

Graduate Theses, Dissertations, and Problem Reports

2004

# Transmitter precoding for multi-antenna multi-user communications

Kirtan N. Modi West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

#### **Recommended Citation**

Modi, Kirtan N., "Transmitter precoding for multi-antenna multi-user communications" (2004). *Graduate Theses, Dissertations, and Problem Reports.* 1501. https://researchrepository.wvu.edu/etd/1501

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

## Transmitter Precoding for Multi-Antenna Multi-User Communications

by

Kirtan N. Modi

Thesis submitted to the College of Engineering and Mineral Resources at West Virginia University in partial fulfillment of the requirements for the degree of

> Master of Science in Electrical Engineering

Daryl S. Reynolds, Ph.D., Chair Matthew C. Valenti, Ph.D. Erdogan Gunel, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia 2004

Keywords: CDMA, MIMO, Transmitter Precoding, Ultra-low Complexity Receivers

Copyright 2004 Kirtan N. Modi

#### Abstract

#### Transmitter Precoding for Multi-Antenna Multi-User Communications

by

Kirtan N. Modi Master of Science in Electrical Engineering

#### West Virginia University

#### Daryl S. Reynolds, Ph.D., Chair

Emerging wireless sensor networks and existing wireless cellular and *ad hoc* networks motivate the design of low-power receivers. Multi-user interference drastically reduces the energy efficiency of wireless multi-user communications by introducing errors in the bits being detected at the receiver. Interference rejection algorithms and multiple antenna techniques can significantly reduce the bit-error-rate at the receiver. Unfortunately, while interference rejection algorithms burden the receiver with heavy signal processing functionalities, thereby increasing the power consumption at the receiver, the small size of receivers, specifically in sensor networks and in downlink cellular communications, prohibits the use of multiple receive antennas. In a broadcast channel, where a central transmitter is transmitting independent streams to decentralized receivers, it is possible for the transmitter to have *a priori* knowledge of the interference. Multiple antennas can be used at the transmitter to enhance energy efficiency. In some systems, the transmitter has access to virtually an infinite source of power. A typical example would be the base station transmitter for the downlink of a cellular system. The power consumption at receivers can be reduced if some of the signal processing functionality of the receiver is moved to the transmitter.

In this thesis, we consider a wireless broadcast channel with a transmitter equipped with multiple antennas and having *a priori* knowledge of interference. Our objective is to minimize the receiver complexity by adding extra signal processing functions to the transmitter. We need to determine the optimal signal that should be transmitted so that interference is completely eliminated, and the benefits that can be obtained by using multiple transmit antennas can be maximized. We investigate the use of linear precoders, linear transformations made on the signal before transmission, for this purpose.

## Acknowledgments

I am pleased to acknowledge my indebtedness to my advisor Dr. Daryl Reynolds for his support, guidance, patience and collaboration during this research. His constructive criticisms on the thesis have helped me make the presentation more clear. The encouragement and the freedom of exploration which he gave me during my stay at WVU has enabled me to grow as a researcher.

I would like to thank Dr. Matthew Valenti for encouraging me while I was working on this thesis and for providing valuable comments on the thesis. I have learnt most of my communication theory from him and it has been very pleasing to use it in my research. I am grateful to Dr. Erdogan Gunel for being on my committee. I have learnt a lot of statistics from him and it has been very useful in my research.

In addition, I would like to acknowledge the help I got from my friends, Amit Sangle and Anirudha Vaidya, while I was writing this thesis. I am indebted to my friend Vivek Balasubramanian for making me laugh everyday during my stay at WVU.

I would like to thank my brother Mitul Modi, sister Jyoti Modi, and my grandfather Bhukhandas Jariwala for their support. Above all, I am thankful to my mother Uravashi Modi, and my father Narendra Modi for all the years of love and caring. My M.S. is nothing but a realization of their dreams and a result of their hardwork, belief and love.

## Contents

A	Acknowledgments iii				
Li	st of	f Figures	vi		
$\mathbf{Li}$	st of	Tables v	iii		
N	otati	ion and Acronyms	ix		
1	$\mathbf{M}\mathbf{u}$	lti-Antenna Multi-User Wireless Communications and the Motivation for			
	Pre	ecoding	1		
	1.1	The Additive White Gaussian Noise Channel Model	3		
		1.1.1 Match Filter Demodulator and the Detection Process	6		
		1.1.2 Bit-Error-Rate of BPSK in an AWGN Channel	10		
	1.2	The Wireless Channel	12		
		1.2.1 Large-Scale Fading: Path Loss Models	15		
		1.2.2 Bandlimited Channel Impulse Response	16		
		1.2.3 Small Scale Fading	17		
		1.2.4 Discrete-Time Statistical Model for a Slowly Fading Wireless Channel	21		
		1.2.5 Performance of BPSK in the presence of Fading	23		
	1.3	Multi-Antenna Communications	25		
		1.3.1 Diversity Techniques	26		
		1.3.2 Diversity via Multi-Antenna Communications	30		
		1.3.3 Precoding Techniques for point to point MIMO communications	31		
	1.4	Multi-User Communications and CDMA	32		
		1.4.1 Essentials of CDMA	34		
	1.5	Precoding for Interference Cancellation in Single and Multi-user Communications	44		
	1.6	Problem Formulation: Precoding for Multi-Antenna Multi-User Communications	46		
<b>2</b>	Pre	ecoding for Flat Fading Channels	49		
	2.1	Introduction	51		
	2.2	Multi-Antenna Technique with CSI at the Transmitter	53		
	2.3	System Model	56		
		2.3.1 Maximum Ratio Weighting and Precoder Design	57		
	2.4	Precoding for Orthogonal Spreading	59		
	2.5	Precoding for Non-Orthogonal Spreading	61		
	2.6	Performance Characterization of Precoding for Flat Fading Channels	63		
	2.7	Conclusion	70		

#### CONTENTS

3	Pre	coding with Partial Channel Information	71			
	3.1	Introduction	72			
	3.2	Precoders with Partial Channel Information	73			
	3.3	Special Case: Partial Channel Information as Old Channel Estimates	75			
	3.4	Performance of Precoding with Partial CSI as Old Channel Estimates	77			
		3.4.1 Simulation Setup	79			
		3.4.2 Results	82			
	3.5	Conclusion	85			
4	Precoding for Frequency Selective Channels 86					
	4.1	Introduction	86			
	4.2	System Model	93			
	4.3	Precoding for Frequency Selective Channels without InterSymbol Interference	96			
		4.3.1 Individually Optimal Precoders	96			
		4.3.2 Jointly Optimal Precoders	102			
	4.4	Precoding for Frequency Selective Channels with Intersymbol Interference 1	105			
	4.5	Performance of Precoding for Frequency Selective Channels	111			
	4.6	Conclusion	115			
<b>5</b>	Sun	nmary, Conclusion and Future Research 1	116			
5 Re	Sun efere	nmary, Conclusion and Future Research 1 nces 1	L16 L18			
5 Re	Sun efere: A	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables 1	116 118 125			
5 Re	Sun efere: A	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables 1       1         A.1       Multidimensional Gaussian Distribution	L16 L18 125 126			
5 Re	Sum efere: A	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1	<b>116</b> 125 126 126			
5 Re	Sun efere A	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1	<b>116</b> 125 126 126 126			
5 Re	Sun efere: A	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1	<b>116</b> 125 126 126 126 126			
5 Re	Sum efere: A	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1	L16 L18 125 126 126 126 128 129			
5 Re	Sun efere: A B	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1	<b>L16</b> 125 126 126 126 128 129 131			
5 Re	Sun efere: A B	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1         B.1       Proof for Condition on Power Scaling Factor       1	L16 L18 125 126 126 126 128 129 131 132			
5 Re	Sun efere: A B C	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         B.1       Proof for Condition on Power Scaling Factor       1         Optimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI       1	<b>L16</b> <b>L18</b> 125 126 126 128 129 131 132 132			
5 Re	Sun efere: A B C	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         B.1       Proof for Condition on Power Scaling Factor       1         Dytimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI 1       1         C.1       Equi-correlated Spreading Sequences       1	<b>L16</b> <b>L18</b> 125 126 126 128 129 131 132 132			
5 Re	Sun efere: A B C D	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1         B.1       Proof for Condition on Power Scaling Factor       1         Optimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI 1       1         C.1       Equi-correlated Spreading Sequences       1         Proof for Partial Channel Information       1	<b>L16</b> <b>L18</b> 125 126 126 126 128 129 131 132 135 135			
5 Re	Sun fere: A B C D E	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1         B.1       Proof for Condition on Power Scaling Factor       1         Optimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI 1       1         C.1       Equi-correlated Spreading Sequences       1         Proof for Partial Channel Information       1         Discrete-Time Model for CDMA Communication over Frequency Selective Channel       1	<b>L16</b> <b>L18</b> 125 126 126 128 129 131 132 132 135 136 137			
5 Re	Sun efere: A B C D E	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1         B.1       Proof for Condition on Power Scaling Factor       1         Optimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI I       1         C.1       Equi-correlated Spreading Sequences       1         Proof for Partial Channel Information       1       1         Discrete-Time Model for CDMA Communication over Frequency Selective Channel       1         E.1       Received signal with Precoding       1	<b>L16</b> <b>L18</b> 125 126 126 128 129 131 132 132 135 136 137 140			
5 Re	Sun fere: A B C D E F	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1         B.1       Proof for Condition on Power Scaling Factor       1         Dytimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI 1       1         C.1       Equi-correlated Spreading Sequences       1         Proof for Partial Channel Information       1         Discrete-Time Model for CDMA Communication over Frequency Selective Channel       1         E.1       Received signal with Precoding       1         Signal Model for Jointly Designed Precoders for Channels without ISI       1	L16 L18 125 126 126 128 129 131 132 132 135 136 137 140 142			
5 Re	Sun efere: A B C D E F	nmary, Conclusion and Future Research       1         nces       1         Gaussian Random Variables and Functions of Gaussian Random Variables       1         A.1       Multidimensional Gaussian Distribution       1         A.2       Conditional Gaussian Distribution       1         A.3       The Q-Function       1         A.4       Rayleigh Random Variable       1         A.5       Expectation of a Q-function with a First-Order Rayleigh Argument       1         Some properties of a Hermitian Matrix       1         B.1       Proof for Condition on Power Scaling Factor       1         Dytimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI II       1         C.1       Equi-correlated Spreading Sequences       1         Proof for Partial Channel Information       1       1         Discrete-Time Model for CDMA Communication over Frequency Selective Channel       1         E.1       Received signal with Precoding       1         Signal Model for Jointly Designed Precoders for Channels without ISI       1         F.1       Constant for Peak to Average Power Ratio       1	L16 L18 125 126 126 126 128 129 131 132 132 135 136 137 140 142 142			

## List of Figures

1.1	Figure illustrates the BER vs SNR performance of BPSK in an AWGN channel and a Rayleigh fading channel	25
1.2	Figure illustrates the BER $vs$ SNR performance of a CDMA downlink system with $K = 10$ users over a Rayleigh fading channel. The spreading code of each user is equally correlated to that of all other users with correlation $\rho = 0.7$ . The receivers are constrained to be simple match filter receivers.	44
2.1	Figure illustrates that full diversity, as indicated by the slope of the BER curve, is achieved using precoding with perfect CSI at transmitter for flat fading channels. There will be a loss in SNR resulting from $\alpha \neq 1$ , which corresponds to the case of non-orthogonal spreading. We have used $\alpha = 0.4$ and $\alpha = 0.1$ and $M_T = 1, 2, 4$ antennas.	68
2.2	Figure illustrates the variation in diversity gain with $\rho$ when perfect CSI is available at the transmitter. There will be a loss in SNR resulting from $\rho \neq 0$ , which corresponds to the case of orthogonal spreading. We have used $\rho = 0.0, 0.4, 0.9$ and	
<u></u>	$M_T = 2,4$ transmit antennas for a system with $K = 4,16$ users	69 60
2.3 2.4	Figure illustrates the variation of $\alpha$ with $\rho$ for $K = 2, 4, 8, 10, 52, \infty$ users Figure illustrates the variation in the loss in SNR = $10 \log_{10} \alpha^2$ with $\rho$ for $K = 2, 4, 8, 16, 32, \infty$ users	09 70
3.1	Figure illustrates the correlation between the past and perfectly known channel and the actual channel as a function of different Doppler frequencies $f_d = 20, 50, 80$ Hz.	78
3.2	Figure illustrates the correlation between the past and perfectly known channel and	70
3.3	the actual channel as a function of the normalized Doppler frequency $f_d \Delta t$ . Figure shows the simulation result for precoding when $M_T = 2$ transmit antennas are used for $K = 10$ users and $\rho = 0.4$ . Different values of $f_d \Delta t$ are used. Performance of MRW with perfect channel information at the transmitter for a single user system	79
3.4	with $M_T = 1, 2$ transmit antennas is shown for comparison	83 83
4.1	Figure shows the filter response of a 2-tap channel, the corresponding pre-Rake filter response and the effective channel output. The scaling factor $\gamma = \frac{1}{\sqrt{ h[1] ^2 +  h[2] ^2}}$ . The taps are spaced $T_c$ seconds apart.	89
	The second s	

- 4.3 Figure illustrates performance achieved with precoders derived individually for each path for a channel with no ISI. *m*-sequences of length N = 15 were used with  $M_T = 2$  transmit antennas, L = 2 paths, and K = 10 users. The BER achieved by using peak to average power ratios  $\beta = 1, 2, 10, 20, 40, \infty$  is shown. The slopes corresponding to the diversity performance of a single user system with L = 1, 2, 4 diversity gain are plotted for comparison.

## List of Tables

3.1	System model for simulation	82
3.2	Maximum delay $\Delta t$ , between actual channel and available estimate, corresponding	
	to $f_d \Delta t < 0.05$ , for different velocities and frequency bands	84

## Notation and Acronyms

We use the following notation and symbols throughout this thesis. Matrices will be represented in bold upper case letters and vectors in bold lower case letters. Some operations and symbols frequently used are given below:

$(\cdot)^*$	:	complex conjugation;
$(\cdot)^T$	:	matrix transpose;
$(\cdot)^H$	:	conjugate of matrix transpose;
$(\cdot)^{\dagger}$	:	pseudo-inverse of a matrix;
$E\{\cdot\} \& E_x\{\cdot\}$	:	ensemble averaging & averaging over the random variable $x$ ;
$\operatorname{vec}\{A\}$	:	column vector generated by stacking columns of matrix $\boldsymbol{A}$ ;
$[\cdot]_{p,q}$	:	element in row $p$ , column $q$ ;
$[\cdot]_{:,q}$	:	column $q$ of matrix;
$\operatorname{diag}(a_{11},\cdots,a_{KK})$	:	$\mathbb{C}^{K \times K}$ diagonal matrix with diagonal elements $a_{11}, \cdots, a_{KK};$
$<\cdot,\cdot>$	:	Euclidean inner product;
$  \cdot  $	:	Euclidean norm;
$\operatorname{tr}\{\cdot\}$	:	trace of a matrix;
$\mathcal{L}\{\cdot\}$	:	Laplace transform operator;
$\mathcal{F}\{\cdot\}$	:	Fourier transform operator;
$\Re\{x\}$	:	Real part of $x$ ;
$\{X_b^{(a)}\}_a$	:	the set $\{X_{h}^{(1)}, X_{h}^{(2)}, \cdots\};$
$\{X^{(a)}\}$	:	the set $\{X^{(1)}, X^{(2)}, \cdots\};$
$\{X_a\}$	:	the set $\{X_1, X_2, \dots\};$
$I_K$	:	$K \times K$ identity matrix;
$oldsymbol{e}_k$	:	column vector that contains all zeros but for a 1 in the $k^{th}$ position;
$\delta[x]$	:	Kroneckor delta function, defined to be
		1 when $x$ is 0 and 0 for all other values of $x$ :
P(A)	:	probability of event A;
$p_X(x)$	:	probability density function (pdf) of random variable X;
$p_e(x)$	:	probability of bit error for a signal to noise ratio (SNR) of x and
		conditioned upon the channel;
$P_e(x)$	:	average probability of error, also referred to as average bit-error-rate
		(BER) given that the average SNR is $x$ ;

K	:	number of users;
N	:	length of spreading sequences;
$oldsymbol{S} \in \mathbb{C}^{N  imes K}$	:	spreading matrix;
$oldsymbol{b} \in \mathbb{C}^{K  imes 1}$	:	information symbols, one for each of the $K$ users;
b[i]	:	$i^{th}$ element of column vector <b>b</b> ;
R	:	spreading signature cross-correlation matrix $(\mathbf{S}^H \mathbf{S})$ ;
M	:	precoding matrix;
$M_T$	:	number of transmit antennas;
$h_i^{(t)}$	:	channel coefficient from transmit antenna $t$ to user $i$ ;
n	:	uncorrelated complex Gaussian random vector, generally used for noise terms;
$\mathcal{N}_c(0,\sigma^2)$	:	complex Gaussian distribution with zero mean and variance $\sigma^2$ per dimension;
$\mathcal{N}(0,\sigma^2)$	:	real Gaussian distribution with zero mean and variance $\sigma^2$ .

