]. We assume a 4-QAM modulation, $M_R = 2$ and an antenna correlation coefficient of $\rho = 0.99$. It can be seen that at a SER of 10^{-2}, we obtain a precoding gain of 5 dB with the same number of transmit and receive antennas. For $M_T = 4$, we obtain a precoding gain

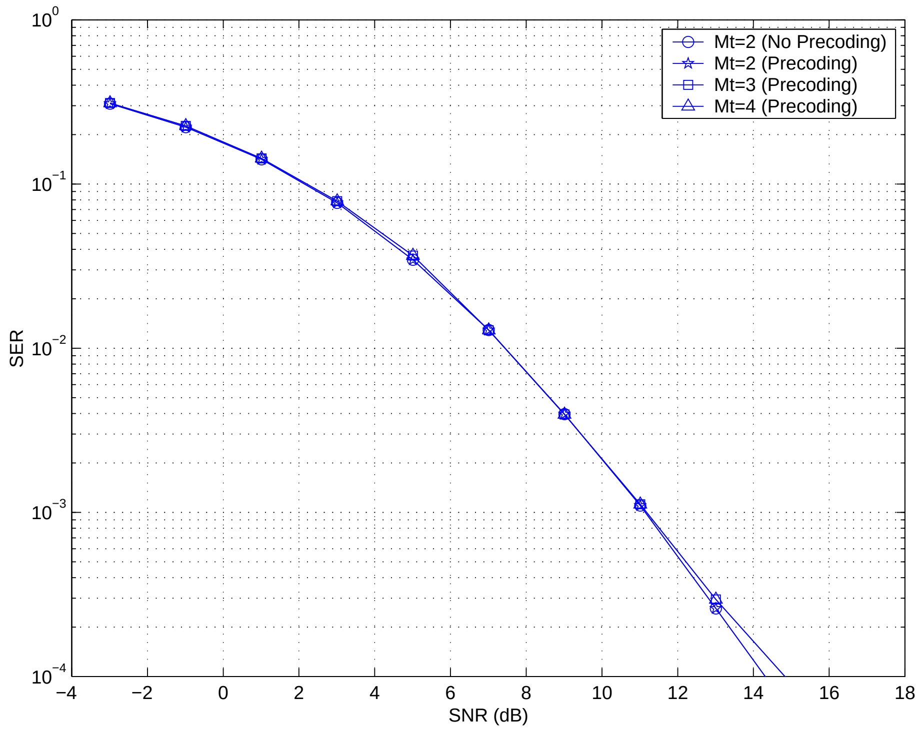

Figure 6.3: Precoding gain for Alamouti scheme assuming $\rho = 0$

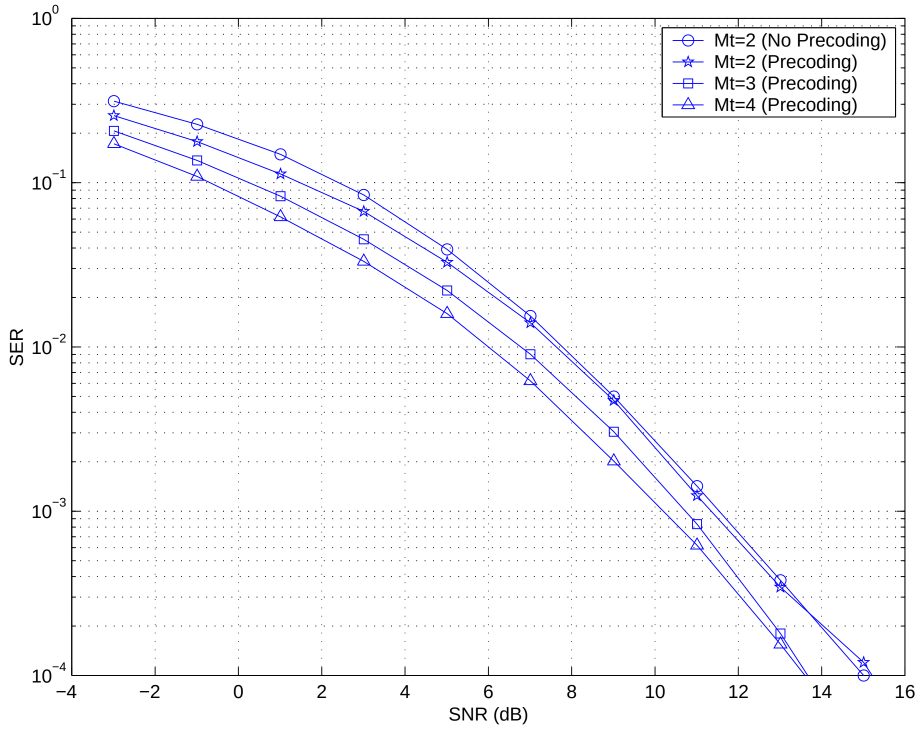

Figure 6.4: Precoding gain for Alamouti scheme assuming $\rho = 0.5$

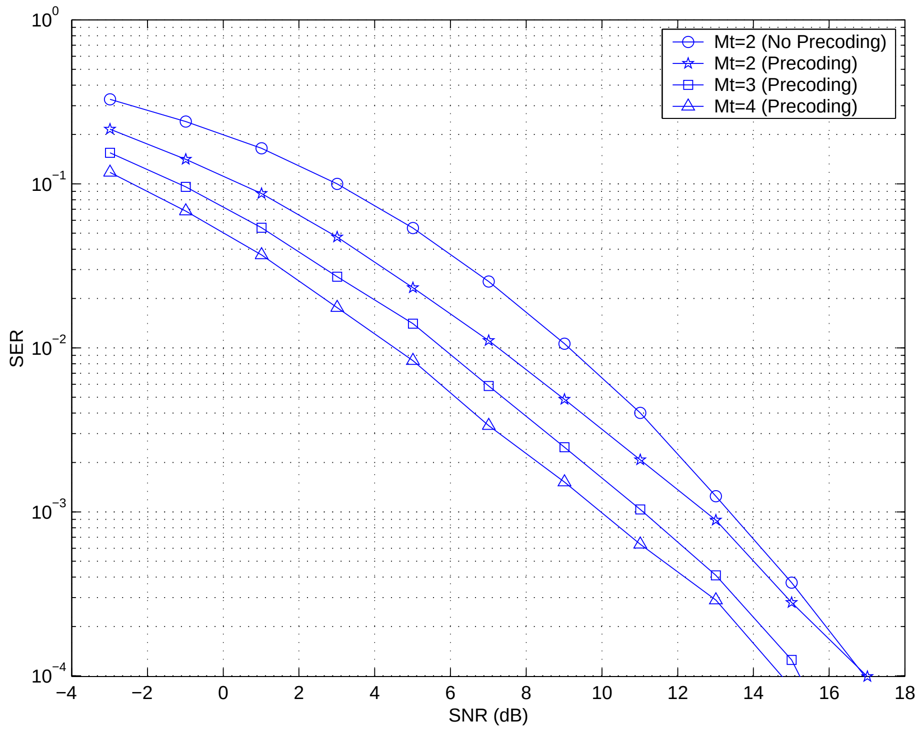

Figure 6.5: Precoding gain for Alamouti scheme assuming $\rho = 0.9$

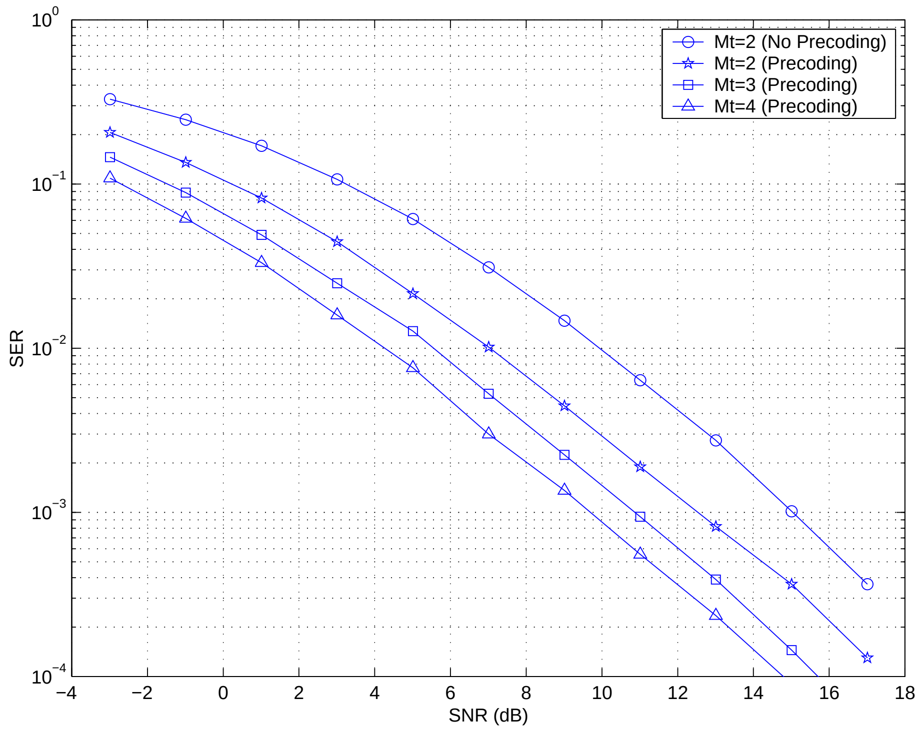

Figure 6.6: Precoding gain for Alamouti scheme assuming $\rho = 0.99$

of 6 dB at the same SER.

Figure 6.8 illustrates the performance improvement of the precoder on a rate 1/2 space-time block code developed for a $M_T = 4$ antennas in [19, 20]. We assume a 4-QAM modulation, $M_R = 2$, and an antenna correlation coefficient of $\rho = 0.99$. It can be seen that at a SER of 10^{-2}, we obtain a precoding gain of 6 dB with the same number of transmit and receive antennas. For $M_T = 5$, we now obtain a precoding gain of 7 dB at the same SER for the same transmit power.