## Multi-Antenna Multi-User Wireless Communications and the Motivation for Precoding

Recent research on the history of wireless communications attributes the invention of radio communication to a West Virginian dentist, Mahlon Loomis [1]. Claims are that he invented it way back in the year 1872. Developments were focused on radio telegraphy for the next few decades. The implementation of the mobile radio telephone by the Detroit police department in 1921, and the beginning of mobile services for public use by Bell Telephone Laboratories in 1946 marked the commercial beginning of mobile telephony [2]. Developments since then have been revolutionary with excellent voice services now available. With the advent of the Internet, the benefits visible from high bandwidth data services, are now driving the developments in the mobile wireless sector.

Right from the beginning of mobile communications, the size and energy efficiency of mobile receivers has received extreme focus. The inherent portability of these receivers incites lightweight and compact designs. They work on batteries which need to be replaced or recharged, thereby motivating energy efficient designs. Cellular phones presently available provide around 3 hours of talk-time before they need to be recharged. This time needs to be increased, while simultaneously bringing down the size and price of such devices to further expand the consumer base for wireless. The presence of multiple users in the communication environment leads to multi-user interference (MUI). This reduces the energy efficiency of the overall system by introducing decoding errors

 $\mathbf{2}$ 

which result in erroneous data at the receiver. Elimination of these errors necessitates frequent retransmissions, or implementation of powerful, high complexity, receiver-based signal processing techniques for interference mitigation. These in turn, increase the total power consumption at the receiver. Reduction in complexity while simultaneously keeping the error rate low will help reduce power consumption, and hence increase the operation time for mobile receivers. The evolution of sensor networks for both the commercial as well as the military sector, requires the implementation of low power consuming transceivers, further motivating the quest for reduced complexity, low power receiver design. The overall energy efficiency of transceivers can be improved by diversity communications. Diversity communication involves providing the receiver with multiple copies of the same signal. The receiver can then combine the signals received to reduce the bit-error-rate (BER), thereby further reducing the power consumption. We often refer to this technique of error reduction as diversity exploitation. One way of achieving diversity communications is by using multiple antennas at the transmitter and/or receiver.

A reduction in receiver complexity can be achieved by moving some signal processing functionality from the receiver to the transmitter. We refer to this transfer of signal processing functionality from the receiver to the transmitter as transmitter based pre-processing. Note that this will result in complicated transmitters. However, in many systems a complicated transmitter can be allowed. A typical example is the downlink of a cellular communication system. Another example would be a wireless access points, deployed in an indoor wireless local area network (LAN). The transmitter in these cases has access to virtually an infinite amount of energy available for signal processing, and hence can accommodate enhanced complexity. The receivers in these cases, cellular phones in case of the cellular downlink and mobile handheld devices, wireless access cards used in laptops in case of wireless LANs, are preferred to have low power consumption and compact size. Transmitter based pre-processing techniques would be excellent for such applications. The design of transmitter based pre-processing techniques has received significant attention in the recent past [3, 4, 5, 6, 7, 8, 9, 42]. In this thesis we will study the design and performance of one such technique called as precoding. Precoding refers to the use of a transformation, linear or non-linear, on the signal to be transmitted. Precoding techniques for multi-user single antenna systems were proposed and investigated in [3, 7, 9]. There the primary goal was interference cancellation using transmitter based pre-processing for mitigating MUI known a priori to the transmitter. Precoding for single user multi-input multi-output (MIMO) antenna communications was proposed in [42], where focus was laid on diversity exploitation. Thus, precoding techniques proposed so far have either focused on multi-user single antenna communications or single user multi-antenna communications.

In this thesis, we will focus on the use of multiple transmit antennas to provide diversity communication in a multi-user system. We will use linear precoding techniques for not only cancelling interference known *a priori* to the transmitter, but also to exploit the diversity made available by the use of multiple transmit antennas. Thereby, we will enable both, reduced receiver complexity as well as increased energy efficiency. We will constrain our receivers to have minimum complexity and often refer to them as *ultra-low complexity* receivers. We will design our precoders to jointly exploit diversity and cancel interference under this constraint on the receiver structure and a total transmit power constraint on the transmitter.

The present chapter focusses on introducing the wireless communication system model and the fundamentals of multi-antenna communications and multi-user communications which are used throughout this thesis. In Section 1.1 we begin by introducing the additive white Gaussian noise (AWGN) model for a simple single path communication channel. We then move on to building a mathematical model for a wireless communication system in Section 1.2. We will analyze the performance of binary phase shift keying (BPSK) over an AWGN and a wireless channel in these Sections. In Section 1.3 we will review the benefits of multi-antenna communications and the relevant research on precoding for point-to-point or single user multi-antenna links conducted in the past. Followed by this, in Section 1.4.1 we will briefly review multi-user communications with a focus on code-division-multiple-access (CDMA) technology. We will review precoding techniques researched in the past for single antenna CDMA communications. Finally, we will look at the factors motivating our research and list the specific contributions of this thesis.

#### 1.1 The Additive White Gaussian Noise Channel Model

In any communication system, the received signal is embedded in random noise, and is thus a random signal itself. The goal of the receiver, in case of digital communications is to detect the transmitted signal. Let us look at the transmitted signal. The transmitted bandpass signal g(t),

#### MIMO Multi-user Wireless Communications and Precoding

in any communication system is given by

$$g(t) = s_m(t) \exp\left(j2\pi f_c t\right), \qquad 0 \le t \le T_s, \tag{1.1}$$

where  $f_c$  is the carrier frequency, and  $s_m(t)$  is the bandlimited information carrying signal often referred to as the baseband signal. In the case of digital communications the baseband signal  $s_m(t)$ is chosen from a set of M signals. We represent this set of M signals by the symbol  $\Xi$ , i.e.,

$$\Xi = \{s_1(t), \cdots, s_m(t)\}.$$
(1.2)

The baseband signal represents a single piece of information, a symbol, which is to be conveyed by the transmitter to the receiver. Typically,  $s_m(t)$  is a time-limited signal and we refer to its time duration as the symbol period, represented as  $T_s$ . When the symbols transmitted correspond to bits 1 and 0, we also refer to the symbol period as the as the bit period  $T_b$ . The generation of a bandpass signal g(t) from a bandlimited signal  $s_m(t)$  as shown in (1.1) is called as up-conversion. Multiplication with a complex exponential in the time-domain corresponds to a frequency shift in the frequency domain. Since the signal  $s_m(t)$  is centered around a frequency of 0 Hz., the signal x(t)will be centered around  $f_c$ , and hence the name up-conversion. The up-converted signal is amplified and transmitted from the transmit antenna. In the simplest case, the channel adds random noise to the transmitted signal. We represent the noise signal as n'(t). The bandpass received signal r'(t)in the simplest model for a communication system can be written as

$$r'(t) = g(t - \tau) + n'(t)$$
  
=  $s_m(t - \tau) \exp(j2\pi f_c(t - \tau)) + n'(t),$  (1.3)

where  $\tau$  is the time taken by the signal to reach the receiver from the transmitter. We call this the path delay from the transmitter to the receiver. Corresponding to the up-conversion process at the transmitter, there is a down-conversion process at the receiver, the result of which will be the baseband received signal  $r_b(t)$ . We assume that the path delay  $\tau$  is perfectly known to the receiver. Then, the signal resulting from the down-conversion process is given as

$$r_b(t) = \left(s_m(t-\tau)\exp\left(j2\pi f_c(t-\tau)\right) + n'(t)\right)\exp\left(-j2\pi f_c(t-\tau)\right) = s_m(t-\tau) + n_b(t),$$
(1.4)

where  $n_b(t)$  is the complex noise signal at the output of the down-converter. We use the subscript b to denote the baseband nature of this signal. Since the receiver knows the path delay, it can shift the received signal to generate the effective baseband received signal  $r(t) = r_b(t + \tau)$ . This signal can be written as

$$r(t) = s_m(t) + n(t), (1.5)$$

where  $n(t) = n_b(t + \tau)$ . When the noise n(t) is white and Gaussian, (1.5) corresponds to the continuous time AWGN channel model [22]. Henceforth, unless specified otherwise, throughout this thesis we will refer to the baseband transmitted signal  $s_m(t)$  as the transmitted signal, and the baseband received signal r(t) as the received signal.

Let us now look at the noise process. There are various sources of noise, the primary one being thermal noise in the electronic components used to build the receiver. In most systems, noise components are present uniformly at all frequencies in the frequency band of interest. Consequently, we assume that the noise process is white, i.e., it has a flat power spectral density like the spectrum of white light. For a wide sense stationary stochastic process, the power spectral density is the Fourier transform of the autocorrelation function. This result is well-known as the Weiner-Khintchine theorem [29]. Consequently, the autocorrelation function will be an impulse function. This means that the samples of the noise process taken at distinct time intervals are uncorrelated. According to the central limit theorem, the random variable formed by adding an infinite number of independently and identically distributed random variables each having a finite mean and variance, has a Gaussian distribution [27]. Adding m independent Gaussian random variables, say  $z_1, \dots, z_m$  having corresponding means  $\mu_1, \dots, \mu_m$  and variances  $\sigma_1^2, \dots, \sigma_m^2$ , results into a Gaussian random variable with resulting mean

$$\sum_{i=1}^{m} \mu_i$$

and variance

$$\sum_{i=1}^m \sigma_i^2.$$

In fact, even if the variables are not identically distributed, the result is still Gaussian [28]. Since the noise is resultant from innumerable molecular sources, we can invoke the central limit theorem and assume the noise process to be Gaussian. The two-sided noise power spectral density constant is given by  $\frac{N_0}{2}$ . Noise samples at different time instants will be uncorrelated, and hence independent (since Gaussian), with mean 0 and variance  $\frac{N_0}{2}$ . As per (1.5), the received signal is embedded in this noise process. Observe that the received signal itself will be a Gaussian process, conditioned on  $s_m(t)$ , which as indicated in (1.2) belongs to the set  $\Xi$ . The goal of the receiver is to detect which signal from the set  $\Xi$  was transmitted. We now see how the receiver detects the transmitted signal.

#### 1.1.1 Match Filter Demodulator and the Detection Process

A sufficient statistic is a statistic such that it contains all the information contained in the random variables it is generated from [30]. As we see in (1.5), the receiver has to estimate  $s_m(t)$ , given the signal r(t). Generation of a statistic can be viewed as a form of data reduction. Hence, if instead of looking at the entire received signal r(t) to detect  $s_m(t)$ , we generate some discrete-time statistic using r(t), the detection process will be considerably simplified. However, this statistic should be sufficient statistic for  $s_m(t)$ .

The receiver consists of two parts following the down-convertor, the demodulator and the detector. The demodulator takes the continuous-time signal from the down-convertor output and uses it to generate sufficient statistics which are fed as input to the detector. To motivate the structure of the demodulator, which generates discrete-time statistics to estimate the transmitted signal, we now see how the transmitted signal can be represented as a discrete-time vector. Let the transmitted signal  $s_m(t)$  belong to the set  $\Xi$  as given in (1.2). Each possible transmitted signal  $s_m(t)$ , can be represented as a discrete-time vector with respect to some basis of the set  $\Xi$ . Let the signals  $p_1(t), \dots, p_U(t)$  form an orthonormal basis of the set  $\Xi$ . Here U is the dimensionality of the set  $\Xi$ . This basis can be found using the Gram-Schmidt procedure [35]. Using the basis we can write every transmitted signal as

$$s_m(t) = \sum_{u=1}^U s_m[u] p_u(t), \qquad m = 1, \cdots, M,$$
 (1.6)

where

$$s_m[u] = \langle s_m(t), p_u(t) \rangle$$
. (1.7)

The symbol  $\langle \cdot, \cdot \rangle$  represents the Euclidean inner product <sup>1</sup>. Thus, we say that the transmitted

$$\langle w(t), v(t) \rangle = \int_{-\infty}^{\infty} v^*(t)w(t)\mathrm{d}t, \qquad (1.8)$$

<sup>&</sup>lt;sup>1</sup>The Euclidean inner product is defined for continuous-time signals as [35]

signal can be represented as discrete-time vector  $s_m$ , i.e.,

$$s_m(t) \leftrightarrow \boldsymbol{s}_m,$$
 (1.10)

where  $\mathbf{s}_m = [s_m[1], \cdots, s_m[U]]^T$  is the vector corresponding to the decomposition of  $s_m(t)$  with respect to the basis of the set  $\Xi$ . We can also refer to  $\mathbf{s}_m$  as the symbol to be transmitted. Consequently, we can say that the receiver needs to detect the discrete-time signal  $\mathbf{s}_m$ . The demodulator computes the vector of sufficient statistics for estimating the transmitted symbol  $\mathbf{s}_m$ , (or equivalently the transmitted signal  $\mathbf{s}_m(t)$ ), using the basis signals  $p_1(t), \cdots, p_U(t)$ . We represent this vector of sufficient statistics by the symbol  $\mathbf{r}$ , with elements given as

$$r[i] = \langle r(t), p_i(t) \rangle$$
 (1.11)  
=  $s_m[i] + n[i],$ 

where  $n[i] = \langle n(t), p_i(t) \rangle$  will be a Gaussian noise sample. The vector  $\boldsymbol{r}$  will be given as

$$\boldsymbol{r} = \boldsymbol{s}_m + \boldsymbol{n},\tag{1.12}$$

where  $\mathbf{n} = [n[1], \dots, n[U]]^T$ , is a Gaussian vector with identically distributed entries. As long as the basis chosen is orthogonal, this vector will consist of independent entries. It has been shown in [22, 30] that when the vector  $\mathbf{r}$  is so generated, it corresponds to the vector of sufficient statistics for the signal  $s_m(t)$ , at least when the background noise signal is Gaussian. Looking at (1.12), we can say that the discrete-time transmitted signal is  $s_m$ . The corresponding discrete-time received signal is the signal  $\mathbf{r}$  formed by the demodulator. Thus, even though continuous-time signals are transmitted and received, we can model the equivalent channel as a discrete-time channel with discrete-time input  $s_m$  and discrete-time notation for the transmitted signal and the received signal throughout this thesis.

The operation indicated in (1.11) can be performed using a filter whose response is matched [29] to the basis signal  $p_i(t)$ . As mentioned above, typically, the signals  $s_m(t)$  are time-limited and and for discrete-time signals as

$$\langle \boldsymbol{w}, \boldsymbol{v} \rangle = \boldsymbol{v}^H \boldsymbol{w}. \tag{1.9}$$

The Euclidean norm  $|| \cdot ||$  of a signal, is defined as the square root of the Euclidean inner product of the signal with itself.

of duration  $T_s$  and so are the basis signals  $p_i(t)$ . In such a case the response of the match filter matched to basis  $p_i(t)$  is given as  $p_i(T_s - t)$ . The demodulator generally consists of a bank of match filters. Each filter from the bank will have its response matched to one of the signals in the basis set. The output of the match-filter is sampled at the symbol rate to obtain the discrete-time sufficient statistics r[i]. We often refer to r as the match filter output, since the demodulator often employs a match filter to generate the decision statistics. As an aside, note that an important property of a the match filter is that it maximizes the achievable signal to noise ratio (SNR), defined in Section 1.1.2, at the output of the demodulator [22].

The detector operates on the statistics generated by the demodulator and forms the decision statistics based on which the final decision about the transmitted symbol is made. We represent this decision statistic by the vector  $\boldsymbol{y}$ , when the transmitted symbol is a vector and by  $\boldsymbol{y}$ , when the transmitted symbol is a vector and by  $\boldsymbol{y}$ , when the transmitted symbol is a scalar. Two optimal detection techniques that are frequently employed are maximum *a posteriori* probability detection (MAP) and maximum likelihood detection (ML) [35]. The MAP detector is the optimal detector and results in the minimum error rate at the receiver. However, it is worth noting the following results which will be used throughout this thesis [16, 22]:

- When the input symbols have the same *a priori* probabilities, the MAP detector is the same as the ML detector.
- For an AWGN channel, ML detection corresponds to minimum Euclidean distance detection. This means that the estimate chosen should be that signal from the transmitted signal set, which is closest in the Euclidean sense to the received signal.

Throughout this thesis, we will assume that the input symbols are uniformly distributed, and hence the optimal detector will correspond to the ML detector.

As an example, consider the transmission, demodulation (match filtering) and detection of a binary phase shift keying (BPSK) symbol  $b = \{\pm 1\}$  using a time-limited pulse p(t) of duration  $T_b$ and given by

$$p(t) \triangleq \begin{cases} \frac{1}{\sqrt{T_b}} & , \quad 0 \le t \le T_b \\ 0 & t > T_b \end{cases}$$
(1.13)

as the basis signal. The baseband transmitted signal is given as

$$s_m(t) = b\sqrt{\mathcal{E}_b p(t)},\tag{1.14}$$

9

where  $\mathcal{E}_b$  is the transmitted energy per bit. Assuming perfect synchronization between the transmitter and the receiver, the received signal will be

$$r(t) = b\sqrt{\mathcal{E}}_b p(t) + n(t), \qquad (1.15)$$

where n(t) is the underlying Gaussian noise process. The match filter response will be [29]

$$p(t-T_b). \tag{1.16}$$

The output of the match filter is sampled at the instant  $T_b$  and fed to the detector as

$$r = b\sqrt{\mathcal{E}}_b + n,\tag{1.17}$$

where  $n \in \mathcal{N}(0, \frac{N_0}{2})$  represents the Gaussian noise variable, where  $N_0$  is the one sided noise power spectral density. Thus, we can say that the discrete-time transmitted signal is b and the corresponding discrete-time received signal is r. Since we have assumed that the inputs are equally likely, the error minimizing detector will correspond to the ML detector. The likelihood function of the received signal is given by the conditional pdf, p(r|b) or any monotonic function of this pdf [22]. Since the noise in (1.17) is zero-mean Gaussian, the likelihood of r will be Gaussian with mean  $\{\pm \sqrt{\mathcal{E}}_b\}$ . The likelihood of r given b = 1 will be

$$p\left(r\middle|b=1\right) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(r-\sqrt{\mathcal{E}}_b)^2}{N_0}\right),\tag{1.18}$$

and that given b = -1 will be

$$p\left(r\middle|b=-1\right) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(r+\sqrt{\mathcal{E}}_b)^2}{N_0}\right).$$
(1.19)

The ML detector chooses the estimate  $\hat{b}$  as 1 if p(r|b=1) > p(r|b=-1), else it chooses -1. This can be expressed as

$$p\left(r\middle|b=1\right) \underset{-1}{\stackrel{1}{\gtrless}} p\left(r\middle|b=-1\right).$$
(1.20)

The condition p(r|b=1) > p(r|b=-1) simply corresponds to the fact that the received signal is closer to  $\sqrt{\mathcal{E}}_b$  than  $-\sqrt{\mathcal{E}}_b$  in the Euclidean sense. Thus, the detection process will simply correspond to taking the sign of the statistic r to generate an estimate of the transmitted symbol b. Since the decision is based on the statistic r, the decision statistic in this case is

$$y = r. (1.21)$$

and the final decision made will be

$$\hat{b} = \operatorname{sign} \{y\} \tag{1.22}$$

and when the noise is complex, decision made will be

$$\hat{b} = \operatorname{sign} \, \Re\{y\}. \tag{1.23}$$

#### 1.1.2 Bit-Error-Rate of BPSK in an AWGN Channel

An important measure of the energy efficiency of a digital communication system is provided by the BER for a given SNR at the receiver. Let us see how the SNR is defined for a digital communication system. In analog communications, power signals, having finite power and infinite energy are used [34]. The figure of merit used to characterize the performance of different analog communication systems is the signal to noise power ratio  $SNR_p$ , defined as

$$SNR_p = \frac{P_s}{P_n},\tag{1.24}$$

where  $P_s$  is the received signal power and  $P_n$  is the noise power at the output of the receive filter. More details of performance measures for analog techniques can be obtained from excellent texts on the topic like [33, 29]. In case of digital communications, note that the received power per bit can given by

$$P_s = \frac{\mathcal{E}_b}{T_b},\tag{1.25}$$

where  $T_b$  is the bit duration and  $\mathcal{E}_b$  is the transmitted energy per bit. If the transmission bandwidth is *B* Hz, the receiver will employ a filter of this bandwidth and the noise power at the output of the receive filter will be,

$$P_n = N_0 B, \tag{1.26}$$

 $N_0$  being the one-sided noise power spectral density. Therefore, the effective signal to noise power ratio is

$$SNR_p = \frac{\mathcal{E}_b R}{N_0 B},\tag{1.27}$$

where  $R = \frac{1}{T_b}$  is the bit rate at which transmission takes place. The factor  $\frac{B}{R}$  is a parameter of bandwidth efficiency and depends on the modulation scheme. Therefore, we can define the normalized SNR as

$$SNR = \frac{\mathcal{E}_b}{N_0} = SNR_p \frac{B}{R}.$$
(1.28)

Another alternative could be to use the ratio of energy per transmitted symbol  $\mathcal{E}_s$ , to  $N_0$ , i.e.,  $\frac{\mathcal{E}_s}{N_0}$ . However, a symbol is not an exact unit of information. A single symbol may correspond to a single bit, like in BPSK or to more than one bit like in higher order modulation schemes like M-ary Phase Shift Keying (MPSK) [22]. Thus, plotting the BER or even the symbol-error-rate vs this parameter would not make sense, since it does not indicate how much energy is actually consumed per information unit (in this case, 1 bit).

We now compute the BER when using BPSK over an AWGN channel. From equations (1.17), (1.21) and (1.23), assuming equally probably inputs, the estimate of b is given as

$$\hat{b} = \operatorname{sign} \, \Re\{b\sqrt{\mathcal{E}}_b + n\}. \tag{1.29}$$

The noise term n will have a distribution given by  $\mathcal{N}(0, \frac{N_0}{2})$ . Assume that b = 1 is transmitted. The pairwise probability of error  $p_e(1 \to -1)$ , is the probability that the receiver estimate is  $\hat{b} = -1$ , given that b = 1 is transmitted. This probability is

$$p_e (1 \to -1) = P \left( \Re\{\sqrt{\mathcal{E}}_b + n < 0\} \right)$$
$$= P \left( \Re\{n\} < -\sqrt{\mathcal{E}}_b \right)$$
$$= Q \left( \sqrt{\frac{2\mathcal{E}_b}{N_0}} \right).$$
(1.30)

Here the Q-function, defined in Appendix A.3, represents the standard tail-function, which gives the area under the tail of a normalized Gaussian random variable <sup>2</sup>. It is easy to see that the pairwise probability of error that the receiver detects  $\hat{b} = 1$ , given that  $\hat{b} = -1$  is transmitted, is the same as that given in (1.30). The inputs are assumed equally probably with probability  $\frac{1}{2}$ . Consequently, the average probability of error  $P_e$ , averaged upon the inputs, for a particular SNR

 $<sup>^2\</sup>mathrm{A}$  normalized Gaussian random variable has zero mean and unit variance

will be

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) = \frac{1}{2}Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right) + \frac{1}{2}Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right)$$
$$= Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\right).$$
(1.31)

#### 1.2 The Wireless Channel

Unlike wireline communications in wireless communications waves emitted by the transmitter propagate in free space. The transmitted signals follow multiple paths from the transmitter to the receiver. There may exist a direct line of sight path between the transmitter and the receiver. However, reception will primarily take place from three phenomena: reflection, diffraction and scattering. Ideal reflection surfaces, which are more of an exception rather than the rule, do not absorb any energy and reflect all the energy incident upon them. Non-ideal reflectors however, absorb energy and result in higher attenuation than simple free space propagation. Diffraction refers to the phenomenon of propagation around an obstruction. As per the Huygen's principle [11], a new wavefront in the direction of propagation results from the combination of secondary wavelets, produced and radiated in all directions by all points on the previous wavefront. Thus, we have multiple waves arriving from around shadowing obstructions in the line of sight path between the transmitter, and/or reflectors and the receiver. Scattering takes place when the reflecting surface is rough (roughness is defined by the Rayleigh criterion [12]). Due to roughness of the reflecting surface, a continuum of waves is generated. This spreading or diffusion of the incident wave by the reflector is called scattering. Sometimes for smoother surfaces, along with diffusion, a reflected component, referred to as the specular component, (the word specular means "reflecting like a mirror" [13]), is also generated. The receiver receives signals through these three mechanisms and in some cases, directly from the transmitter when a line of sight path exists between the receiver and transmitter.

The signals received from different paths have different path lengths, and hence different path delays in general. As a result they are out of phase. Consider the addition of a reflected and an incident wave along the path between the transmitter and the reflecting surface. At any point along the path, the signals add constructively and destructively, depending on their phases at the point,

thereby forming a pattern of crests and troughs. The crests are separated from adjacent troughs by a distance of  $\lambda/4$ , where  $\lambda = c/f$ , is the wavelength of the carrier frequency f. As the receiver moves along this path, it passes through these crests and troughs, and experiences fluctuations in the received energy as it moves. In a real scenario, the received signal is a combination of almost an infinite number of waves arriving at the receiver. Wireless cellular communication standards cdma2000, W-CDMA, TD-CDMA, etc. [48, 52, 53, 54] and wireless local area network (WLAN) standards like IEEE 802.11a/b/g [55], and personal area network (PAN) standards like Bluetooth [56] are specified to operate in the frequency bands from 1GHz to 6 GHz. The corresponding wavelengths range from 30 cm. to 5 cm., all fractions of a meter. Consequently, obstructions, close to the receive antenna on the meter scale, and acting as scatterers/reflectors, will significantly affect the electromagnetic field impinging the receiver antenna. Due to mobility, either of the receiver or of nearby objects, path lengths and hence the phases of each of these signals change rapidly. The vector addition of such signals will result in variations in the received energy level. Consequently, the received energy changes with distance and surroundings, and hence with time. This fluctuation in the received energy with time is known as fading. Based on the time scale under consideration, fading is classified into two categories [14, 15]:

1. Large-scale fading: As a signal travels through free space, the intensity of the corresponding electric field varies inversely with the distance traversed. The power flux density, measured in watts per sq. m., at a particular point in space is proportional to the square of the electric field intensity at that point. The signal incident at the receiver, arrives from different paths as mentioned above. Thus, the power flux density at the receiver, and hence the received power, decay with a factor of 2 or more with distance from the transmitter. In fact, it has been shown in [20] that assuming uniformly random obstacles and exponential distribution of inter-obstacle distances, the received power decays exponentially with distance. It can be shown by simple computations that for standard systems, when travelling at speeds of around 25 miles/sec., a relatively significant change in the average received power, (where the averaging is over a large time interval), takes at least a few hundred milliseconds. However, the detection and estimation of signals takes place on the microsecond scale. The important thing to note is that the time-scale over which the received power changes significantly, is

large as compared to the time-scale over which detection of transmitted signals takes place. Hence, the phenomenon of change in received power with change in transmitter-receiver (T-R) separation is frequently referred to as large-scale fading. In Section 1.2.1, we will see different path loss models used to model the mean signal strength at the receiver. These models closely approximate the attenuation that the transmitted signal undergoes before reaching the receiver.

2. Small-scale fading: Changes in path length corresponding to a few centimeters will significantly change the phase of the signal received over that path. Thus, small-scale changes in the environment surrounding the receiver, which cause the path length to change by a few centimeters, cause a significant phase shift in the signals arriving at the receiver. Due to vector addition of signals having random phases, the receiver experiences randomly varying energy levels over very small distances corresponding to time spans on the order of a few symbol durations. The received energy level is high when there are a large number of in-phase multipath elements and low otherwise. This phenomenon, occurring over a small time scale, is appropriately referred to as small-scale fading, and sometimes as multipath fading [21]. We will define critical characteristics of small-scale fading in Section 1.2.3. In this thesis, we are primarily concerned with exploiting the distortion introduced by this phenomenon.

To characterize the received signal, we need to consider the electromagnetic field impinging the receiver antenna. In Section 1.1.1, we saw that the transmitted and the received signals can be represented as discrete-time signals because of the discretization process (demodulation) at the receiver. The discrete-time output at the receiver can be written as a function of a discrete-time channel filter and a discrete-time input. It has been shown that using accurate approximations to the electromagnetic field incident on the receiver, the discrete-time, baseband input-output channel can be modelled as a linear, but time-varying filter as far as detection of the transmitted signals is concerned [12, 20, 22]. Due to fading, the filter response of the discrete-time filter in our model will not only be time-varying, but will also be random. In Section 1.2.2, we will see statistical models based on the random characteristics of the channel filter response. The performance measure of energy efficiency of a communication system is the BER for a particular SNR at the receiver. In Section 1.2.5, we will see the effect of multipath fading on the BER performance of the system.

#### 1.2.1 Large-Scale Fading: Path Loss Models

As the separation between the transmitter and the receiver increases, the signal gets more and more attenuated. Path loss refers to the average loss in received power, based on the T-R separation. We denote it by the symbol  $\overline{PL}$  and it is measured in dB. It is useful for computing the coverage area of a transmitter for a broadcast channel. Note that in this thesis, by a broadcast we refer to a channel where a central transmitter is transmitting independent data streams to independent users. A typical example of this is the cellular downlink. Path loss models for macrocellular systems, where antennas are placed at large heights [21], have been studied in [23], while those for micro-cellular systems, having antennas at low heights, are studied in [24]. Since we are not going to address the problem of large-scale fading we will only highlight significant features of some important path-loss models.

The complexity of analysis based upon solving the electromagnetic equations to determine the exact electromagnetic field incident upon the receiving antenna is extremely high. Hence, in designing practical models we use empirical measurements along with some analytical inferences. The energy received from the line of sight communication decays inversely with the square of the T-R separation. Reflection from non-ideal surfaces results in higher losses, depending on the reflection material characteristics and on the total distance travelled by the signal before reaching the receiver. Losses due to scattering and diffraction are higher than reflection losses [12]. At large distances the path difference between the direct path and the reflection path from the ground surface is relatively small. However, the ground surface introduces a phase shift of  $\pi$ . The resulting destructive addition of waves at the receiver causes the energy reduction to be proportional to the fourth power of the distance traversed by the signal. Considering all these factors, the average path-loss for a transmitter receiver separation of d is given by [12],

$$\overline{PL}(d) = \overline{PL}(d_0) + 10n \log\left(\frac{d}{d_0}\right) \text{ dB.}$$
(1.32)

Here  $d_0$  is the reference distance for which the loss is computed via actual measurement [25] and n is the path loss exponent, which varies from 2 to 10. The most common value used is 3 and 10 is used for highly dense environments. This is the log-distance path loss model.

The log-distance path loss model in (1.32) does not account for the randomness of the environment for fixed T-R separations. This randomness is accounted for by the log-normal shadowing model. Measurements indicate that it suffices to add a Gaussian random variable, say  $X_{\sigma} \sim (0, \sigma^2)$ , given in dB, to the path loss given by the log-distance model [12, 26]. Thus, the path loss in dB is normally distributed with mean  $\overline{PL}$  and variance  $\sigma^2$ . In literature it is often said that the path loss is log-normally distributed. The corresponding received power in dB is given by

$$P_r(d) = P_t - \overline{PL}(d) - X_\sigma \, \mathrm{dB},\tag{1.33}$$

where  $P_t$  is the transmitted power. This is known as the log-normal model.

Throughout this thesis, we will not be concerned with the attenuation the signal undergoes from the transmitter to the receiver. In our multi-user system model, we will assume that log-normal shadowing is alleviated by power allocation and control schemes.