In conclusion, we see that as $\rho \to 1$ the linear precoder obtains a gain of $10log_{10}M_T$ dB of array gain over a non-precoded system by transmitting on the dominant eigenvector. As $\rho \to 0$, the channel becomes i.i.d and the linear precoder pours power uniformly in all directions. In such a case, we see no performance improvement over a non-precoded system.

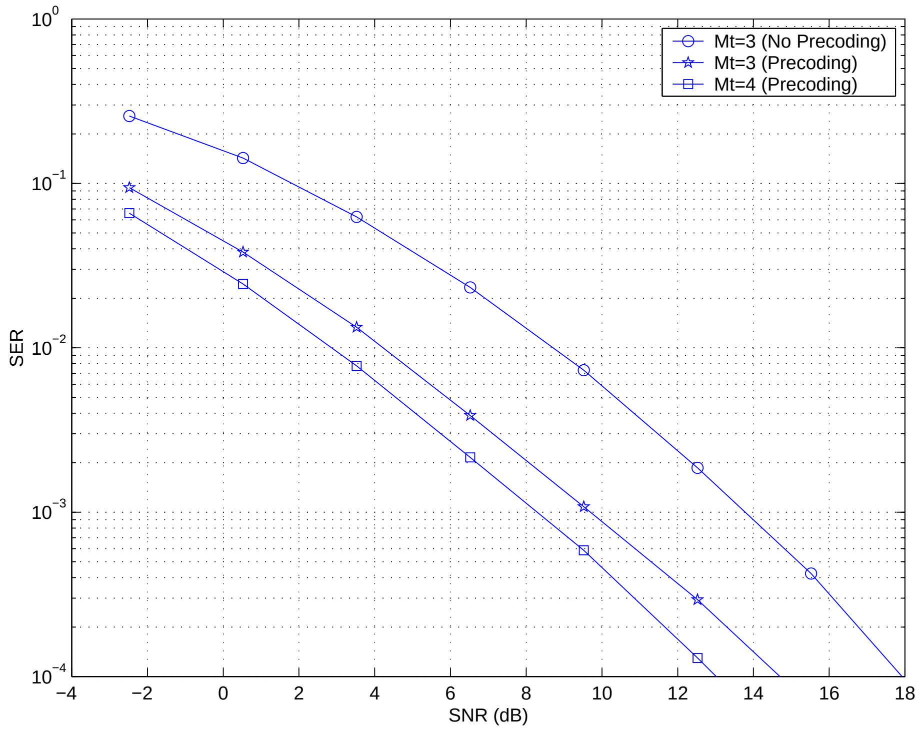

Figure 6.7: Precoding gain for rate 3/4 space-time code assuming $\rho = 0.99$

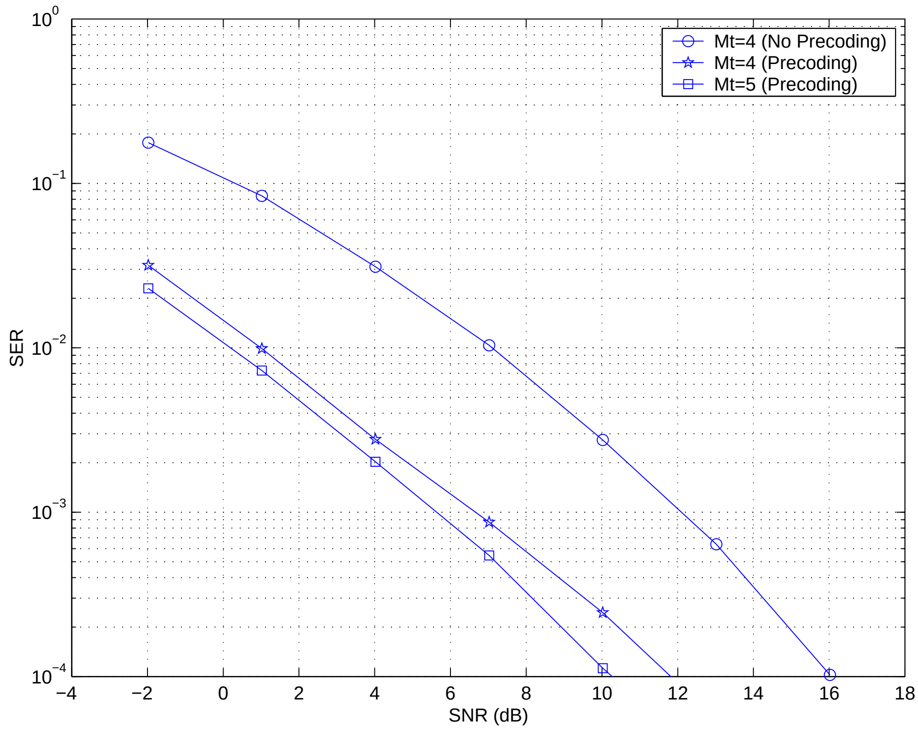

Figure 6.8: Precoding gain for rate 1/2 space-time code assuming $\rho = 0.99$

Chapter 7

Conclusions

Space-time coding and spatial multiplexing are prime candidates for achieving high data rates and link quality in wireless links. However, both the schemes assume no channel knowledge at the transmitter. In a number of applications, channel knowledge can be made available at the transmitter. A natural question to ask is how to use these channel estimates to further optimize the transmitter. There can be several ways to linearly or non-linearly optimize the transmitter and receiver depending on channel knowledge. In this thesis, we consider linear optimization schemes at the transmitter (precoding) and receiver (decoding) to improve performance.