#### 1.2.2 Bandlimited Channel Impulse Response

Consider the transmission of a single symbol as given in Section 1.1. The baseband transmitted signal, the up-converted and transmitted signal will be the same as given in (1.1) and (1.2). However, the bandpass received signal will be unlike (1.3), since the transmitted signal now passes through a tetherless medium taking multiple paths from the transmitter to the receiver. Since each path has a different path length, the signal arrives with a different path delay and different attenuation along every path. We call the coefficient representing the attenuation over a particular path as the path gain for that path. Due to mobility, the path lengths will be time-varying and so will be the corresponding path delays and the path gains. The net received signal is a linear combination of the signals received over different paths and the noise signal n'(t). The channel can be classified as a discrete multipath channel or as a continuous multipath channel [22]. In case of a discrete multipath channel, the different paths are resolvable and we can write the bandpass received signal as a sum of discrete components. Thus,

$$r'(t) = \sum_{i} a_i(t) s_m \left( t - \tau_i(t) \right) \exp\left( j 2\pi f_c(t - \tau_i(t)) \right) + n'(t), \tag{1.34}$$

where  $a_i(t)$  represents the time-varying path gain and  $\tau_i(t)$  represents the time-varying path delay of the  $i^{th}$  path. In case of a continuous multipath channel, which occurs when scattering is dominant, a continuum of waves is received. The bandpass received signal has the form

$$r'(t) = \int_{-\infty}^{\infty} a(\tau, t) s_m(t - \tau) \exp\left(j2\pi f_c(t - \tau)\right) dt + n'(t),$$
(1.35)

#### Chapter 1 MIMO Multi-user Wireless Communications and Precoding

where  $a(\tau, t)$  represents the time-varying path gain and  $\tau$  represents the path delay. Since there are multiple paths, unlike the down-conversion for a single path channel given in (1.4) the down-convertor, does not include information of the path delays. In case of a discrete multipath channel the baseband received signal after down-conversion will be

$$r(t) = \left(\sum_{i} a_{i}(t)s_{m}(t - \tau_{i}(t)) \exp\left(j2\pi f_{c}(t - \tau_{i}(t))\right) + n'(t)\right) \exp(-j2\pi f_{c}t) + n_{b}(t)$$

$$= \sum_{i} a_{i}(t)s_{m}(t - \tau_{i}(t)) \exp(-j2\pi f_{c}\tau_{i}(t)) + n(t),$$
(1.36)

where n(t) will be the noise at the output of the down-convertor. The effective continuous-time, bandpass (time-varying) channel impulse response for a discrete multipath channel can be given by

$$h(\tau, t) = \sum_{i} a_{i}(t) \exp(-j2\pi f_{c}\tau_{i}(t))\delta(\tau - \tau_{i}(t)), \qquad (1.37)$$

where  $\delta(\tau)$  is the well-known Dirac delta function such that

$$\delta(\tau) = 0, \qquad \tau \neq 0$$

$$\int_{-\infty}^{\infty} \delta(t) = 1.$$
(1.38)

The effective continuous-time, bandpass (time-varying) channel impulse response for a continuous multipath channel can be obtained similarly as

$$h(\tau, t) = a(\tau, t) \exp(-j2\pi f_c \tau).$$

$$(1.39)$$

The continuous-time, bandlimited received signal can be given as

$$r(t) = \int_{-\infty}^{\infty} a(\tau, t) s_m(t - \tau) \exp(-j2\pi f_c \tau) dt + n(t).$$
(1.40)

Thus, we can say that the channel can be modelled as a time-varying linear filter. In the following section we will motivate the different parameters of a fading channel based on the time varying nature of this filter and finally obtain the discrete-time statistical models for a wireless communication system in Section 1.2.4.

#### 1.2.3 Small Scale Fading

We have seen that small scale fading results from small scale changes in the surroundings of the receiver. From the time varying filter model for the channel we infer that there are at least two critical characteristics of the channel which need to be parameterized. The time-varying nature of the channel, represented as a variation of the channel impulse response with t, necessitates a measure of the temporal coherence of the channel. The filter nature of the channel on the other hand, necessitates a measure of the frequency coherence of the time-varying frequency response of the channel filter  $H(f,t) = \mathcal{F}\{h(\tau,t)\}$ . Here f represents the Fourier domain corresponding to the  $\tau$  domain. For the Fourier domain corresponding to the t domain we will use the symbol  $\omega$ .

From the filter model for the discrete multipath channel given in (1.37), we observe that each tap of the filter has the form

$$a_i(t) \exp\left(-j2\pi f_c \tau_i(t)\right).$$

The changes in the taps with time t, are reflected significantly in the phase change, the change being directly proportional to the carrier frequency  $f_c$  [20]. A change in the path delay  $\tau_i(t)$  by  $\frac{1}{4f_c}$ will result in a significant phase change of  $\frac{\pi}{2}$  in the signal received from the path i. The change in path delay by  $\frac{1}{4f_c}$  corresponds to a change in path length by  $\lambda_c/4$ , where  $\lambda_c = \frac{c}{f_c}$  is the carrier wavelength. The path length is changing from relative motion between either the obstructions or the transmitter, and the receiver. Different paths will experience different relative velocities. Let the maximum relative velocity experienced by at least one of the paths be v. The change in path delay corresponding to a path length change of  $\lambda_c/4$ , will be  $\frac{c}{4f_c}/v = \frac{c}{4f_cv} = \frac{1}{4B_d}$ , where

$$B_d = \frac{f_c v}{c}.\tag{1.41}$$

We refer to  $B_d$  as the Doppler spread of the channel. The parameter

$$T_c = \frac{1}{4B_d} \tag{1.42}$$

is called the coherence time of the channel, and defined as the time for which the channel remains almost constant. Here we have defined a phase change of  $\frac{\pi}{2}$  as a significant change to get the denominator in (1.42) as 4. but we could use different definitions for significant change and use the denominator accordingly. This discussion also holds for a continuous multipath channel. As per (1.39), the path delay change corresponds to the change in  $\tau$  as compared to the change in  $\tau_i(t)$  for the discrete multipath channel.

The summation in (1.37) is over distinguishable paths. We define the multipath delay spread,  $T_d$  as the time difference between the longest and the shortest of these paths, i.e.,

$$T_d = \max_i \tau_i(t) - \min_i \tau_i(t). \tag{1.43}$$

In case of the continuous multipath channel, the multipath delay spread corresponds to the value of  $\tau$  beyond which the amplitude  $a(\tau, t)$  is relatively negligible. The coherence bandwidth,  $B_c$ , is the frequency interval over which the channel is highly correlated. It is inversely proportional to the delay spread  $T_d$ . The constant of proportionality used varies in a manner similar to that in the relation between coherence time and Doppler frequency. For a discrete multipath channel having a corresponding single tap filter structure, the corresponding frequency response H(f,t) will exhibit a non-resolvable frequency flat behavior. The multi-tap nature of the channel injects frequency selectivity in the channel behavior. Note that the coherence bandwidth is analogous to coherence time, while the Doppler spread is analogous to the multipath delay spread.

Further clarity and justification on the relations given above is obtained by looking at the channel correlation functions and the Doppler power spectrum of the channel impulse response which are defined below. Since the multipath channel has a large number of paths contributing to the total response, and each path has a random phase and attenuation, we can invoke the central limit theorem and say that the band-limited channel response  $h(\tau, t)$  will be a complex valued Gaussian random process [22]. A complex Gaussian random process is composed of two real, independent, Gaussian random processes as in-phase and quadrature components. Assuming the channel to be a wide-sense stationary process, we can define the channel autocorrelation function  $\phi_h(\tau_1, \tau_2, \Delta t)$ , as

$$\phi_h(\tau_1, \tau_2, \Delta t) \triangleq \frac{1}{2\sigma_h^2} E\left\{h^*(\tau_1, t)h(\tau_2, t+\Delta t)\right\}, \qquad (1.44)$$

where  $\sigma_h^2 = E\{|h(\tau,t)|^2\}/2$ . Since the channels associated with different paths are uncorrelated, this simplifies to

$$\phi_h(\tau_1, \tau_2, \Delta t) = \phi_h(\tau_1, \Delta t)\delta(\tau_1 - \tau_2).$$
(1.45)

For fixed time instant t or  $\Delta t = 0$ , the above autocorrelation function gives the average received power as a function of  $\tau$ . This is referred to as the multipath intensity profile of the channel and frequently denoted by  $\phi_h(\tau)$ . Note that from the definition of multipath delay spread  $T_d$  as given above, significant portion of the intensity profile is constrained within the interval  $0 \leq \tau \leq T_d$ .

The spaced-frequency autocorrelation function is the autocorrelation function de-

fined as

$$\phi_H(f_1, f_2, \Delta t) \triangleq \frac{1}{2} E(H^*(f_1, t)H(f_2, t+\Delta t)).$$
 (1.46)

This can be shown to be the Fourier transform of the correlation function defined in (1.44) [22], i.e.,

$$\phi_H(\Delta f, \Delta t) = \int_{-\infty}^{\infty} \phi_h(\tau_1, \Delta t) \exp\left(-j2\pi \Delta f\tau_1\right) d\tau_1, \qquad (1.47)$$

where  $\Delta f = f_2 - f_1$ . Thus, for a fixed time instant t, we see that the spaced-time spaced-frequency autocorrelation function is the Fourier transform of the multipath intensity profile. This further justifies the relation between the coherence bandwidth and the multipath delay spread indicated above.

Based on the relation between the coherence bandwidth and the signalling bandwidth  $B_s = \frac{1}{T_s}$ , where  $T_s$  is the symbol period, we classify the channel as flat or frequency selective fading. If  $B_s \leq B_c$  or equivalently,  $T_d \leq T_s$ , all the components of the transmitted spectrum experience the same channel response. We say that the channel is flat fading. For purposes of detection and estimation of transmitted symbols, the equivalent channel can be considered as a single tap filter. On the other hand if this constraint is not met, different parts of the transmitted frequency spectrum experience different channel responses. In this case we say that the effective channel is frequency selective. The equivalent channel response is now modelled as having multiple taps. The multi-tap filter nature results into intersymbol interference (ISI) as we now see. When  $T_d > T_s$ , the symbol spreads in time beyond the symbol duration and hence parts of the signal are received for more than a single symbol interval. When the transmitter transmitter symbols continuously, adjacent symbols interfere with each other and we refer to this as ISI. If the transmitter pads zeros between successive symbols ISI can be eliminated, but observe that the effective data transfer rate is reduced. When the transmitter pads zeros between symbols to avoid ISI, we refer to the resulting channel as frequency selective without ISI. When no zero-padding is done to avoid ISI, we refer to the channel as frequency selective with ISI.

Next, we fix the frequency f, i.e., take  $\Delta f = 0$  in (1.47), to get  $\phi_H(\Delta t) = \phi_H(0, \Delta t)$ . This will characterize the temporal variations of the channel response. The Fourier transform given by

$$S_H(\omega) = \int_{-\infty}^{\infty} \phi_H(0, \Delta t) \exp\left(-j2\pi\omega \Delta t\right) d\Delta t, \qquad (1.48)$$

gives the Doppler power spectrum of the channel. The range of values over which it will be essentially non-zero will be the Doppler spread which we have seen in (1.41). The Fourier transform relation between the time-autocorrelation  $\phi_H(\Delta t)$ , and the Doppler spectrum justifies the relationship between the Doppler spread and the coherence time of the channel. Note that if the channel is static, then the Fourier transform  $S_H(\omega)$  will be an impulse  $\delta(\omega)$ .

Based on the coherence time, we classify the channel as slow and fast fading. If the coherence time is less than a symbol period, i.e.,  $T_c < T_s$ , or equivalently,  $B_d > B_s$ , we say the channel is fast fading else we say it is slow fading. Another term frequently used in literature is quasi-static fading, which refers to a slowly fading channel such that the coherence time is more than a few symbol durations. In this case, the channel is essentially constant for a frame of transmission. Throughout this thesis, we will assume the channel to be quasi-static.

In this section we have characterized the important parameters of a wireless channel corresponding to small-scale fading. In the following section we build a discrete-time model for a slowly fading wireless channel.

#### 1.2.4 Discrete-Time Statistical Model for a Slowly Fading Wireless Channel

Similar to the discrete-time AWGN channel model, we can form a discrete-time model for a wireless channel. We saw that in case of wireless communications, the channel from the transmitter to the receiver corresponds to a complex Gaussian random process. Thus, in a wireless system, besides being embedded in noise, the transmitted signal is filtered via a random process before arriving at the receiver as indicated in (1.36) and (1.40). Consider the signal for a continuous multipath channel as given in (1.40). The signal r(t) can be expressed in terms of the Fourier transform of the transmitted signal  $S_m(f)$ , and the time-varying frequency response of the channel filter H(f,t) as [22]

$$r(t) = \int_{-\infty}^{\infty} H(f,t) S_m(f) \exp(j2\pi ft) \,\mathrm{df} + n(t).$$
(1.49)

Now assume that the channel is slow and flat fading, such that it is essentially constant during the transmission of the signal  $s_m(t)$ . Then, the time-varying filter response can be written as a constant H(f,t) = H(0,t) = h and hence,

$$r(t) = hs_m(t) + n(t)$$
(1.50)

where, h will now represent a complex Gaussian random variable since we have assumed the channel to be complex Gaussian. As per (1.11), the output of the match filter k at the demodulator will be

$$r[k] = \langle r(t), p_k(t) \rangle$$
 (1.51)  
=  $hs_m[k] + n[k].$ 

Correspondingly we can write the discrete-time received signal at the output of the demodulator as

$$\boldsymbol{r} = h\boldsymbol{s}_m + \boldsymbol{n}.\tag{1.52}$$

This is the discrete-time model for a wireless slowly flat fading channel. Proceeding on similar lines as (1.49), a slow and frequency selective fading channel, which as we saw results from  $T_d > T_s$ , can be modelled as a discrete-time time-invariant finite impulse response (FIR) filter [22] for a time-duration during which we assume the channel to be constant. This filter response can be represented as

$$h(\tau) = \sum_{l=0}^{L-1} h[l]\delta\left(\tau - \frac{l}{B}\right)$$
(1.53)

where B is the bandwidth of the transmitted bandlimited signal  $s_m(t)$ . The time-variation of the channel is modelled as time-variations of the filter coefficients h[l]. We assume that the channel is constant at least for a duration of  $T_d$  seconds. The corresponding channel output can be given as

$$r(t) = \sum_{l=0}^{L-1} h[l]s_m\left(t - \frac{l}{B}\right) + n(t), \qquad 0 \le t \le T_d.$$
(1.54)

Note that the tap spacing is 1/B, and we often say that the channel can be modelled as a having L paths with a path delay of 1/B between consecutive paths. For both flat, as well as frequency selective channels, depending on the filter coefficients, h in case of a flat fading channel and  $h[l], \forall l$ , in case of a frequency selective channel, the channel can be classified as follows:

1. Rayleigh fading channel: The mean of each of the complex dimensions of h, (each h[l] for a frequency selective channel), is 0. Thus, as given in (A-24), the envelope |h|, (each |h[l]| for a frequency selective channel), will have a generalized Rayleigh distribution with n = 2. In this thesis we will stick to the Rayleigh model because of its simplicity. This occurs when no specular components are present in the received signal.

2. Rician fading channel: In the presence of a specular component, the mean of the each of the complex dimensions of h, (each h[l] for a frequency selective channel), is non-zero. In this case the envelope will have a Rician distribution. This is a two parameter distribution and useful for modelling outdoor channels, where a direct path exists between the transmitter and receiver.

Besides these models the two-parameter Nakagami-m model [31] is frequently used. It has been shown that this model in fact provides the best fit to empirically available data for urban wireless channels [22, 32]. Throughout this thesis, we will assume the channel to be Rayleigh fading, however most of the results can be directly extended to other channel models. The following section gives the BER performance of a Rayleigh flat fading channel.

#### 1.2.5 Performance of BPSK in the presence of Fading

As we now see in a wireless system, the BER reduces linearly with an increase in the SNR. Assuming BPSK transmission, from (1.52) the received signal in case of a slow, flat fading channel can be written as

$$r = hb + n, \tag{1.55}$$

where  $h \sim \mathcal{N}_c(0, \sigma_h^2)$ , is the channel coefficient, and  $n \sim \mathcal{N}_c(0, \frac{N_0}{2})$  is the noise coefficient. We assume coherent detection, i.e., detection assuming that the channel is completely known at the receiver. The detector uses the knowledge of the phase distortion introduced by h to generate the decision statistic given by  $h^*r$ , and the bit estimate given as

$$b = \operatorname{sign} \Re \{h^* r\}$$
  
= sign  $\Re \{|h|^2 \sqrt{\mathcal{E}_b} + h^* n\}.$  (1.56)

We observe that real part of the decision statistic  $h^*r$ , has a conditional Gaussian distribution, conditioned upon h, with mean,  $|h|^2 \sqrt{\mathcal{E}_b}$  and variance  $|h|^2 \frac{N_0}{2}$ . The pairwise probability of error conditioned upon h is given by

$$p_{e}\left(1 \rightarrow -1 \left|h\right) = P\left(\Re\left\{|h|^{2}\sqrt{\mathcal{E}_{b}} + h^{*}n\right\} < 0\right)$$
$$= Q\left(\frac{|h|^{2}\sqrt{\mathcal{E}_{b}}}{\sqrt{\frac{|h|^{2}N_{0}}{2}}}\right)$$
$$= Q\left(\sqrt{\frac{2|h|^{2}\mathcal{E}_{b}}{N_{0}}}\right)$$
$$= Q\left(\sqrt{2x}\right),$$
$$(1.57)$$

where  $x = \frac{|h|^2 \mathcal{E}_b}{N_0}$  is a Chi-square random variable, (since each dimension of h is Gaussian), with probability density function (p.d.f.) given by

$$p_X(x) = \frac{1}{m} \exp\left(-\frac{x}{m}\right), \qquad x \ge 0, \tag{1.58}$$

where  $m = 2\sigma_h^2 \frac{\mathcal{E}_b}{N_0}$ . The average pairwise error probability is obtained by averaging (1.57) over the p.d.f. of x. Assuming equally likely inputs, this average will also be the average BER represented as  $P_e$ . Thus for a given SNR, the average BER is

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) = \int_0^\infty \frac{1}{m} \exp\left(-\frac{x}{m}\right) Q\left(\sqrt{2x}\right) \mathrm{dx}.$$
(1.59)

We use the Laplace transform method as shown in Appendix A.3. The Laplace transform of the p.d.f. of X is

$$\phi_X(s) = \mathcal{L}\left(p_X(x)\right) = \frac{1}{1+ms}$$

Substituting this in (A-20), we get

$$P_{e}\left(\frac{\mathcal{E}_{b}}{N_{0}}\right) = \int_{0}^{\pi/2} \frac{1}{\pi} \frac{\sin^{2} t}{\sin^{2} t + m} dt$$
  
$$= \frac{1}{\pi} \left[ t - \sqrt{\frac{m}{1+m}} \tan^{-1} \left(\sqrt{\frac{1+m}{m}} \tan t\right) \right]_{0}^{\pi/2}$$
  
$$= \frac{1}{2} \left( 1 - \sqrt{\frac{2\sigma_{h}^{2} \frac{\mathcal{E}_{b}}{N_{0}}}{1 + 2\sigma_{h}^{2} \frac{\mathcal{E}_{b}}{N_{0}}}} \right).$$
 (1.60)

Note that  $\sigma_h^2$  is the variance of the real as well as the imaginary parts of the channel coefficient h. Figure 1.1 is a plot of the BER expressions given in equations (1.31) and (1.60). Observe that unlike the Gaussian channel, where the BER varies exponentially with an increase in the SNR (from

the bound on Q-function given in (A-18)), with fading it varies linearly. To observe the linearity for fading channels, we can simply use the bound computed in (A-28), using  $a = \sqrt{\frac{2\mathcal{E}_b}{N_0}}$  and n = 2, since the coefficient  $|h|^2$  in (1.57) has a generalized Rayleigh distribution of order 2. This bound on the BER is obtained as

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) \le \frac{1}{\left(1 + 2\sigma_h^2 \frac{\mathcal{E}_b}{N_0}\right)} \approx \frac{1}{\left(\frac{\mathcal{E}_b}{N_0}\right)} \tag{1.61}$$

where the approximation holds for high SNRs when  $\sigma_h^2 = 1/2$ . In the following section, we will see how the multipath nature of the channel can be exploited to improve the energy efficiency of a wireless system by increasing the slope of the BER curve.



Figure 1.1: Figure illustrates the BER vs SNR performance of BPSK in an AWGN channel and a Rayleigh fading channel

#### **1.3** Multi-Antenna Communications

The performance of a communication system can be measured using throughput, defined as the number of successfully received information bits per transmission. The achievable throughput depends on the BER. For lower values of BER, since errors are minimal, throughput is large. The BER, as we observed in the above section, depends on the SNR and hence on the transmission
energy. If infinite energy was available for transmission, then high throughput could be easily obtained. However, this not being possible, we have to impose an energy constraint on the transmission. With an energy constraint on the transmissions, we target BER minimization so as to maximize the effective throughput. This corresponds to minimization of the BER for a given SNR. We saw in the previous section that as we move from the Gaussian channel to the fading channel, the exponential decay of error rate with increase in the SNR becomes linear. Specifically, from (1.61), we can say that at high SNRs,

$$P_e(\text{SNR}) \propto \text{SNR}^{-1}.$$
 (1.62)

We have observed that multipath fading is an inherent characteristic of wireless channels. The randomness introduced via multipath fading can be exploited to improve the energy efficiency of wireless links. The use of multipath to improve energy efficiency falls under the broad category of diversity techniques. Diversity techniques focus on improving the slope of BER curve, specifically at high SNRs, so that the BER falls faster with a rise in SNR. The goal of diversity techniques is maximizing the diversity gain d, defined as the slope of the BER vs SNR plot in the high SNR region,

$$P_e(\text{SNR}) \propto \text{SNR}^{-d}.$$
 (1.63)

We will choose d as our figure of merit throughout this thesis. Higher values of d imply lower error rates for a given SNR indicating improved energy efficiency. We use this parameter as a measure of the energy efficiency of a system. Diversity techniques are reviewed in the following section.

### 1.3.1 Diversity Techniques

In diversity communications, multiple versions of the same signal are made available to the receiver. The receiver as we will see, combines the different versions in a manner so as to minimize the BER. Sometimes parallel channels can be created between the transmitter and the receiver so as to provide the receiver with multiple versions of the transmitted signal. Such parallel channels can alternately be used to maximize the rate of communications by sending distinct signals across each of the channels [37, 39]. Our focus will be on the diversity part. Note that creating parallel channels is not necessary for achieving diversity communications. As we saw in Section 1.2.4, a frequency

selective channel can be modelled as a multi-tap filter. If the responses resulting from each of the taps can be isolated at the receiver, then diversity combining can be achieved. Diversity can be obtained via space, time, frequency, polarization and other methods. Time and frequency diversity are achieved by signalling in different time and frequency slots respectively. Frequency diversity can also be attained by using the frequency selective nature of wireless channels. Polarization diversity is achieved by reception using multiple antennas each have a different polarization. A form of diversity, which has received wide-spread attention over the past decade is spatial diversity obtained by using multiple antennas at the transmitter (transmit diversity) and/or at the receiver (receive diversity). In this thesis we will focus on multiple antenna techniques, which are extremely popular. They are well known as multi-input multi-output (MIMO) communication techniques. The fundamental idea behind the use of diversity techniques is as follows: Let the probability that a signal received across a particular channel experiences a deep fade, i.e., a very low instantaneous SNR, be p. Assume that all the parallel channels created have the same statistics and hence the same probability of experiencing a deep fade p. The signal passes through each of these parallel channels and arrives at the receiver experiencing independent fading across each channel. Then, the probability that all L copies available at the receiver have experienced a deep fade is  $p^{L}$ . Note that for optimal use of diversity techniques, it is essential that the signals received from different parallel channels have experienced independent fades, and have widely varying instantaneous SNRs.

Once the different versions of the transmitted signal are available at the receiver, different diversity combining techniques can be used [12, 21]. The diversity combining takes place at the receiver, primarily at the detector. One way of using this diversity to reduce the BER is to choose the signal with the largest instantaneous SNR. This is called selection diversity (SD) combining [22]. The optimal combining technique, should result into a ML receiver. Let the transmitted signal set be a continuous time waveform  $x(t) = s_m(t)$ , where  $s_m(t) \in \Xi, m = 1 \dots M$  is a continuous-time signal having a discrete time representation  $s_m$ , with respect to the basis of the set  $\Xi$ . Let us assume that the receiver is provided with L independently faded versions of the transmitted signal. Let the fading experienced by the  $i^{th}$  copy of the signal be given by  $h_i$ . The net received signal at the receiver can then be written as

$$r(t) = \sum_{i=0}^{L-1} h_i x(t) + n_i(t), \qquad (1.64)$$

where  $n_i(t)$  is a zero-mean, white, complex Gaussian noise process corresponding to the  $i^{th}$  path. The noise processes  $n_i(t), i = 0, \dots, L-1$  are independent. Note that the above model corresponds to a slow frequency selective fading channel when the receiver can separate the distinct paths. After match filtering with respect to an orthogonal basis for the set of signals  $\Xi$ , the discrete-time received signal can be written as

$$\boldsymbol{r} = \sum_{i=0}^{L-1} h_i \boldsymbol{x} + \boldsymbol{n}_i = \sum_{i=0}^{L-1} \boldsymbol{r}_i, \qquad (1.65)$$

where the complex Gaussian white noise vectors  $\mathbf{n}_i$  are independent of each other, and each is assumed to have a covariance matrix  $\sigma^2 \mathbf{I}$ . The symbol  $\mathbf{x}$  corresponds to the discrete-time transmitted symbol  $\mathbf{s}_m \in \{\mathbf{s}_1, \dots, \mathbf{s}_M\}$ . It is clear that  $\mathbf{r}_i$  is the signal received over the  $i^{th}$  path and each  $\mathbf{r}_i$  is perfectly known to the receiver. The instantaneous SNR of the signal received over path i is a function of  $|h_i|^2$ . Maximum ratio combining (MRC) consists of taking a weighted average of the signals received from different paths, with weights proportional to the instantaneous SNR along that path. As we now prove, MRC results into a ML receiver. Since the noise is zero-mean, the likelihood function [36] of the received signal given that the signal  $\mathbf{s}_n$  was transmitted, is given by [21]

$$p(\mathbf{r}|\mathbf{s}_n) = \frac{1}{(2\pi\sigma^2)^{LU/2}} \exp\left(-\frac{\sum_{i=0}^{L-1} |\mathbf{r}_i - h_i \mathbf{s}_n|^2}{2\sigma^2}\right),$$
(1.66)

where U, as mentioned in Section 1.1, is the dimensionality of the set  $\Xi$ . ML detection corresponds to choosing  $s_n$ , which maximizes the likelihood given by (1.66). Thus, we choose  $s_n$  which minimizes the function

$$l(\mathbf{s}_{n}) = \sum_{i=0}^{L-1} |\mathbf{r}_{i} - h_{i}\mathbf{s}_{n}|^{2}$$
  
= 
$$\sum_{i=0}^{L-1} (|\mathbf{r}_{i}|^{2} - 2Re\{h_{i}^{*}\mathbf{s}_{n}^{H}\mathbf{r}_{i}\} + |h_{i}|^{2}\mathcal{E}_{n}),$$
 (1.67)

where  $\mathcal{E}_n$  is the transmitted signal energy given by

$$\mathcal{E}_n = \langle \boldsymbol{s}_n, \boldsymbol{s}_n \rangle \,. \tag{1.68}$$

Note that the function  $l(s_n)$  represents the Euclidean distance between the received signal and the transmitted signal, assuming that the channel is perfectly known to the receiver. Assuming an

equal energy constellation, i.e.,

$$\langle \boldsymbol{s}_i, \boldsymbol{s}_i \rangle = \langle \boldsymbol{s}_j, \boldsymbol{s}_j \rangle, \quad \forall i, j,$$
 (1.69)

we observe that the ML estimate of the transmitted symbol is given by

$$\hat{\boldsymbol{s}}_m = \arg\max_{\boldsymbol{s}_n} \sum_{i=0}^{L-1} Re\{h_i^* \boldsymbol{s}_n^H \boldsymbol{r}_i\}.$$
(1.70)

Note that this also corresponds to

$$\hat{\boldsymbol{s}}_m = \arg\max_{\boldsymbol{s}_n} \sum_{i=0}^{L-1} Re\{h_i^* \boldsymbol{r}_i^H \boldsymbol{s}_n\}.$$
(1.71)

Observe that multiplying  $\mathbf{r}_i$  with  $h_i^*$ , corresponds to the weighting it with the instantaneous SNR of the  $i^{th}$  channel (which is directly proportional to  $|h_i|^2$ ). This technique of weighting and combining independent replicas of the transmitted signal corresponds to MRC. Hence, we say that MRC results into ML performance. Equations (1.70) and (1.71) are referred to as pre-detection and post-detection combining respectively [21].

We now see the achievable BER with BPSK signalling at a given SNR when using this technique. Consider that we are using simple BPSK signalling, as seen in Section 1.1.1, where the transmitted symbol is  $b = \{\pm 1\}$ . Assume that b = 1 is transmitted. With BPSK signalling the decision statistic will be

$$\hat{b} = \operatorname{sign} \, \Re\left\{ \sum_{i=0}^{L-1} |h_i|^2 b \sqrt{\mathcal{E}_b} + h_i^* n_i \right\},$$
(1.72)

where  $\mathcal{E}_b$  is the transmitted energy per bit. The complex Gaussian noise terms is assumed to have a variance of  $\frac{N_0}{2}$  per dimension. Since the inputs are equally-likely, as we saw in (1.31), the pairwise error probability will be the actual BER. The corresponding probability of error conditioned upon the channel, is then given by

$$P\left(\Re\left\{\sum_{i=0}^{L-1}|h_{i}|^{2}\sqrt{\mathcal{E}_{b}}+h_{i}^{*}n_{i}\right\}<0\Big|\{h_{j}\}\right)=Q\left(\frac{\sum_{i=1}^{L}|h_{i}|^{2}\sqrt{\mathcal{E}_{b}}-0}{\sqrt{\sum_{i=0}^{L-1}|h_{i}|^{2}N_{0}/2}}\right)$$
$$=Q\left(\sqrt{\frac{2\sum_{i=0}^{L-1}|h_{i}|^{2}\mathcal{E}_{b}}{N_{0}}}\right).$$
(1.73)

The average BER for a given SNR is thus given by

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) = E_{\{h_j\}}\left\{Q\left(\sqrt{\frac{2\sum_{i=0}^{L-1}|h_i|^2\mathcal{E}_b}{N_0}}\right)\right\},\tag{1.74}$$

where the expectation is over the channel coefficients  $\{h_0, \dots, h_{L-1}\}$  Assuming Rayleigh fading, we observe that the term  $\sqrt{\sum_{i=0}^{L-1} |h_i|^2}$  has a generalized Rayleigh distribution with n = 2L degrees of freedom. An upper bound for this expectation is given in (A-28) where n = 2L and m = Land  $a = \sqrt{\frac{2\mathcal{E}_b}{N_0}}$ . A high SNR region approximation to the closed form expression for the above expectation is given in various texts [12, 22], as

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) \approx \binom{2L-1}{L} \left(\frac{1}{4\gamma_c}\right)^L,\tag{1.75}$$

where

$$\gamma_c = E\left\{|h_i|^2 \frac{\mathcal{E}_b}{N_0}\right\} = 2\sigma_h^2 \frac{\mathcal{E}_b}{N_0}.$$
(1.76)

An exact expression can be obtained from (A-30), by using n/2 = L and  $a = \sqrt{\frac{2\mathcal{E}_b}{N_0}}$ . For differential phase shit keying (DPSK), which is employed in Chapter 3, the approximate probability of error in the high SNR region is given as [12, 22]

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) \approx \binom{2L-1}{L} \left(\frac{1}{2\gamma_c}\right)^L. \tag{1.77}$$

Throughout this thesis we will use these standard approximations for the BER unless specified otherwise. Note that for both (1.75) and (1.77), we can write that at sufficiently high SNRs

$$P_e (\text{SNR}) \propto \text{SNR}^{-L}.$$
 (1.78)

As per (1.63), the maximum order of diversity or diversity gain d that can be obtained will be L. Other, sub-optimal, forms of diversity combining are selection diversity and equal gain combining [17]. Detailed analysis of these techniques are given available in standard texts [12, 21]. In this thesis, we will target MRC performance when designing our precoders. In the following section we observe how multiple antennas are used to achieve spatial diversity.