First, assuming perfect channel knowledge at the transmitter and receiver, and a flat-fading channel,we derived a linear precoding and decoding scheme that minimizes any *weighted sum of output symbol estimation errors*, assuming a total transmit power constraint across all antennas. The optimum design forced transmission only on the eigenmodes of the MIMO channel, for any set of error weights. The power allocation on the eigenmodes depended on the error weights, which can be varied depending on the application. We showed how to vary the error weights to obtain many of the previously known designs as well as new designs. The weighted MMSE criteria thus provides a unified framework for designing jointly optimal linear precoders and decoders, assuming perfect channel knowledge at the transmitter.

Next, we developed an optimal linear precoder and decoder designs based on the (a) Pairwise error probability (PEP) criteria assuming a total power constraint (b) Geometric mean square error criteria assuming a total power constraint and (c) Maximum information rate criteria subject to a peak power constraint. For criteria (a) and (b), the optimum design forced eigenmode transmission as also suggested by the weighted MMSE design. However for criteria (c), we find that eigenmode transmission need not be the optimum strategy. This is a consequence of the peak power constraint that forces the precoder to pour power on modes other the eigenmodes.

Next, we showed how to extend the linear precoder and decoder framework to delay spread channels and multicarrier systems employing OFDM modulation. In the former case, we employ block transmission with guard symbols inserted between data blocks to prevent inter-block interference. In OFDM modulation, redundancy is added in the form of a cyclic prefix to handle delay spread.

Next, we presented a novel Finite Impulse Response (FIR) precoder structure to pre-equalize the MIMO channel with delay spread assuming a zero-forcing constraint and subject to a transmit power constraint. FIR precoding is different from block precoding considered above in that it uses an FIR filter structure that operates continuously on the input symbol vector stream and requires no guard symbol transmission.

Finally, we considered the design of precoders assuming partial channel knowledge at the transmitter and perfect channel knowledge at the receiver. In this context, we developed an optimum linear precoder for a space-time coded system, assuming knowledge of only the transmit antenna correlations. The antenna correlation properties are typically stationary over time, and can be estimated and fed back to the transmitter. Assuming a flat fading channel and a maximum likelihood receiver, we show that the optimum precoder forces eigenmode transmission on the transmit antenna correlation matrix.

In solving the above problems, we noticed a common theme in the optimum transmission strategy. For the case involving a total power constraint across all transmit antennas, the optimum precoder and decoder invariably forced eigenmode transmission (on the MIMO channel or on some derivative of the channel such as antenna correlation matrices). The only difference between different designs is the power allocation on the eigenmodes. The exception to this rule was for the problem with peak power constraint which forced the linear precoder to pour power in modes other than the eigenmodes.

In solving the above problems, we introduced several novel mathematical techniques that can potentially be very useful in solving other problems in MIMO.

Suggestions for Future Research:

- In practice, the channel and its parameters (such as antenna correlations) cannot be estimated perfectly. The performance of linear precoders and decoders needs to be characterized with channel estimation errors.
- In this thesis, we have not considered code design. Instead we have focussed only on the signal processing aspect of transmit optimization. The design of space-time codes that can be used in conjunction with linear precoders need to be investigated.
- Non-linear precoder and decoder designs need to be considered and their performance evaluated with respect to linear designs.
- Finally, the design of precoders with partial channel knowledge need to be further investigated. For example, the optimal precoder design when only the mean or variance of the channel are known is an interesting research problem.

Appendix A

Proofs of Results

A.1 Proof of Lemma 1

Consider the most general expressions for $\mathbf{F}$ and $\mathbf{G} = \bar{\mathbf{G}}\mathbf{R_{nn}}^{-1/2}$:

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f} + \tilde{\mathbf{V}}\tilde{\mathbf{\Phi}}_\mathbf{f} = \mathbf{F}_\| + \mathbf{F}_\perp \tag{A.1}$$

$$\bar{\mathbf{G}} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1/2} + \tilde{\mathbf{\Phi}}_\mathbf{g}\tilde{\mathbf{V}}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1/2} = \mathbf{G}_\| + \mathbf{G}_\perp \tag{A.2}$$

where $\mathbf{F}_\|$ and $\mathbf{G}_\|$ are each in the range space of $\mathbf{R_{nn}^{-1/2}H}$, and $\mathbf{F}_\perp$ and $\mathbf{G}_\perp$ are each in the null-space of $\mathbf{R_{nn}}^{-1/2}\mathbf{H}$; $\tilde{\mathbf{\Phi}}_\mathbf{f}$ is any $(M_T - B) \times B$ matrix, and $\tilde{\mathbf{\Phi}}_\mathbf{g}$ is any $B \times (M_T - B)$ matrix. Note that the above decomposition for $\mathbf{G}$ is valid since $\mathbf{R_{nn}}^{-1/2}$ is square and full-rank, and imposes no restrictions in our search for optimal

$\mathbf{G}$. Furthermore, we have

$$\mathbf{F}_{\perp}{}^{*}\mathbf{F}_{\|} = 0; \;\; \mathbf{F}_{\|}{}^{*}\mathbf{F}_{\perp} = 0; \;\; \mathbf{R_{nn}}^{-1/2}\mathbf{H}\mathbf{F}_{\perp} = 0; \tag{A.3}$$

$$\mathbf{G}_{\perp}\mathbf{G}_{\|}{}^{*} = 0; \;\; \mathbf{G}_{\|}\mathbf{G}_{\perp}{}^{*} = 0; \;\; \mathbf{G}_{\perp}\mathbf{R_{nn}}^{-1/2}\mathbf{H} = 0 \tag{A.4}$$

Substituting the pair $(\mathbf{G}\bar{\mathbf{R}}_{\mathbf{nn}}{}^{-1/2}, \mathbf{F})$ in equations (3.14) and (3.15) respectively, we get:

$$\mathbf{HF} = \mathbf{HFF}^{*}\mathbf{H}^{*}\mathbf{R_{nn}}^{-1/2}\bar{\mathbf{G}}^{*} + \mathbf{R_{nn}}^{1/2}\bar{\mathbf{G}}^{*} \tag{A.5}$$

$$\mathbf{W}\bar{\mathbf{G}}\mathbf{R_{nn}}^{-1/2}\mathbf{H} = \mathbf{F}^{*}\mathbf{H}^{*}\mathbf{R_{nn}}^{-1/2}\bar{\mathbf{G}}^{*}\mathbf{W}\bar{\mathbf{G}}\mathbf{R_{nn}}^{-1/2}\mathbf{H} + \mu\mathbf{F}^{*} \tag{A.6}$$

Pre-multiply equation (A.5) by $\mathbf{G}_{\perp}\mathbf{R_{nn}}^{-1/2}$ and use (A.4) to get

$$\bar{\mathbf{G}}_{\perp}\bar{\mathbf{G}}^{*} = 0 \;\; \Rightarrow \bar{\mathbf{G}}_{\perp}\bar{\mathbf{G}}_{\perp}{}^{*} = 0 \;\; \Rightarrow \bar{\mathbf{G}}_{\perp} = 0. \tag{A.7}$$

Post-multiply equation (A.6) by $\mathbf{F}_{\perp}$ and use (A.3) to get

$$\bar{\mathbf{F}}_{\perp}^{*}\mathbf{F} = 0 \;\; \Rightarrow \bar{\mathbf{F}}_{\perp}^{*}\mathbf{F}_{\perp} = 0 \;\; \Rightarrow \bar{\mathbf{F}}_{\perp} = 0. \tag{A.8}$$

Using $\bar{\mathbf{F}}_{\perp} = \bar{\mathbf{G}}_{\perp} = 0$ in (A.1) and (A.2), we get the desired result.