### 1.3.2 Diversity via Multi-Antenna Communications

In the absence of sufficient number of distinct paths between the two ends of the communication link, it is natural to induce more paths by using multiple antennas. Multiple antennas can be used at either the transmitter (MISO - multiple input single output), called transmit diversity, or at the receiver (SIMO - single input multiple output), called receive diversity, or at both ends (MIMO - of multiple paths resulting from the use of multiple antennas as parallel channels carrying distinct signals in order to enhance the rate. The multiplexing gain, say m, achievable by a MIMO system is defined as [40]

$$m = \lim_{\text{SNR}\to\infty} \frac{R(\text{SNR})}{\log(\text{SNR})},\tag{1.79}$$

where R(SNR) is the rate (bits/channel use) of the code set being used for communication [43]. Most MIMO research focuses on improving the performance of point to point links by enhancing the diversity or rate of communications (by spatial multiplexing) using multiple antennas. Interestingly, it has been discovered that the spatial multiplexing gain and the diversity gain cannot be simultaneously maximized and there exists a fundamental tradeoff [40] between maximizing these two benefits of MIMO communications. Typical examples of popular space-time communication techniques are space-time block codes (STBC) [38, 80], and Bell Layered Space-Time Architecture (BLAST) first proposed by Foschini in [39]. While BLAST is a spectrally efficient technique, STBCs are used to maximize the available diversity gains and thus, the reliability. Coding and decoding strategies that achieve the optimal trade-off have recently been invented [41]. It is important to note that most of this work is based on the availability of channel state information (CSI) at the receiver. Multiplexing gain, as defined in (1.79), is proportional to the minimum of the number of transmit and receive antennas [40]. For MISO and SIMO this minimum is 1. In this thesis, focus has been laid on MISO systems (see Chapter 2). We look for techniques maximizing the diversity gain that can be achieved using multiple transmit antennas and CSI available only at the transmitter for a multiuser system. Significant research has been conducted on designing precoders for the point-to-point case [42, 44, 45], which can be viewed as a specific case of the multi-user model we are going to pursue in this thesis. In the following section we will review the precoding techniques researched for point-to-point links.

### 1.3.3 Precoding Techniques for point to point MIMO communications

Consider a one-way communication link with  $M_T$  transmit and  $M_R$  receive antennas at the two ends. Linear precoders, transformations on the symbol set at the transmitter before transmission,

and linear decoders, mapping the received signals to symbol estimates can be designed for this system model. The transmitted signal  $\boldsymbol{x} \in \mathbb{C}^{M_T}$ , the corresponding discrete-time received signal at the output of the match filters at the receiver  $\boldsymbol{r} \in \mathbb{C}^{M_R}$ , and the decision statistic  $\boldsymbol{y}$ , generated by the decoder for the transmitted signal  $\boldsymbol{s}$ , are given as

$$\boldsymbol{x} = \boldsymbol{F}\boldsymbol{s} \tag{1.80}$$

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{1.81}$$

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{r}, \tag{1.82}$$

where  $\boldsymbol{H} \in \mathbb{C}^{M_R \times M_T}$  is the channel matrix, with  $[\boldsymbol{H}]_{i,j}$  as the channel between transmit antenna j, and receive antenna i. In case of a flat Rayleigh fading channel, each entry in  $\boldsymbol{H}$  is complex Gaussian with zero mean and variance  $\sigma_h^2$ . Matrices  $\boldsymbol{F}$  and  $\boldsymbol{G}$ , form the precoder and decoder pair to be used at the transmitter and the receiver, respectively. In [42], CSI was assumed available at both, the transmitter and the receiver, and used to derive optimal precoder-decoder pairs under various constraints. One of the optimality criterion chosen there is maximizing the diversity gain. Unconstrained optimization results in the trivial solution of transmitting with infinite power. Constraints can be imposed on the average and/or peak transmit power by constraining the trace or the eigenvalues of the precoding matrix. Subsequent literature [44, 45], which appeared while the work presented in this thesis was in progress, deals with non-linear Tomlinson-Harashima type precoding techniques [46, 47] combined with decision feedback equalization (DFE) [70, 71] at the receivers for improving the diversity performance.

As is evident, this precoding technique reviewed above focused on essentially point to point links and assumed joint channel knowledge at the transmitter and receiver. Our work on the other hand, is in the context of multi-user communications under the assumption that CSI is available only at the transmitter. The following section is a brief review of multi-user communications and its special form, CDMA.

### 1.4 Multi-User Communications and CDMA

Multi-user communications in general refers to the scenario where multiple nodes or users are communicating with each other over a common channel. The simplest example of a multi-user

communication system is two-way communication over a common wired/wireless link. In this thesis we address the downlink or broadcast channel, where different nodes are receiving distinct information via a common channel from a single central transmitter. The dual of this model is the uplink or multiple-access channel, where multiple nodes are trying to transmit to a central receiving node over a shared medium. Common multiple access techniques are: Time Division Multiple Access (TDMA), and Frequency Division Multiple Access (FDMA), wherein the available resources, viz. time span and frequency band for communication, are divided orthogonally among all the users. By orthogonal division, we mean that the time slots and frequency bands assigned to the users are non-overlapping in nature. Consequently, the effective channel for each user appears like a single user channel, since the signals of one user do not interfere with the signals of any other user. These essentially static (in the sense that sharing of resources is pre-determined by the allocation of slots), and slotted techniques based on the principle of dividing the channel into orthogonal slots are spectrally inefficient, specifically when the nature of the traffic is bursty. In case the traffic is bursty, the assigned slots will be unused when no communication is taking place. Random access techniques like those used in Ethernet are suitable only for highly bursty channels [22]. A more popular technique for multiple access communications, which has found widespread acceptance in cellular standards being defined worldwide is Code Division Multiple Access (CDMA). In CDMA, all users are allowed to overlap in both frequency and time. The benefits of this technology have motivated its use in the radio interface specifications for International Mobile Telecommunications-2000 (IMT-2000) [48, 49] by the International Telecommunication Union (ITU). This specification is for Third Generation (3G) cellular communications. CDMA standards included are the Wideband CDMA (WCDMA) for Universal Mobile Telephone System (UMTS) [50, 51], cdma2000 [52, 53] and Time-Division Synchronous CDMA (TD-SCDMA) [54]. Under idealized conditions, assuming equal rate assignments for all users, the capacity of CDMA is identical to that of TDMA and FDMA [22]. However, when the assigned rates are not equal, the achievable sum rates of all users may be higher with CDMA [22]. There are various other benefits of this technology that make it attractive. A great deal of literature has been devoted to the benefits of CDMA and can be found in [12, 57, 58] and references therein. Briefly, the primary advantages are, soft limit on user capacity which depends on the number of users, multipath resolution properties, which enable diversity combining rake reception [59], and dynamic resource sharing which increases

overall system efficiency [30]. We will see the essentials of CDMA in the following sub-sections.

### 1.4.1 Essentials of CDMA

In CDMA, each user is assigned a unique signature waveform. Typically, the signature waveform of a particular user, k, is represented<sup>2</sup> as  $s_k(t)$ . These signature waveforms, which may or may not be orthogonal in nature, are modulated by the user data before transmission. If the signature waveforms are orthogonal, then under ideal conditions, the data of a particular user can be obtained by simply correlating the received waveform with the signature waveform of the corresponding user. No signal can be truly bandlimited as well as time-limited [60]. However, we can employ softer definitions by defining the available bandwidth B and the symbol period  $T_s$  as the frequency and time interval within which 98% of the total energy is constrained. For a given bandwidth B, and symbol rate  $R_s = 1/T_s$ , the number of orthogonal signals that can be constructed is given as  $2B/R_s = 2T_s B$  [61]. Thus, the number of users K, each communicating at the rate  $R_s$ , that can be supported with orthogonal signalling is given by

$$K = 2T_s B = 2\frac{B}{R_s}.$$

Note that high duration-bandwidth product,  $(T_s B = B/R_s)$ , implies the ability to have more users. Signals with high duration- bandwidth product are also called spread spectrum signals [30], and we will often refer to signature waveforms as spreading waveforms. Spread spectrum signals have bandwidths that are much higher than the information rate. Spreading has various other benefits associated with it like robustness to channel distortion, low probability of intercept, better anti-jamming performance and message privacy [22].

Non-orthogonal signalling can replace orthogonal signalling, thereby supporting more users under the duration and bandwidth constraints. Non-orthogonality would also result from asynchronous reception, where the receiver and transmitters are not perfectly synchronous, or from multipath propagation. Non-orthogonality, whether introduced by the channel and other means like asynchronism, or intentionally to increase user capacity, causes interference. Techniques for

<sup>&</sup>lt;sup>2</sup>Note that in Sections 1.1, 1.2 and 1.3 we referred to the transmitted baseband signal as  $s_m(t)$  and the discretetime transmitted symbol as  $s_m$ . Henceforth, throughout this thesis, we assume that the transmitted signal is either +1 or -1 and represented as b, while  $s_m(t)$  represents the spreading waveform for user m. Spreading sequences are represented as  $s_k$  and throughout the thesis we will use  $s_k$  to refer to spreading sequences rather that discrete-time transmitted symbols.

mitigating the associated interference are collectively called multi-user detection [30]. The high spectral efficiency and larger user support resulting from non-orthogonal signalling has motivated our research of precoding techniques for multi-user systems. Before looking at precoding design, first we will look at the generation of spreading waveforms in order to get a deeper understanding of CDMA communications.

Typically, the spreading waveform for user k is generated by using a spreading sequence or spreading code  $\mathbf{s}_k \in \mathbb{C}^{N \times 1}$ . N is also called the spreading gain. The relationship between the spreading code and the spreading waveform is given as [30]

$$s_k(t) = A_k \sum_{i=0}^{N-1} \psi_i(t, \boldsymbol{s}_k),$$
 (1.83)

where  $A_k$  is the transmitted amplitude for user k. We can say that the spreading waveform of each user k is generated using the set of signals  $\Phi_k = \{\psi_i(t, \mathbf{s}_k), i = 0, \dots, N-1\}$ , whose elements are related to the corresponding spreading code  $\mathbf{s}_k$ . An important property of the elements of the set  $\Phi_k$  is that they are orthogonal to each other in some sense i.e.,

$$\langle g(\psi_i(t, \boldsymbol{s}_k)), g(\psi_l(t, \boldsymbol{s}_k)) \rangle = 0, \qquad (1.84)$$

where,  $g(\psi_i(t, \boldsymbol{s}_k))$  is some function of  $\psi_i(t, \boldsymbol{s}_k)$ . Furthermore,

$$\langle \psi_i(t, s_k), \psi_i(t, s_k) \rangle = 1,$$
 (1.85)

so that the generated spreading waveform has unit energy. The power transmitted for user k is then equal to  $A_k^2$ . For simplicity, in the remainder of this thesis we assume that all the receivers receive equal average powers. We can say that  $A_k = 1$ . Different implementations of CDMA are possible depending on the definition of the relation between the signals in the set  $\Phi_k$ , and the corresponding sequence  $s_k$ . These are given below.

### 1. Direct Sequence CDMA (DS-CDMA):

In this case the signals in the set  $\Phi_k$ , are orthogonal in time and linear functions of the elements of the sequence  $s_k$ . Specifically, the functions are given by

$$\psi_i(t, \boldsymbol{s}_k) = s_k[i]p(t - iT_c), \qquad (1.86)$$

where  $s_k[i]$  is the *i*<sup>th</sup> element of the spreading sequence and p(t) is a simple time-limited pulse

$$p(t) = \begin{cases} \zeta & 0 \le t < T_c \\ 0 & 0 > t, t \ge T_c \end{cases} .$$
 (1.87)

Here  $\zeta \neq 0$ , is proportional to the spreading length and the duration  $T_c$ . The duration  $T_c$ over which the pulse has support is called the chipping period and the pulse itself is called a chip or a chipping waveform. In this case, it is easy to see that

$$<\psi_i(t, s_k), \psi_l(t, s_k)> = 0, \quad i \neq l.$$
 (1.88)

Importantly, observe that the orthogonality stems from the time-limited nature of the pulse p(t). To satisfy the normalization condition on the chipping waveform as imposed by (1.85), we choose  $\zeta = \sqrt{\frac{1}{T_c}}$ . For a K user system, all the sets  $\Phi_k, k = 1, \dots, K$ , have elements similarly related to the corresponding sequences  $s_k$ . The definition of the waveforms p(t) can be chosen differently for each sequence, however, we will assume that it is the same for all the sets  $\Phi_k$ . There is a bandwidth expansion associated with DS-CDMA. Observe that  $T_s = NT_c$  when one spreading code is used to spread one transmitted symbol. If the baseband signal bandwidth is  $B = \frac{1}{T_s}$ , then the bandwidth of the signal after spreading will be  $B_{spreading} = \frac{1}{T_c} = NB$ . Therefore, to increase N means to expand the transmission bandwidth.

### 2. Multi-carrier CDMA (MC-CDMA):

Alternately, we can choose the signals in the set  $\Phi_k$ , such that they are orthogonal in the frequency domain. The definition of the elements  $\psi_i(t, \mathbf{s}_k)$  is given in terms of their Fourier transforms  $\Psi_i(f, \mathbf{s}_k)$ . As an example,

$$\Psi_i(f, \boldsymbol{s}_k) = s_k[i]P(f - if_c), \qquad (1.89)$$

where the pulse P(f) is bandlimited in a manner analogous to the time-limited pulse waveform in (1.87) and is given as

$$P(f) = \begin{cases} \zeta & 0 \le f < f_c \\ 0 & 0 > f, f \ge f_c \end{cases} .$$
(1.90)

In this case, it is easy to see that

$$<\Psi_i(f, s_k), \Psi_l(f, s_k) > = 0, \quad i \neq l,$$
 (1.91)

where the orthogonality now stems from the bandlimited nature of P(f). In this case the function  $g(\psi_i(t, \mathbf{s}_k))$  given in (1.84) corresponds to the Fourier transform of the signal  $\psi_i(t, \mathbf{s}_k)$ . This form is actually the frequency domain dual of DS-CDMA.

3. Frequency Hopping CDMA (FH-CDMA) Note that in both DS-CDMA and MC-CDMA, the signals in the set  $\Phi_k$  are linear functions of the elements of the spreading sequence. Non-linear relations can also be defined and one such definition is given below in (1.92). The signals in the set  $\Phi_k$ , can be generated by using the elements of the spreading sequence  $s_k$ , to modulate the frequency of time-limited waveforms. As an example, we can use

$$\psi_i(t, \mathbf{s}_k) = \begin{cases} \cos\left(2\pi s_k[i]t\right) & iT_c \le t \le (i+1)T_c \\ 0 & t > (i+1)T_c, t < iT_c \end{cases}$$
(1.92)

Again it is easy to see that

$$\langle \psi_i(t, \boldsymbol{s}_k), \psi_l(t, \boldsymbol{s}_k) \rangle = 0, \quad i \neq l.$$

Since every chip is now transmitted with a different center frequency, and hence the name frequency hopping.

Note that like DS-CDMA, for MC-CDMA as well as FH-CDMA, the same relationship holds between the signals in the set  $\Phi_k$ , and the corresponding sequence  $s_k$ , for all  $k = 1, \dots, K$ . In this thesis we will focus on DS-CDMA. In most cases the discussion will hold equally for MC-CDMA but for a small difference. As we see in Section 1.4.1, this difference stems from the noise statistics. In case of FH-CDMA the signals in the set  $\Phi_k$  are not linearly related to the spreading sequence  $s_k$ , and the discussion will not hold for it. It has been mentioned here simply for completeness. In the following discussion, unless specified otherwise, we use the term CDMA and DS-CDMA interchangeably. In case of CDMA, irrespective of the set of chipping waveforms used, with perfect timing as well as frequency synchronization between the transmitter and the receiver, orthogonality of different spreading waveforms will depend simply on the orthogonality of the different spreading sequences. We now study the generation of these sequences.

### Spreading Sequences

Ideally, the signal of each interfering user appears like Gaussian noise, thereby indicating that the spreading waveform generating sequence elements should be Gaussian samples [58]. For practical purposes, where reduced complexity is the key, each element of the sequence is selected randomly from the set  $\{+\frac{1}{\sqrt{N}}, -\frac{1}{\sqrt{N}}\}$ . The binary sequence of  $\{\pm 1\}$  is generated and then divided by  $\sqrt{N}$  in order to normalize the spreading sequence to have unit norm, i.e.,

$$||s_k|| = \langle s_k, s_k \rangle^{1/2} = 1.$$
(1.93)

Since generating truly random sequences is practically infeasible, pseudo-random sequences, deterministic sequences which have certain key properties of random Bernoulli sequences [62], are used. The design of optimal spreading sequences depends on various parameters of the system like transmission power, available bandwidth, detection schemes, quality of service (QOS), and other issues. A lot of literature has been devoted to this topic and key contributions can be found in [63, 64, 65]. In this thesis we restrict ourselves to *m*-sequences [68] and to random sequences. Binary sequences are generated using a linear feedback binary shift register. If n - 1 is the length of the shift register, the maximum possible period of the sequence is given by  $2^n - 1$  [67]. If the design is such, the resulting spreading sequences are called maximum length linear feedback shift register sequences or *m*-sequences. In case of random sequences, the sequence elements are independently and randomly chosen from the set  $\{+\frac{1}{\sqrt{N}}, -\frac{1}{\sqrt{N}}\}$ , with each element of the set being equally probable. Random sequences are used in our simulations particularly to act as an upper bound for the average performance of the system.

### Spreading Matrix and Spreading Cross-Correlation Matrix

In dealing with multi-user DS-CDMA we often use matrix notation. This section described the two most important matrices used throughout this thesis. For a K-user system, the spreading sequences used  $s_k, k = 1, \dots, K$ , are represented as columns of the spreading matrix S, given as

$$\boldsymbol{S} = \left[ \begin{array}{cccc} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \cdots & \boldsymbol{s}_K \end{array} \right]. \tag{1.94}$$

Orthogonal spreading sequences have cross-correlation, given as

$$\langle \boldsymbol{s}_j, \boldsymbol{s}_i \rangle = \boldsymbol{s}_i^H \boldsymbol{s}_j = \delta[i-j],$$
(1.95)

39

where  $\delta[i - j]$  is the Kroneckor delta function. The spreading sequence cross-correlation matrix is defined as

$$\mathbf{R} = \mathbf{S}^{H} \mathbf{S}$$

$$= \begin{bmatrix} \langle \mathbf{s}_{1}, \mathbf{s}_{1} \rangle & \langle \mathbf{s}_{2}, \mathbf{s}_{1} \rangle & \cdots & \langle \mathbf{s}_{K}, \mathbf{s}_{1} \rangle \\ \langle \mathbf{s}_{1}, \mathbf{s}_{2} \rangle & \langle \mathbf{s}_{2}, \mathbf{s}_{2} \rangle & \cdots & \langle \mathbf{s}_{K}, \mathbf{s}_{2} \rangle \\ \vdots & \vdots & \vdots \\ \langle \mathbf{s}_{1}, \mathbf{s}_{K} \rangle & \langle \mathbf{s}_{2}, \mathbf{s}_{K} \rangle & \cdots & \langle \mathbf{s}_{K}, \mathbf{s}_{K} \rangle \end{bmatrix}.$$
(1.96)

Observe that the diagonal elements will be unity, since the spreading sequences are normalized, while the non-diagonal elements of the matrix correspond to the cross-correlation between the spreading sequences, i.e.,

$$[\mathbf{R}]_{ij} = \langle \mathbf{s}_j, \mathbf{s}_i \rangle, \qquad i \neq j.$$

$$(1.97)$$

For orthogonal sequences, this matrix will correspond to a  $K \times K$  identity matrix, i.e.,

$$\boldsymbol{R} = \boldsymbol{I}_K. \tag{1.98}$$

When designing spreading codes, the goal is to minimize non-diagonal entries while maximizing the number of sequences. Using m-sequences, we can generate a set of sequences with cross-correlation matrix given as

$$\boldsymbol{R} = \begin{bmatrix} 1 & \frac{1}{\sqrt{N}} & \cdots & \frac{1}{\sqrt{N}} \\ \frac{1}{\sqrt{N}} & 1 & \cdots & \frac{1}{\sqrt{N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} & \cdots & 1 \end{bmatrix}.$$
 (1.99)

### Downlink Communication Using CDMA

Let us see the exact transmission and detection process when DS-CDMA with spreading sequence  $s_1$  is used for a single user BPSK system over a flat fading channel. The channel is assumed static for at least one bit duration  $T_b = NT_c$ . The spreading waveform generated using  $s_1$  will be

$$s_1(t) = \sqrt{\mathcal{E}}_b \sum_{i=0}^{N-1} s_1[i]p(t-iT_c), \qquad 0 \le t \le (N-1)T_c.$$
(1.100)

The transmitted signal will be

$$x(t) = b\sqrt{\mathcal{E}}_{b}s_{1}(t), \qquad 0 \le t \le (N-1)T_{c}$$
  
=  $b\sqrt{\mathcal{E}}_{b}\sum_{i=0}^{N-1} s_{1}[i]p(t-iT_{c}),$  (1.101)

where b is the bit to be transmitted and p(t) is the time-limited pulse as given in (1.87). This transmitted signal has a corresponding discrete-time representation. To see this we will first look at the received signal. The received signal  $r_1(t)$  will be

$$r_1(t) = h_1 x(t) + z_1(t), (1.102)$$

where  $h_1$  is the channel coefficient having a complex Gaussian distribution, and  $z_1(t)$  is a white Gaussian noise process. The receiver front-end will consist of a match filter whose response is matched to the pulse p(t). The output of the filter is sampled at the chipping rate  $R_c = \frac{1}{T_c}$ . N sampled outputs are collected in a vector to form the discrete-time received vector  $\mathbf{r}_1$ , which is a sufficient statistic for estimating the transmitted bit b. Assuming perfect synchronization between the transmitter and receiver, the elements of this vector can be obtained using (1.101) as

$$r_{1}[i] = \int_{iT_{c}}^{(i+1)T_{c}} r_{1}(t)p(t-iT_{c})dt$$
  
=  $\int_{iT_{c}}^{(i+1)T_{c}} (bh_{1}\sqrt{\mathcal{E}}_{b}\sum_{i=0}^{N-1} s_{1}[i]p(t-iT_{c}) + z_{1}(t))p(t-iT_{c})dt$  (1.103)  
=  $h_{1}s_{1}[i]b\sqrt{\mathcal{E}}_{b} + z_{1}[i],$ 

where we have used the normalization property of the pulse as given in (1.85) and (1.87).  $z_1[i]$  is the noise at the match filter output. The vector  $\mathbf{r}_1$  is given as

$$\boldsymbol{r}_1 = h_1 \boldsymbol{s}_1 b \sqrt{\mathcal{E}_b} + \boldsymbol{z}_1, \tag{1.104}$$

where  $z_1$  is a vector of i.i.d. noise terms. The independence of noise terms arises from the orthogonality of the signals in the set  $\Phi_1$ . From (1.104), observe that we can write the transmitted signal in discrete-time as

$$\boldsymbol{x} = \boldsymbol{s}_1 b \sqrt{\mathcal{E}}_b. \tag{1.105}$$

The receiver will correlate the vector  $\mathbf{r}_1$  to the spreading sequence of user 1 to obtain the decision statistic

$$y = \boldsymbol{s}_1^H \boldsymbol{r}_k \tag{1.106}$$

$$= h_1 b \sqrt{\mathcal{E}}_b + n_1, \tag{1.107}$$

where  $n_1 = s_1^H z_1$  has a complex Gaussian distribution. The final decision can now be made using the detector for a single user channel. The operation of correlating the discrete-time received signal with the spreading sequence is called match filtering with respect to the spreading sequence and, sometimes, it is also called as de-spreading. The operation of match filtering the continuous-time received signal  $r_1(t)$  with respect to the pulse shape p(t) as indicated in (1.104) and then correlating the vector  $\mathbf{r}_1$  with the sequence  $\mathbf{s}_1$  as given in (1.106) to generate the decision statistic  $y_1$  can be achieved by a single match filter matched to the spreading waveform  $s_1(t)$  given in (1.100). In CDMA communications if the statistic generated after de-spreading is used as the final decision statistic, the receiver is called a match filter receiver. It is sometimes also referred to as a single user receiver. Multiuser detectors further process the sufficient statistic generated by the match filter before generating the decision statistic. Thus all multi-user detectors have the match-filter

detector embedded within. Observe that the match filter receiver structure has minimal complexity as it does no further processing besides generating the decision statistic using the match filter.

Now consider the downlink transmission using DS-CDMA for a K-user BPSK system over a slow, flat fading channel. The transmitted signal will be given as

$$x(t) = \sum_{k=1}^{K} s_k(t) b[k] \sqrt{\mathcal{E}}_b, \qquad (1.108)$$

where b[k] is the bit to be transmitted for the  $k^{th}$  user. The signal received by user k,  $r_k(t)$ , will be

$$r_k(t) = h_k x(t) + z_k(t), (1.109)$$

where  $h_k$  is the channel coefficient from the transmitter to the receiver and  $z_k(t)$  is a white Gaussian noise process at the receiver of user k. Since the same pulse p(t) is used by all users, the resulting signal at the output of the match filter i at user k,  $r_k[i]$ , will be

$$r_{k}[i] = \int_{iT_{c}}^{(i+1)T_{c}} \left( h_{k} \sum_{j=1}^{K} s_{j}(t) \sqrt{\mathcal{E}}_{b} b[j] + z_{k}(t) \right) p(t - iT_{c}) dt$$

$$= \int_{iT_{c}}^{(i+1)T_{c}} b \left( h_{k} \sum_{j=1}^{K} \sum_{i=0}^{N-1} s_{j} \sqrt{\mathcal{E}}_{b} b[i] p(t - iT_{c}) + z_{k}(t) \right) p(t - iT_{c}) dt \qquad (1.110)$$

$$= h_{k} \sum_{j=1}^{K} s_{j}[i] \sqrt{\mathcal{E}}_{b} b[j] + z_{k}[i],$$

where  $z_k[i]$  is the noise component at the match filter output. The vector  $\mathbf{r}_k$  formed by collecting the match filter outputs will be

$$\boldsymbol{r}_{k} = h_{k} \begin{bmatrix} \sum_{j=1}^{K} s_{j}[1] \sqrt{\mathcal{E}}_{b} b[j] \\ \vdots \\ \sum_{j=1}^{K} s_{j}[N] \sqrt{\mathcal{E}}_{b} b[j] \end{bmatrix} + \underbrace{\begin{bmatrix} z_{k}[1] \\ \vdots \\ z_{k}[K] \end{bmatrix}}_{\boldsymbol{z}_{k}}$$

$$= h_{k} \sqrt{\mathcal{E}}_{b} \begin{bmatrix} s_{1}[1] \cdots s_{K}[1] \\ \vdots \\ s_{1}[N] \cdots s_{K}[N] \end{bmatrix} \underbrace{\begin{bmatrix} b[1] \\ \vdots \\ b[K] \end{bmatrix}}_{\boldsymbol{b}} + \boldsymbol{z}_{k}$$

$$= h_{k} \boldsymbol{S} \boldsymbol{b} \sqrt{\mathcal{E}}_{b} + \boldsymbol{z}_{k},$$

$$(1.111)$$

where we have used the notation for spreading matrix given in (1.94). Therefore, we can write the discrete-time transmitted signal as

$$\begin{aligned} \boldsymbol{x} &= \sum_{k=1}^{K} \boldsymbol{s}_{k} \boldsymbol{b}[k] \sqrt{\mathcal{E}}_{b} \\ &= \boldsymbol{S} \boldsymbol{b} \sqrt{\mathcal{E}}_{b}. \end{aligned} \tag{1.112}$$

From now on, unless specified otherwise we will incorporate the factor  $\sqrt{\mathcal{E}}_b$  in the vector **b**, i.e., we will say that  $b[k] = \{\pm \sqrt{\mathcal{E}}_b\}$ . Thus, we will write the transmitted signal as

$$\boldsymbol{x} = \boldsymbol{S}\boldsymbol{b}.\tag{1.113}$$

and the signal received at user k as

$$\boldsymbol{r}_k = h_k \boldsymbol{S} \boldsymbol{b} + \boldsymbol{z}_k. \tag{1.114}$$

Note that with perfect synchronization, for MC-CDMA the received signal can be written in a similar form. However, the noise vector will have correlated entries, since, the bandlimited pulse in (1.90) is not time-limited. But we do have the same discrete-time model as (1.104). Thus, even though we have constrained ourself to DS-CDMA, most of the discussion in this thesis will apply equally to MC-CDMA. Only the colored nature of the vector  $\boldsymbol{n}$  will need to be accounted for.

The detector at receiver k processes the vector  $\mathbf{r}_k$  in order to estimate the bit transmitted for user k. The properties of  $\mathbf{r}_k$  will depend on those of the spreading sequences, the fading channel and the background noise. The performance and the structure of the receiver will depend on the properties of the spreading codes. We will assume that the receiver has complete knowledge of the spreading code of user k. The receiver will correlate the vector  $\mathbf{r}_k$  to the spreading sequence of user k to obtain the decision statistic

$$y[k] = \boldsymbol{s}_k^H \boldsymbol{r}_k \tag{1.115}$$

$$= h_k[\mathbf{R}]_{1,:}\mathbf{b} + n[k], \qquad (1.116)$$

where  $\mathbf{R} = \mathbf{S}^H \mathbf{S}$  as given in (1.96), is the spreading sequence cross-correlation matrix and n[k] is the noise after match-filtering. If the spreading codes are orthogonal, i.e.,

$$\boldsymbol{R} = \boldsymbol{I},\tag{1.117}$$

then

$$y[k] = \mathbf{s}_k^H \mathbf{r}_1 = h_k b[k] + n[k], \qquad (1.118)$$

where we have used the fact that the spreading codes have unit energy. The noise will be Gaussian and we will effectively have a single user fading channel. The final decision can simply be made as

$$\hat{b}[k] = \operatorname{sign} \, \Re\{y[k]\}. \tag{1.119}$$

However, if the spreading codes are not orthogonal, we see from (1.116) that the decision statistic in (1.118), will contain interference terms. The BER performance will be limited by the interference introduced. Figure (1.2), shows the effect of this interference for a CDMA system with K = 10 users. The spreading codes are assumed to have equal cross-correlation of  $\rho = \langle s_i, s_j \rangle = 0.7, i \neq j$ . We can stack the decision statistics formed at each of the receivers to form the vector of decision statistics  $\boldsymbol{y}$  given as

$$\boldsymbol{y} = \begin{bmatrix} y[1] & \cdots & y[K] \end{bmatrix}^T.$$
(1.120)

In the following chapter we will see how this vector can be expressed in terms of the spreading matrix, the spreading cross-correlation matrix and the channel of each user.

To eliminate this multi-user interference, the receiver can use advanced signal processing techniques. This resulting complex receivers are also referred to as multi-user receivers and excellent research has been conducted on these [30, 69]. Precoding techniques for multi-user communications researched in the past were focused on using the *a priori* knowledge of interference available to the transmitter to cancel the same. Some of the precoding techniques researched in the past emphasized reducing receiver complexity. In the following section, we review precoding solutions proposed for interference cancellation over scalar (single user, single antenna) channels, as well as for broadcast communications.



Figure 1.2: Figure illustrates the BER vs SNR performance of a CDMA downlink system with K = 10 users over a Rayleigh fading channel. The spreading code of each user is equally correlated to that of all other users with correlation  $\rho = 0.7$ . The receivers are constrained to be simple match filter receivers.