A.2 Proof of Lemma 2

Pre-multiplying equation (3.14) by $\mathbf{G}$ and post-multiplying equation (3.15) by $\mathbf{F}$, and using the expressions for $\mathbf{F}$ and $\mathbf{G}$ from Lemma 1 in the so-obtained equations, we get:

$$\begin{aligned}
\mathbf{\Phi_g V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_f} &= \mathbf{\Phi_g V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_f \Phi_f}^*\mathbf{V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_g}^* + \\
&\qquad \mathbf{\Phi_g V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_g}^* \\
\mathbf{W \Phi_g V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_f} &= \mathbf{\Phi_f}^*\mathbf{V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_g}^*\mathbf{W \Phi_g V^* H^* R_{nn}}^{-1}\mathbf{H V \Phi_f} + \\
&\qquad \mu \mathbf{\Phi_f}^*\mathbf{V^* V \Phi_f}
\end{aligned} \tag{A.9}$$

Using the EVD of $\mathbf{H^* R_{nn}}^{-1}\mathbf{H}$, and the fact that $\mathbf{V^* V} = \mathbf{I}$, the above equations reduce to:

$$\mathbf{\Phi_g \Lambda \Phi_f} = \mathbf{\Phi_g \Lambda \Phi_f \Phi_f}^* \mathbf{\Lambda \Phi_g}^* + \mathbf{\Phi_g \Lambda \Phi_g}^* \tag{A.10}$$

$$\mathbf{W \Phi_g \Lambda \Phi_f} = \mathbf{\Phi_f}^* \mathbf{\Lambda \Phi_g}^* \mathbf{W \Phi_g \Lambda \Phi_f} + \mu \mathbf{\Phi_f}^* \mathbf{\Phi_f} \tag{A.11}$$

Note from equation (A.10) that $\mathbf{\Phi_g \Lambda \Phi_f}$ must be Hermitian since the other terms in the equation are Hermitian. Using the same argument in equation (A.11), we conclude that $\mathbf{W \Phi_g \Lambda \Phi_f}$ must also be Hermitian. But since $\mathbf{W}$ is a diagonal matrix, this implies that $\mathbf{\Phi_g \Lambda \Phi_f}$ must be real diagonal, assuming that $\mathbf{W}$ has distinct diagonal elements. For the case when $\mathbf{W}$ has repeated diagonal elements w_{ii}, we can consider $\tilde{\mathbf{W}} = \mathbf{W} + \mathbf{\Delta_W}$ where $\mathbf{\Delta_W}$ is a perturbation matrix that ensures that w_{ii} are distinct. Since the optimal design matrices $\mathbf{G}$ and $\mathbf{F}$ are continuous functions of $\mathbf{W}$, we have that $\lim_{\mathbf{\Delta_W} \to 0} \mathbf{G}(\tilde{\mathbf{W}}) = \mathbf{G}(\mathbf{W})$, and similarly for $\mathbf{F}$. This means that we can take $\mathbf{\Phi_g \Lambda \Phi_f}$ to be real diagonal for any $\mathbf{W}$. Note that the case of $\mathbf{W} = \mathbf{I}$ has been well studied in the literature [48, 56] where the fact that $\mathbf{\Phi_g \Lambda \Phi_f}$ is diagonal was shown to be true. Next, we use $\mathbf{D_1}, \mathbf{D_2}$ etc to denote real-valued diagonal matrices of appropriate dimensions. We showed above that

$$\mathbf{\Phi_g \Lambda \Phi_f} = \mathbf{D_1} \tag{A.12}$$

Then, it follows from (A.10) and (A.11) that $\mathbf{\Phi_g \Lambda \Phi_g}^* = \mathbf{D_2} \geq 0$; $\mathbf{\Phi_f}^* \mathbf{\Phi_f} = \mathbf{D_3} \geq 0$ or equivalently

$$\mathbf{\Phi_g}^* = \mathbf{\Lambda}^{-1/2} \mathbf{U_g} \mathbf{D_2}^{1/2}; \;\; \mathbf{\Phi_f} = \mathbf{U_f} \mathbf{D_3}^{1/2} \tag{A.13}$$

where $\mathbf{U_g}$ and $\mathbf{U_f}$ are unitary matrices that are arbitrary for now and the symbol ≥ 0 denotes a positive semidefinite matrix. Inserting (A.13) into (A.12) yields: $\mathbf{D_2}^{1/2} \mathbf{U_g}^* \mathbf{\Lambda}^{1/2} \mathbf{U_f} \mathbf{D_3}^{1/2} = \mathbf{D_1}$ or, equivalently,

$$\mathbf{U_g}^* \mathbf{\Lambda}^{1/2} \mathbf{U_f} = \mathbf{D_4} \tag{A.14}$$

In particular, (A.14) implies that

$$\mathbf{U_g}^* \mathbf{\Lambda} \mathbf{U_g} = \mathbf{D_4}^2; \;\; \mathbf{U_f^*} \mathbf{\Lambda} \mathbf{U_f} = \mathbf{D_4}^2 \tag{A.15}$$

For the case when the eigenvalues of $\mathbf{\Lambda}$ are distinct, (A.15) along with (A.14) and the fact that $\mathbf{U_g}$ and $\mathbf{U_f}$ are unitary give

$$\mathbf{U_g} = \mathbf{U_f} = diag([e^{j\Theta_1}, ..., e^{j\Theta_b}]) = \mathbf{\Theta} \tag{A.16}$$

for arbitrary $\Theta_k \epsilon [0, 2\pi]$. The case of repeated eigenvalues in $\mathbf{\Lambda}$ can be handled by means of a perturbation argument similar to the one used in the analysis of (A.10) and (A.11) to obtain (A.12). Hence, we conclude that (A.16) holds in the general case. Using (A.16) in (A.13) yields:

$$\mathbf{\Phi_g} = \mathbf{\Theta}^* \mathbf{D_2}^{1/2} \mathbf{\Lambda}^{-1/2}; \;\; \mathbf{\Phi_f} = \mathbf{D_3}^{1/2} \mathbf{\Theta} \tag{A.17}$$

It is readily checked that the $\mathbf{\Theta}$ in (A.17) does not affect the Lagrangian function in (3.9) in anyway. Accordingly, setting $\mathbf{\Theta} = \mathbf{I}$ in (A.17) imposes no restriction in our optimization search. With this observation the proof is concluded.

A.3 Proof of Theorem 1

Substituting (A.17) with $\mathbf{\Theta} = \mathbf{I}$ in (A.10), (A.11) and (A.12), we get:

$$\mathbf{D_1} = \mathbf{D_1}^2 + \mathbf{D_2} \tag{A.18}$$

$$\mathbf{W}\mathbf{D_1} = \mathbf{W}\mathbf{D_1}^2 + \mu\mathbf{D_3} \tag{A.19}$$

$$\mathbf{D_2}\mathbf{D_3}\mathbf{\Lambda} = \mathbf{D_1}^2 \tag{A.20}$$

From (A.18) and (A.19) we get, respectively

$$\mathbf{D_2} = \mathbf{D_1}(\mathbf{I} - \mathbf{D_1}); \quad \mu\mathbf{D_3} = \mathbf{W}\mathbf{D_1}(\mathbf{I} - \mathbf{D_1}) \tag{A.21}$$

Inserting (A.21) in (A.20), we get, $\mathbf{W}\mathbf{D_1}^2(\mathbf{I} - \mathbf{D_1})^2\mathbf{\Lambda} = \mu\mathbf{D_1}^2$ or equivalently the solution (assuming $\mathbf{D_1} > 0$):

$$\mathbf{D_1} = (\mathbf{I} - \mu^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2})_+ \tag{A.22}$$

Using (A.22) in (A.21) yields, for $\mu > 0$, $\mathbf{D_2} = (\mu^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2} - \mu\mathbf{\Lambda}^{-1}\mathbf{W}^{-1})_+$ and $\mathbf{D_3} = \mathbf{W}(\mu^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2} - \mathbf{\Lambda}^{-1}\mathbf{W}^{-1})_+$. The above expressions for $\mathbf{D_2}$ and $\mathbf{D_3}$ along with (A.17) (with $\mathbf{\Theta} = \mathbf{I}$) lead to (3.20) and (3.19), which concludes the proof for the case when $\mathbf{D_1} > 0$.

A.4 Proof of Lemma 5

We have from (4.5) that

$$\mathbf{SNR}(\mathbf{F},\mathbf{G}) = \mathbf{F}^*\mathbf{H}^*\mathbf{G}^*(\mathbf{G}\mathbf{R_{nn}}\mathbf{G}^*)^{-1}\mathbf{G}\mathbf{H}\mathbf{F}$$

$$= \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1/2}\mathbf{R_{nn}}^{1/2}\mathbf{G}^*(\mathbf{G}\mathbf{R_{nn}}\mathbf{G}^*)^{-1}\mathbf{G}\mathbf{R_{nn}}^{1/2}\mathbf{R_{nn}}^{-1/2}\mathbf{H}\mathbf{F} \tag{A.23}$$

Noting that $\mathbf{\Pi}_{\mathbf{R_{nn}^{1/2}}\mathbf{G}^*} = \mathbf{R_{nn}^{1/2}}\mathbf{G}^*(\mathbf{G}\mathbf{R_{nn}}\mathbf{G}^*)^{-1}\mathbf{G}\mathbf{R_{nn}^{1/2}} \leq \mathbf{I}$ is the orthogonal projector onto the range space of $\mathbf{R_{nn}}^{1/2}\mathbf{G}^*$, we get from (A.23) that

$$\mathbf{SNR}(\mathbf{F},\mathbf{G}) = \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1/2}\mathbf{\Pi}_{\mathbf{R_{nn}}^{1/2}\mathbf{G}^*}\mathbf{R_{nn}}^{-1/2}\mathbf{H}\mathbf{F} \leq \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}\mathbf{F} \tag{A.24}$$

The $\mathbf{SNR}(\mathbf{F},\mathbf{G})$ reaches its upper-bound if and only if $\mathbf{R_{nn}^{1/2}}\mathbf{G}^* = \mathbf{R_{nn}}^{-1/2}\mathbf{H}\mathbf{F}\mathbf{\Phi_g}^*$, where $\mathbf{\Phi_g}$ is an invertible arbitrary $B \times B$ matrix. Alternatively, the upper bound is reached if and only if

$$\mathbf{G} = \mathbf{\Phi_g}\mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} \tag{A.25}$$

Consider the most general expression for $\mathbf{F}$

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f} + \tilde{\mathbf{V}}\tilde{\mathbf{\Phi}}_\mathbf{f} = \mathbf{F}_\parallel + \mathbf{F}_\perp \tag{A.26}$$

where $\mathbf{F}_{\|}$ is in the range space of $\mathbf{R_{nn}^{-1/2}H}$, or $\mathbf{\Lambda^{1/2}V^*}$ and $\mathbf{F}_{\perp}$ is in the null-space of $\mathbf{R_{nn}}^{-1/2}\mathbf{H}$ or $\mathbf{\Lambda^{1/2}V^*}$; $\tilde{\mathbf{\Phi}}_{\mathbf{f}}$ is any $(M_T - B) \times B$ matrix, and $\mathbf{\Phi_f}$ is any $B \times B$ matrix. Substituting (A.26) in (A.24) and using the fact that $\mathbf{F}_{\perp}$ is in the null-space of $\mathbf{R_{nn}}^{-1/2}\mathbf{H}$ we get:

$$\begin{aligned}\mathbf{SNR}(\mathbf{F}) = \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{HF} &= (\mathbf{F}_{\|} + \mathbf{F}_{\perp})^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}(\mathbf{F}_{\|} + \mathbf{F}_{\perp}) \\ &= \mathbf{F}_{\|}{}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{HF}_{\|} = \mathbf{SNR}(\mathbf{F}_{\|}) \end{aligned} \tag{A.27}$$

Furthermore we have

$$Tr(\mathbf{FF}^*) = Tr(\mathbf{F}_{\|}\mathbf{F}_{\|}^* + \mathbf{F}_{\perp}\mathbf{F}_{\perp}^*) \geq Tr(\mathbf{F}_{\|}\mathbf{F}_{\|}^*) \tag{A.28}$$

with equality if and only if $\mathbf{F}_{\perp} = 0$. Hence, without any loss of generality, we can assume

$$\mathbf{F} = \mathbf{F}_{\|} = \mathbf{V\Phi_f} \tag{A.29}$$

Using the EVD of $\mathbf{H}^*\mathbf{R_{nn}^{-1}H}$, and $\mathbf{F} = \mathbf{V\Phi_f}$ in (A.27) we can write:

$$\mathbf{SNR}(\mathbf{F}) = \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{HF} = \mathbf{\Phi_f\Lambda\Phi_f^*} \tag{A.30}$$

Hence, the optimization problem posed in (4.8) reduces to the following:

$$\max_{\mathbf{\Phi_f}} \mathcal{J} = \lambda_{min}\left(\mathbf{\Lambda \Phi_f \Phi_f^*}\right)$$
$$\textit{subject to}: \quad Tr(\mathbf{\Phi_f \Phi_f^*}) = p_0 \tag{A.31}$$

The solution is obtained as $\mathbf{\Lambda \Phi_f \Phi_f^*} = \alpha^2 \mathbf{I}$, so that $\lambda_{min}\left(\mathbf{\Lambda \Phi_f \Phi_f^*}\right) = \lambda_{max}\left(\mathbf{\Lambda \Phi_f \Phi_f^*}\right)$. Alternatively, assuming λ_i are distinct [1] , we can write, $\mathbf{\Phi_f} = \alpha \mathbf{\Lambda}^{-1/2}$. From the power constraint we have $\alpha^2 = \frac{p_0}{\sum_{i=1}^{B} \lambda_i^{-1}}$. This concludes our proof.

A.5 Proof of Lemma 6

Denote $\mathbf{R_{zz}} := \mathbf{HFF^*H^*} + \mathbf{R_{nn}}$. Adding and subtracting $\mathbf{F^*H^*R_{zz}^{-1}HF}$ to the error covariance matrix $\mathbf{R_e}(\mathbf{G}, \mathbf{F})$

$$\begin{aligned}\mathbf{R_e}(\mathbf{G},\mathbf{F}) &= (\mathbf{G} - \mathbf{F^*H^*R_{zz}^{-1}})\mathbf{R_{zz}}(\mathbf{G} - \mathbf{F^*H^*R_{zz}^{-1}})^* + \mathbf{I} - \mathbf{F^*H^*R_{zz}^{-1}HF} \\ &\geq \mathbf{I} - \mathbf{F^*H^*R_{zz}^{-1}HF} = (\mathbf{I} + \mathbf{F^*H^*R_{nn}^{-1}HF})^{-1}\end{aligned} \tag{A.32}$$

The lower bound is achieved if and only if $\mathbf{G} - \mathbf{F^*H^*R_{zz}^{-1}} = 0$. We then get from

[1]For the case when $\mathbf{\Lambda}$ has repeated diagonal elements λ_i, we can consider $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda} + \mathbf{\Delta_\Lambda}$ where $\mathbf{\Delta_\Lambda}$ is a perturbation matrix that ensures that λ_i are distinct. Since the optimal design matrices $\mathbf{G}$ and $\mathbf{F}$ are continuous functions of $\mathbf{\Lambda}$, we have $\lim_{\mathbf{\Delta_\Lambda} \to 0} \mathbf{G}(\tilde{\mathbf{\Lambda}}) = \mathbf{G}(\mathbf{\Lambda})$ and similarly for $\mathbf{F}$.

(A.32)

$$\mathbf{G} = \mathbf{F}^*\mathbf{H}^*(\mathbf{H}\mathbf{F}\mathbf{F}^*\mathbf{H}^* + \mathbf{R_{nn}})^{-1} \tag{A.33}$$

$$\mathbf{R_e}(\mathbf{F}) = (\mathbf{I} + \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}^{-1}}\mathbf{H}\mathbf{F})^{-1} \tag{A.34}$$

Consider the most general expression for $\mathbf{F}$

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f} + \tilde{\mathbf{V}}\tilde{\mathbf{\Phi}}_\mathbf{f} = \mathbf{F}_\parallel + \mathbf{F}_\perp \tag{A.35}$$

where $\mathbf{F}_\parallel$ is in the range space of $\mathbf{R_{nn}^{-1/2}}\mathbf{H}$, or equivalently $\mathbf{\Lambda}^{1/2}\mathbf{V}^*$ and $\mathbf{F}_\perp$ is in the null-space of $\mathbf{R_{nn}}^{-1/2}\mathbf{H}$ or equivalently $\mathbf{\Lambda}^{1/2}\mathbf{V}^*$; $\tilde{\mathbf{\Phi}}_\mathbf{f}$ is any $(M_T - B) \times B$ matrix, and $\mathbf{\Phi_f}$ is any $B \times B$ matrix.

We can verify that $\mathbf{R_e}(\mathbf{F}) = \mathbf{R_e}(\mathbf{F}_\parallel)$. and $Tr(\mathbf{F}\mathbf{F}^*) = Tr(\mathbf{F}_\parallel\mathbf{F}_\parallel^* + \mathbf{F}_\perp\mathbf{F}_\perp^*) \geq Tr(\mathbf{F}_\parallel\mathbf{F}_\parallel^*)$ with equality if and only if $\mathbf{F}_\perp = 0$. Hence without any loss of generality, we can let $\mathbf{F}_\perp = 0$. Hence we have $\mathbf{F} = \mathbf{V}\mathbf{\Phi_f}$. To show that $\mathbf{\Phi_f}$ is diagonal, we can use the EVD of $\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}$ in (A.34) to get

$$\begin{aligned} det[\mathbf{R_e}(\mathbf{F})^{-1}] = det(\mathbf{I} + \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}^{-1}}\mathbf{H}\mathbf{F}) &= det(\mathbf{I} + \mathbf{\Lambda^{1/2}}\mathbf{\Phi_f}\mathbf{\Phi_f^*}\mathbf{\Lambda^{1/2}}) \\ &\leq \prod_{i=1}^{B}[1 + (\mathbf{\Phi_f}\mathbf{\Phi_f^*})_{ii}\lambda_i]) \end{aligned} \tag{A.36}$$

where the above inequality is the *Hadamard inequality* [2]. Assuming λ_i are distinct, the upper bound in (A.36) is attained if and only if $\mathbf{\Phi_f}$ is diagonal[3].

Using the matrix inversion lemma, the expression for $\mathbf{G}$ in (A.33) can be written as:

$$\mathbf{G} = \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} - \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{HF}[\mathbf{I} + \mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{HF}]^{-1}\mathbf{F}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} \quad \text{(A.37)}$$

Using the EVD of $\mathbf{H}^*\mathbf{R_{nn}}^{-1}\mathbf{H}$ and $\mathbf{F} = \mathbf{V\Phi_f}$ the above equation can be simplified as:

$$\mathbf{G} = [\mathbf{I} - \mathbf{\Phi_f}^2\mathbf{\Lambda}(\mathbf{I} + \mathbf{\Phi_f}^2\mathbf{\Lambda})^{-1}]\mathbf{\Phi_f}^*\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} \quad \text{(A.38)}$$

where we have used the fact that $\mathbf{V}^*\mathbf{V} = \mathbf{I}$. Letting $\mathbf{\Phi_g} := [\mathbf{I} - \mathbf{\Phi_f}^2\mathbf{\Lambda}(\mathbf{I} + \mathbf{\Phi_f}^2\mathbf{\Lambda})^{-1}]\Phi_f^*$, we finally get:

$$\mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R_{nn}}^{-1} \quad \text{(A.39)}$$

[2] For an $m \times m$ matrix A, $det(A) \leq \prod_{i=1}^{m} a_{ii}$

[3] For the case when some λ_i are identical, we can introduce a perturbation $\delta\lambda_i$ so that $\lambda_i + \delta\lambda_i$ are distinct and take the limit $\delta\lambda_i \to 0$.

To compute the optimal power allocation policy, we first form the Lagrangian,

$$
\begin{aligned}
L' &= -det[\mathbf{R_e}(\mathbf{F})^{-1}] + \mu(\sum_{i=1}^{B} \phi_{f,i}^2 - p_0) \\
&= \prod_{i=1}^{B} [1 + (\mathbf{\Phi_f \Phi_f^*})_{ii}\lambda_i] + \mu(\sum_{i=1}^{B} \phi_{f,i}^2 - p_0)
\end{aligned} \tag{A.40}
$$

Setting $\frac{\partial L'}{\partial \phi_{f,i}^2} = 0$, we get the power allocation policy. This concludes our proof.

A.6 Proof of Lemma 7

The cost function to be maximized is:

$$
J = det(\mathbf{I} + \frac{\mathbf{R}_{a,T}^{1/2}\mathbf{FEE^*F^*}\mathbf{R}_{a,T}^{1/2}}{4\sigma^2}) \tag{A.41}
$$

Using the SVD of $\mathbf{R}_{a,T}^{1/2}$ and EVD of $\mathbf{E}$, equation (A.41) can be re-written as:

$$
\begin{aligned}
J &= det(\mathbf{I} + \frac{1}{4\sigma^2}\mathbf{U_r \Lambda_r V_r^* F V_e \Lambda_e V_e^* F^* V_r \Lambda_r U_r^*}) \\
&= det(\mathbf{I} + \frac{1}{4\sigma^2}\mathbf{\Lambda_r V_r^* F V_e \Lambda_e V_e^* F^* V_r \Lambda_r}) \\
&\leq \prod_{i=1}^{B} [1 + \frac{1}{4\sigma^2}(\mathbf{\Lambda_r V_r^* F V_e \Lambda_e V_e^* F^* V_r \Lambda_r})_{ii}]
\end{aligned} \tag{A.42}
$$

where the last inequality is due to the *Hadamard inequality.*

The upper bound is achieved if we choose $\mathbf{F}_{opt} = \mathbf{V_r \Phi_f V_e^*}$ in which case the cost function reduces to

$$J = det[\mathbf{I} + \frac{\mathbf{\Lambda_r^2 \Lambda_e \Phi_f^2}}{4\sigma^2}] = \prod_{i=1}^{B} [1 + \frac{\lambda_{e,i}\lambda_{r,i}^2\phi_{f,i}^2}{4\sigma^2}] \tag{A.43}$$

Now for any $\bar{N} \leq B$, it holds that

$$\begin{aligned} \prod_{i=1}^{\bar{N}} (1 + \frac{\lambda_{e,i}\lambda_{r,i}^2\phi_{f,i}^2}{4\sigma^2})^{1/\bar{N}} &= \prod_{i=1}^{\bar{N}} (\frac{\lambda_{e,i}\lambda_{r,i}^2}{4\sigma^2})^{1/\bar{N}} \ \prod_{i=1}^{\bar{N}} (\frac{4\sigma^2}{\lambda_{e,i}\lambda_{r,i}^2} + \phi_{f,i}^2)^{1/\bar{N}} \\ &\leq \prod_{i=1}^{\bar{N}} (\frac{\lambda_{e,i}\lambda_{r,i}^2}{4\sigma^2})^{1/\bar{N}} \ \frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} (\frac{4\sigma^2}{\lambda_{e,i}\lambda_{r,i}^2} + \phi_{f,i}^2) \end{aligned} \tag{A.44}$$

Let $\frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} (\frac{4\sigma^2}{\lambda_{e,i}\lambda_{r,i}^2} + \phi_{f,i}^2) = \gamma$. The upper bound is achieved if and only if $(\frac{4\sigma^2}{\lambda_{e,i}\lambda_{r,i}^2} + \phi_{f,i}^2) = \gamma$ are constant for all $i = 1, 2, ..\bar{N}$. Equivalently, we have that

$$\mathbf{\Phi_f}^2 = (\gamma \mathbf{I} - 4\sigma^2 \Lambda_r^{-2} \Lambda_e^{-1})_+ \tag{A.45}$$

where $(.)_+$ means that only non-negative values are allowed. This concludes our proof.

Bibliography

[1] H. Sampath, P. Stoica and A. Paulraj. "Generalized Linear Precoder and Equalizer design for MIMO channels using the weighted MMSE criterion." *IEEE Transactions on Communications*, accepted, April 2001.

[2] H. Sampath, P. Stoica and A. Paulraj. "Generalized Linear Precoder and Equalizer design for MIMO channels using the weighted MMSE criterion." *Proceedings of the Asilomar conference on signals, systems and computers*, Monterey, CA, 2000.

[3] A. Paulraj and H. Sampath. "Space-Time Wireless Communications (a.k.a Smart Antennas)." *Managing Telecommunications and Networking Technologies in the 21st Century: Issues and Trends*, Idea Group Publishing, 2001.

[4] H. Sampath and A. Paulraj. "Precoding for space-time coded systems assuming knowledge of transmit antenna correlations." In preparation.

[5] H. Sampath and A. Paulraj. “Joint Transmit and Receive Optimization for high data rate wireless communications using multiple antennas”. *Proceedings of the Asilomar conference on signals, systems and computers*, vol.1, Monterey, CA, 1999.

[6] A. Scaglione, P. Stoica, G.B. Giannakis, S. Barbarossa and H. Sampath. “Optimal Designs for Space-Time Linear Precoders and Equalizers.” *IEEE Transactions on Signal Processing*, submitted, October 2000.

[7] H. Sampath, H. Bolcskei and A. Paulraj. ”Pre-Equalization for MIMO channels with delay spread.” *IEEE Vehicular Technology Conference*, vol.3, Boston, 2000.

[8] H. Sampath and A. Paulraj. ”Space-Time Processing TDMA Wireless Testbed.” *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol.4, Phoenix, March 1999.

[9] H. Sampath and A. Paulraj. ”Achievable rate regions for spatial multiplexing system using the MMSE criterion.” *IEEE International Symposium on Information Theory*, Sorrento, Italy, June, 2000.

[10] S.M. Kay. ”Fundamentals of statistical signal processing.” vol 2, *Detection Theory*, Prentice Hall, 1993.

[11] H. Bölcskei, "Oversampled Filter Banks and Predictive Subband Coders." *PhD thesis*, Vienna University of Technology, Nov. 1997.

[12] J. Yang and S. Roy. "Joint Transmitter-Receiver Optimization for Multi-Input Multi-Output Systems with Decision Feedback." *IEEE Trans. Info. Theory*, vol. 40, No. 5, September 1994.

[13] L. Vandendorpe, J. Louveaux, B. Maison and A. Chevreuil. "About the Asymptotic Performance of MMSE MIMO DFE for Filter-Bank Based Multicarrier Transmission." *IEEE Trans Comm*, vol 47, No. 10, October 1999.

[14] H. Lutkepohl. "Handbook of matrices." John Wiley, 1996.

[15] V. Tarokh, N. Seshadri and A.R. Calderbank. "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction." *IEEE Transactions on Information Theory*, vol 44, No.2, March 1998.

[16] V. Tarokh, A.F. Naguib, N. Seshadri and A.R.Calderbank. "Combined Array Processing and Space-Time Coding." *IEEE Transactions on Information Theory*, vol.45, No.4, May 1999.

[17] V. Tarokh and H. Jafarkhani. "A differential detection scheme for transmit diversity." *IEEE Journal on Selected Areas in Communications*, vol. 18 7, July 2000.

[18] A.F. Naguib, V. Tarokh, N. Seshadri and A.R. Calderbank. "A space-time coding modem for high-data-rate wireless communications." *IEEE Journal on elected Areas in Communications*, vol. 16 8 , Oct. 1998.

[19] V. Tarokh, H. Jafarkhani and A.R Calderbank. " Space-time block coding for wireless communications: performance results." *IEEE Journal of Sel. Areas in Comm*, vol. 17 3 , March 1999.

[20] V. Tarokh, H. Jafarkhani and A.R Calderbank. "Space-time block codes from orthogonal designs." *IEEE Trans. Info Theory*, vol. 45 5 , July 1999.

[21] E. Telatar. "Capacity of Multi-Antenna Gaussian Channels." *Technical Memorandum*, ATT Bell Laboratories, 1995.

[22] J.K. Tugnait. "FIR Inverses To MIMO Rational Transfer Functions With Application to Blind Equalization." *Thirtieth Asilomar Conference on Signals, Systems and Computers*, vol. 1, 1997.

[23] H. Olofsson, M. Almgren, M. Hook. "Transmitter diversity with antenna hopping for wireless communication systems." *IEEE 47th Vehicular Technology Conference*, vol. 3, 1997.

[24] R.W. Heath., Jr and A. Paulraj. "A simple scheme for transmit diversity using partial channel feedback." *Asilomar Conference on Signals, Systems and Computers*, Vol 2, 1998.

[25] R.W. Heath Jr and A. Paulraj. "Switching between Diversity and Multiplexing." *IEEE Communication Letters*, submitted.

[26] D.Gore, R.Nabar and A.Paulraj, "Selecting an Optimal Set of Transmit Antennas for a Low Rank Matrix Channel." *IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, June 2000.

[27] R.Nabar. D.Gore and A.Paulraj, "Selection and Use of Optimal Transmit Antennas in Wireless Systems." ICT, Acapulco, Mexico, May 2000.

[28] G. Ganesan and P. Stoica. "Space-Only and Space-Time Processing for Wireless Channels." *Proceedings of RVK*, Karlskrona, Sweden, June 1999.

[29] J.H. Winters. "The diversity gain of transmit diversity in wireless systems with Rayleigh fading." *IEEE Transactions on Vehicular Technology*, vol. 47 1, Feb. 1998.

[30] E. Lindskog and A. Paulraj. "A transmit diversity scheme for channels with intersymbol interference." *IEEE International Conference on Communications*, vol. 1 , 2000.

[31] A. Wittneben "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation." *IEEE International Conference on Communications*, Geneva, Volume: 3 , 1993.

[32] B.M. Hochwald and T.L Marzetta. "Unitary Space-Time modulation for multiple-antenna communications in Rayleigh fading channel." *IEEE Trans. Information Theory*, submitted.

[33] T.L. Marzetta and B.M. Hochwald. "Capacity of a mobile multiple-antenna communication link in Rayleigh flat-fading." *IEEE Trans Inform. Theory*, vol. 45, no. 1, Jan 1999.

[34] N. Seshadri and J.H. Winters. "Two schemes for improving the performance of frequency-division duplex (FDD) transmission systems using transmitter antenna diversity." *Int. Journal Wireless Inform. Networks*, vol. 1, no. 1, pp. 49-60, Jan 1994.

[35] G.J. Foschini and M.J. Gans. "On the limits of wireless communication in a fading environment when using multiple antennas." *Wireless Personal Communications*, vol 6, no.3, March 1998.

[36] R. Negi, A.M. Tehrani and J. Cioffi. "Adaptive antennas for space-time coding over block-time invariant multipath fading channels." *IEEE Vehicular Technology Conference*, vol. 1 , 1999.

[37] A.M Tehrani, R. Negi and J. Cioffi. "Space-time coding and transmission optimization for wireless channels." *Asilomar Conference on Signals, Systems and Computers*, vol. 2 , 1998.

[38] Jiann-Ching Guey; Fitz, M.P.; Bell, M.R.; Wen-Yi Kuo. "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels." *IEEE Transactions on Communications*, vol. 47, April 1999.

[39] S.M. Alamouti. "A simple transmit diversity technique for wireless communications." *IEEE Journal on Selected Areas of Communications*, vol 16, No.8, Oct. 1998.

[40] R.H. Roy. "Spatial division multiple access technology and its application to wireless communication systems." *IEEE 47th Vehicular Technology Conference* vol: 2 , 1997.

[41] D. Gerlach and A. Paulraj. "Adaptive transmitting antenna arrays with feedback." *IEEE Signal Processing Letters*, vol. 1 10 , Oct. 1994.

[42] G.W. Wornell and M.D. Trott. "Efficient signal processing techniques for exploiting transmit antenna diversity on fading channels." *IEEE Trans. on Signal Processing*, vol. 45 1 , Jan. 1997.

[43] A. Paulraj and T. Kailath. "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)." US Patent No. 5,345,599, 1993.

[44] . A.J.Paulraj and C.B. Papadias. "Space-time processing for wireless communications." *IEEE Signal Processing Magazine*, vol. 14 6 , Nov. 1997.

[45] G.J. Foschini. "Layered Space-Time Architecture for Wireless Communication in a Fading Environment when using Multi-Element Antennas." *Bell Labs Technical Journal*, Autumn 1996.

[46] W. Jang, B.R. Vojcic, R.L. Pickholtz. "Joint Transmitter-Receiver Optimization in Synchronous Multiuser Communications over Multi path channels." *IEEE Transactions on Communications, vol.46, No 2. February 1998*

[47] A. Scaglione, S. Barbarossa and G.B. Giannakis. "Filter bank Transceivers Optimizing Information Rate in Block Transmissions over Dispersive Channels." *IEEE Transactions on Information Theory*, vol 45, No.3, April 1999.

[48] A. Scaglione, G.B. Giannakis and S. Barbarossa. "Redundant Filter bank Precoders and Equalizers Part I: Unification and Optimal Designs." *IEEE Transactions on Signal Processing*, vol 47, No.7, July 1999.

[49] D. Agarwal, V. Tarokh, A. Naguib and N. Seshadri. "Space-Time Coded OFDM for High Data-Rate Wireless Communications Over Wide band Channels." *48th IEEE Vehicular Technology Conference*, vol.3, May 1998.

[50] B.L. Hughes. "Differential space-time modulation." *IEEE Trans. on Information Theory*, vol. 46 7 , November 2000.

[51] G.D. Dudevoir, J.S. Chow, J.M. Cioffi and S. Kasturia. "Vector coding for T1 transmission in the ISDN subscriber loop." *IEEE Int. Conf. Communications*, Boston, MA, 1989.

[52] J.R. Magnus and H. Neudecker. "Matrix Differential Calculus With Applications in Statistics and Econometrics." Wiley Publications, 1998.

[53] S. Boyd and L. Vandenberghe. "Course Reader: Introduction to Convex Optimization with Engineering Applications." Stanford University.

[54] T. Cover and J. Thomas. "Elements of Information Theory." New York : Wiley, 1991.

[55] Goldsmith. A. and Chua S.G. "Adaptive Coded Modulation for Fading Channels." *IEEE Transactions on Communications*, pp 595-602, May 1998.

[56] J. Yang and S. Roy. "On Joint Transmitter and Receiver Optimization for Multiple-Input-Multiple-Output (MIMO) Transmission Systems." *IEEE Transactions on Communications*, vol 42. No 12, December 1994.

[57] W.C. Jakes. "Micorwave Mobile Communications." IEEE Press, 1993.

[58] T.S Rappaport. "Wireless Communications - Principles and Practice." IEEE Press, 1996.

[59] A.J.Rustako, Jr., N. Amitay, G.L. Owens, and R.S. Roman. "Radio propagation at microwave frequencies for line-of-sight microcellular mobile and personal communications." *IEEE Trans. Vehic. Technol.*, Vol VT-40, No. 1, Feb. 1991.

[60] W.C.Y. Lee. "Mobile Communications Engineering." New York: Mc-Graw-Hill, 1982.

[61] A.J. Goldsmith and L.J. Greenstein. "A measurement-based model for predicting coverage areas of urban microcells." *IEEE J. Selected Areas Commun.*, vol SAC-11, No. 7, Sept. 1993.

[62] . L.J. Greenstein, D.G. Michelson, and V. Erceg. "Moment-Method Estimation of the Ricean K-Factor." *IEEE Communication Letters*, vol. 3 no.6, 1999.

[63] . L.J. Greenstein, S. Ghassemzadeh, V. Erceg and D.G.Michelson. "Ricean K-factors in narrowband fixed wireless channels." *Proceedings of WPMC*, Amsterdam, 1999.

[64] D.S. Baum, D. Gore, R. Nabar, S. Panchanathan, K.V.S Hari, V. Erceg and A.J. Paulraj. "Measurement and characterization of broadband MIMO Fixed Wireless Channels at 2.5 GHz." *Proceedings of ICWPC*, Hyderabad, India, 2000.

[65] W. Mohr. "Radio propagation for local loop applications at 2 GHz." *Proceedings of ICUPC*. 1994, pp.119-123.

[66] V.Erceg, D.G. Michelson, S.S Ghassemzadeh, L.J. Greenstein, et al. "A model for the multipath delay profile of fixed wireless channels." *IEEE Journal on Selected Areas in Communications*, vol. 17, no.3, 1999.

[67] V.Erceg, L.J.Greenstein, S.Y. Tjandra, S.R. Parkoff, et al. "An empirically based path loss model for wireless channels in suburban environments." *IEEE Journal of Selected Areas in Communications*, vol. 17, no. 7, 1999.

[68] M. Hata. "Empirical formula for propagation loss in land mobile radio services." *IEEE Trans. Vehic. Technol.*, vol VT-29, No.3, pp. 317-325, Aug. 1980.

[69] V. Erceg et al. "MIMO channel measurement campaign at Stanford University." 1999, 2000.

[70] G. Raleigh and J. Cioffi. "Spatio-Temporal Coding for Wireless Communications." *IEEE Transactions on Communications*,vol 46. No3, March 1998.

[71] H. Bolcskei, D. Gesbert, and A. Paulraj. "On the capacity of wireless systems employing OFDM-based spatial multiplexing." *IEEE Transactions on Comm*, accepted.

[72] K.Sheikh, D.Gesbert, D.Gore and A.Paulraj, "Smart Antennas for Broadband Wireless Access Networks." *IEEE Communications Magazine*, Nov'99.

[73] D. Shiu and J.M. Kahn. "Power Allocation Strategies for Wireless Systems with Multiple Transmit Antennas." *IEEE Trans. Commun.*, in revision.

[74] D. Shiu, J.G. Foschini, M. Gans, J.M. Kahn, "Fading Correlation and Its Effect on the Capacity of Multi-Element Antenna Systems." *IEEE Trans. Communications*, to appear.

[75] E. Biglieri, J. Proakis, and S. Shamai. "Fading channels: information-theoretic and communications aspects." IEEE Trans. Info. Theory, pp. 2619–2692, Oct. 1999.

[76] B. Hochwald and W. Sweldens. "Differential unitary space time modulation." Tech. rep., Bell Laboratories, Lucent Technologies, Mar. 1999.

[77] J.G. Proakis. *Digital Communications*, 2nd. ed. New York: McGraw-Hill, 1989.

[78] A. Paulraj, C. Papadias and D. Gesbert. "Space-Time Processing for Wireless Communication." preprint, 1998.

[79] S.P Wu and S. Boyd. "sdpsol: A parser/solver for semidefinite programs with matrix structure." *In Recent Advances in LMI Methods for Control*, SIAM,chapter 4, pp.79-91, 2000.

[80] L. Vandenberghe and S. Boyd. "Determinant maximization with linear matrix inequality constraints." *SIAM Journal on Matrix Analysis and Applications*, 19(2):499-533, 1998.

[81] S. Boyd and L. El Ghaoui and E. Feron and V. Balakrishnan "Linear Matrix Inequalities in System and Control Theory." *SIAM Studies in Applied Mathematics*, Vol. 15, Society of Industrial and Applied Mathematics (SIAM), 1994.

[82] *Proceedings of the Smart Antenna Workshop*, Stanford University, 1994-1999.