## 1.5 Precoding for Interference Cancellation in Single and Multiuser Communications

The idea of precoding for interference cancellation is not new and finds it roots in papers by Tomlinson [46], and Harashima *et al* [47], both written more than 3 decades ago for eliminating ISI in single user wireline channels. Decision feedback equalizers, which use a feedforward and feedback filter pair to cancel ISI at the receiver, are often used in wireline channels [70, 71]. In a technique called pre-equalization [22], the feedback filter of the DFE is placed at the transmitter in order to completely eliminate propagation errors occurring from the use of a feedback filter at the receiver. This however, led to the problem of increased transmission power [22]. Tomlinson and Harashima proposed the use of non-linear precoding, consisting of pre-subtracting the ISI interference from the symbol to be transmitted and then applying modulo arithmetic to control the total transmitted power. Information theoretic results were obtained a decade later in a paper by Costa, where an excellent analogy was made with "writing on dirty paper" [72]. It says, if we are writing on a piece of dirty paper, we would write so as to avoid overlap with the dirty parts. In a similar manner, we can pre-adapt the transmission over a dirty channel (channel with interference), in order to avoid any form of interference we know the receiver will experience. Costa proved that for a scalar AWGN channel with Gaussian interference, known non-causally at the transmitter and unknown to the receiver, the achievable capacity is identical to that of the scalar AWGN channel without interference. The technique of coding based on pre-subtracting interference known a priori to the transmitter, is since known as Costa precoding. Costa precoding was recently used in [73] to show that the sum capacity of a vector Gaussian broadcast channel [43], is identical to a that of a multi-antenna channel with cooperating receivers, which corresponds to a multi-access channel. Recently, a generalization of Tomlinson-Harashima precoding was proposed in the form of trellis precoding [74]. Note that in a broadcast channel, each receiver signal acts as interference for the other receivers. The authors in [74] proposed a technique to approach the achievable sum rates in a vector Gaussian broadcast channel. All these approaches are information-theoretic and targeted towards computing or maximizing the achievable rate with multiple antennas. Multiple antennas can be used for cancelling interference as well. In [75], a tradeoff between achievable rate, interference cancellation and diversity has been shown to exist for a multiple-access channel with multiple antennas. No results have been provided on the existence of a similar tradeoff for the downlink channel yet. The existing capacity-achieving techniques are very complex at both ends and do not necessarily provide diversity. In this thesis we are concerned with the design of low complexity techniques for the broadcast channel. Our primary target is minimization of receiver complexity and better energy efficiency via diversity communications. Unlike trellis-type non-linear, high complexity techniques, we look for simple linear precoders. Similar to these techniques, our goal is to cancel known interference, but we also target diversity maximization under transmit power constraints. The following section defines the problem of precoding as addressed in this thesis.

## 1.6 Problem Formulation: Precoding for Multi-Antenna Multi-User Communications

In Section 1.3.3, we saw how pairs of linear precoders and decoders are used for single user MIMO channels. These linear precoders and decoders can jointly minimize the mean square error (MSE) between the transmitted and received symbols [42]. Joint transmission schemes exploit the availability of channel knowledge at both, the transmitter and receiver, to increase either the rates, or maximize the diversity while keeping the rate fixed, have been considered for single user MIMO channels [42]. Precoders for Gaussian, as well as frequency selective, broadcast channels are derived in [9] where Rake receivers and perfect receiver CSI is assumed for the multipath channels. In [3] the authors develop an adaptive precoding strategy with blind channel estimation at the transmitter, but due to the absence of Rake receivers or any other diversity combining strategy, no diversity gain is available. In [8], pre-rake diversity combining has been introduced which maximizes the diversity gain for frequency selective channels without receiver CSI. Precoding techniques for single antenna frequency selective channels were also considered in [7], where the precoders were applied to spread CDMA waveforms. All of these techniques are essentially linear and some of these focus on reducing receiver complexity [3, 7, 8]. Based upon the above review of linear precoding techniques, we make the following observations:

- 1. Precoding has so far been studied for point-to-point, multi-antenna channels and for single antenna broadcast channels. The study of linear precoding techniques for multi-user, multiantenna communications has not been addressed yet.
- 2. The multi-user precoding techniques studied so far have focused on interference cancellation. There was no emphasis on pre-processing techniques for diversity exploitation for single antenna but multipath channels without receiver CSI.
- 3. Most of the results presented so far have assumed perfect CSI availability at the transmitter and/or receiver. Modelling the performance of precoding when partial CSI is available to the transmitter has neither been studied for single antenna multi-user systems, nor for point-to-

point multi-antenna systems.

In our approach we consider linear precoding for multi-user systems corresponding to the multiinput single-output (MISO) CDMA downlink. Our task is to reduce the complexity of the receiver. So far we have been referring to ultra-low complexity receivers as having minimal complexity. Receivers employing MRC to exploit diversity require channel knowledge. To minimize the receiver complexity, we will assume only a single antenna receiver having no channel information and focus on using transmit antenna diversity techniques. As we will see in Chapter 2, the use of a single receive antenna is further motivated by the size of mobile receivers. Based on the discussion in Section 1.1.1 and 1.4.1, we saw that the receiver for a single user CDMA consists of a match filter matched to the chipping waveform followed by match filtering with respect to the spreading sequence of the user. As mentioned in Section 1.4.1, this structure called match filter receiver or single user receiver is embedded in all multi-user receivers. Hence, we say that the match-filter receiver has minimal complexity. The use of only match filter receivers was first proposed in the form of ultralow complexity receivers in [3]. We constrain our receivers to be such ultra-low complexity receivers. Our goal will be interference suppression and simultaneous diversity exploitation. Unlike point to point MIMO, joint decoding using signals received at different antennas cannot be performed in this case. We therefore employ CDMA to enable signal separation at the receivers. We define linear precoders as linear transformations on the transmitted signal for suppressing interference while maximizing the diversity gain. Our objective is to minimize the receiver complexity using optimal linear precoders at the transmitter. Our optimality constraint is to minimize the average BER at all the receivers, while keeping the transmit power fixed. The primary contributions of this thesis will be:

- We propose the use of linear precoders to exploit multi-antenna diversity and simultaneously cancel interference for a multi-user multi-antenna CDMA downlink system using ultra-low complexity receivers. Part of this work has been published in [100]
- 2. We derive and analyze the performance of linear precoders for multi-user multi-antenna CDMA communications over flat fading channels assuming perfect CSI at the transmitter. Part of this work has been published in [100].

- 3. We derive and analyze the performance of linear precoders for multi-user multi-antenna CDMA communications over flat fading channels assuming partial CSI at the transmitter. The precoders are derived for a general model on partial CSI modelled as knowledge of some quantity statistically correlated to the channel. The performance resulting from applying these precoders to a system where partial CSI is available to the transmitter in the form of old channel estimates is evaluated. Part of this work has been published in [100, 108].
- 4. We derive and analyze the performance of linear precoders for multi-user multi-antenna CDMA communications over frequency selective fading channels with and without ISI assuming perfect CSI at the transmitter. The precoders are derived to exploit multi-antenna diversity as well as frequency diversity. Part of this work has been published in [100, 109].
- 5. We derive linear precoders for multi-user multi-antenna CDMA communications over frequency selective fading channels with and without ISI assuming partial CSI at the transmitter. The precoders are derived to exploit multi-antenna diversity as well as frequency diversity. Part of this work has been published in [100, 109].

In Chapter 2, we will look at precoding techniques for flat fading channels. There we will derive precoders assuming perfect CSI availability at the transmitter and analyze their BER performance. In Chapter 3, we will derive precoders assuming that partial CSI is available to the transmitter. We will evaluate the resulting performance using simulations. In Chapter 4, we will derive precoders for interference cancellation and diversity maximization over frequency selective channels. Simulation results are presented to evaluate the performance of the precoders derived there. The work in this thesis is summarized in [100, 108, 109].

# **Precoding for Flat Fading Channels**

We know from Chapter 1 that non-orthogonality of spreading sequences results in multi-user interference (MUI). However, non-orthogonal sequences provide the benefit of enhanced user capacity. Even if orthogonal sequences are used, the multipath nature of the wireless channel and asynchronous transmissions destroy the orthogonality of spreading sequences. Thus, multi-user detection (MUD) techniques for cancelling multi-user interference need to be implemented at the receiver. Temporal spreading of the signal in case of frequency selective channels causes intersymbol interference (ISI). Therefore, besides MUD techniques, channel equalization techniques for cancelling the ISI introduced are needed. MUD techniques are implemented using either the knowledge of spreading sequences of all users, or more complicated blind/semi-blind approaches to MUI cancellation [69]. Equalization techniques require the decentralized receivers to know their downlink channel coefficients, which requires channel estimation at the receiver. Channel estimation is required in case MRC is to be employed for diversity combining. The known complexity of these techniques motivates the use of transmitter based signal processing to reduce receiver complexity. This of course, will require channel knowledge at the transmitter. Channel knowledge can be easily obtained at the transmitter provided a time-division-duplex (TDD) system, where the uplink and the downlink use the same band of frequencies to communicate in orthogonal time-slots, is implemented. In such systems either training based, or blind channel estimation algorithms could be implemented. In systems where the uplink and downlink use orthogonal frequency bands, the receiver could echo, i.e., transmit the received waveform samples over the uplink [7]. In any case, channel state information (CSI) can be made available to the transmitter without having the receiver perform any kind of channel estimation. In this chapter, we will assume that the transmitter has somehow obtained perfect CSI. In later chapters, we assume that the receiver has limited CSI. Our main objective in this chapter is to design precoders for minimizing the receiver complexity. Therefore, we design precoders to not only cancel the MUI resulting from non-orthogonal spreading, but also to simultaneously exploit diversity provided by the use of multiple transmit antennas. We assume the channels of all users to be flat and slowly fading, so that the available CSI can be used to design the precoders.

In this chapter, we will consider a CDMA system employing non-orthogonal spreading. Our transmitter is equipped with multiple antennas while each receiver has a single antenna. The use of a single antenna at each mobile is reasonable. To achieve good diversity gain from the use of multiple receive antennas, the corresponding paths over which the signal is received should be uncorrelated. To achieve uncorrelated paths it is necessary that any two antennas be spatially apart by a few wavelengths. Omnidirectional antennas should be at least  $0.4\lambda_c$  apart [76], whereas loop antennas should be spaced  $0.6\lambda_c$  apart [76], where  $\lambda_c$  is the carrier wavelength. In the band between 1 GHz to 6 GHz, the wavelength varies from 30cm to 5cm. Receivers for CDMA downlink presently available are less than 5cm in length. In sensor networks, miniature size sensors are used. These prohibit the use of more than a single antenna, thereby not only justifying, but also motivating our single receive antenna assumption. In Chapter 1, we saw that the receiver of each user match filters the received signal with respect to the spreading sequence it has been assigned. We constrain our receivers to be simple match filters performing no additional processing to form the final decision statistic.

We begin with a brief introduction to existing precoding techniques for CDMA. Next, we look at the choice of a suitable diversity exploiting technique for our transmit diversity scenario. We will describe our system model in Section 2.3. The following Sections 2.4 and 2.5, will focus on deriving the precoders. Section 2.6 will focus on performance analysis and finally we will conclude. Part of the work presented here appears in [100].

## 2.1 Introduction

We saw in (1.113), that for a K-user CDMA system, the discrete-time, downlink transmitted signal can be written as

$$\boldsymbol{x} = \boldsymbol{S}\boldsymbol{b},\tag{2.1}$$

where S is the spreading matrix and b contains the K data symbols of the K users. With channel knowledge available to the transmitter, two primary techniques are possible. One is spreading code optimization while the other is precoding. In case of spreading code optimization, depending on the CSI available, spreading codes are chosen to optimize performance under some suitable optimization criteria. Implementing adaptive codes would require joint optimization of the filters to be used at the transmitter and the receiver. This necessitates receiver participation and would not reduce the receiver complexity. Precoding for single-antenna multi-user communications was first introduced in [9]. A precoder was applied to the data to be transmitted, before spreading so as to achieve interference suppression with simple match filter receivers. Note that since the precoding filter is applied to the data before spreading, we refer to this technique as a data-filtering approach. For a Gaussian channel, it suffices to use a match filter receiver. For a multipath channel though, that work requires a Rake receiver [59] (see Chapter 4), which needs channel estimation. Another technique, called, pre-filtering, which, in contrast with the data filtering approach used in [9], used a waveform filtering approach, was proposed in [7]. By waveform filtering, we mean that the preprocessing filter is applied to the waveform generated after the final spreading operation. If we denote the linear pre-processing operation by a matrix M, the transmitted signal when employing the data-filtering approach will be

$$\boldsymbol{x} = \alpha \boldsymbol{S} \boldsymbol{M} \boldsymbol{b}, \tag{2.2}$$

and that when employing the waveform filtering approach, will be

$$\boldsymbol{x} = \alpha \boldsymbol{M} \boldsymbol{S} \boldsymbol{b}. \tag{2.3}$$

The constant  $\alpha$ , is used to represent the scaling operation required when a transmit power constraint is imposed at the transmitter. The constraint can be on the on individual transmit powers or on the total transmit power. In the former case, we replace  $\alpha$  in (2.2) and (2.3) by a diagonal matrix, depending on the target user signal to noise ratio (SNR) specifications. When all users are to have the same target SNR, we can use a scalar  $\alpha$  to constrain the total transmit power. In this thesis, we assume that all the users have the same target SNRs and hence, use a scalar  $\alpha$ . We will choose  $\alpha$  such that the resulting transmit power with precoding is identical to that without precoding. We will refer to the two techniques illustrated in (2.2) and (2.3) as **SMb**-precoding and **MSb**-precoding respectively. Assume that we are employing **SMb**-precoding. From (2.1) and (2.2), the transmitted power  $P_{no-precoding}$ , in the absence of precoding will be given as

$$P_{no-precoding} = E_{\boldsymbol{b}} \{ \boldsymbol{b}^{H} \boldsymbol{S}^{H} \boldsymbol{S} \boldsymbol{b} \}$$

$$= E \{ \boldsymbol{b}^{H} \boldsymbol{R} \boldsymbol{b} \},$$
(2.4)

and that with precoding  $P_{precoding}$  is given as

$$P_{precoding} = E_{\boldsymbol{b}} \{ \alpha^2 \boldsymbol{b}^H \boldsymbol{M}^H \boldsymbol{S}^H \boldsymbol{S} \boldsymbol{M} \boldsymbol{b} \}$$

$$= E \{ \alpha^2 \boldsymbol{b}^H \boldsymbol{M}^H \boldsymbol{R} \boldsymbol{M} \boldsymbol{b} \}.$$
(2.5)

Equating the two we have

$$\alpha = \sqrt{\frac{E\{\boldsymbol{b}^{H}\boldsymbol{R}\boldsymbol{b}\}}{E\{\boldsymbol{b}^{H}\boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b}\}}}.$$
(2.6)

Since  $b[i] = \pm \sqrt{\mathcal{E}}_b$  and since the symbols of different users are uncorrelated,

$$E\{\boldsymbol{b}\boldsymbol{b}^H\} = \mathcal{E}_{\boldsymbol{b}}\boldsymbol{I}_k,\tag{2.7}$$

and for any general matrix  $\boldsymbol{Z}$ ,

$$E\{\boldsymbol{b}^{H}\boldsymbol{Z}\boldsymbol{b}\} = E\{\operatorname{tr}\{\boldsymbol{b}\boldsymbol{b}^{H}\boldsymbol{Z}\}\} = \operatorname{tr}\{\boldsymbol{Z}\}.$$
(2.8)

Since the spreading sequences are normalized as given in (1.93), the diagonal elements of  $\mathbf{R}$  will be 1. Hence, using (2.8) we get

$$P_{no-precoding} = E\{\boldsymbol{b}^{H}\boldsymbol{R}\boldsymbol{b}\} = K\mathcal{E}_{b}, \qquad (2.9)$$

and

$$E\{\boldsymbol{b}^{H}\boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b}\} = \operatorname{tr}\{\boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{M}\}.$$
(2.10)

Using (2.9) and (2.10) in (2.6), we get

$$\alpha = \sqrt{\frac{K\mathcal{E}_b}{\operatorname{tr}\{\boldsymbol{M}^H \boldsymbol{R} \boldsymbol{M}\}}}.$$
(2.11)

Note that this only illustrates the computation when a single antenna is used. For multiple antennas the term  $P_{precoding}$ , corresponding to total transmitted power with precoding needs to incorporate multiple terms. We will see this in Section 2.6.

Data-filtering precoders were derived in [3] for frequency selective channels, where ultra-low complexity, match filter receivers were assumed. The absence of a Rake receiver in this case resulted in no exploitation of the available multipath diversity. Besides, neither [3], nor [7, 9] considered multi-antenna transmissions for diversity exploitation. This has motivated our problem of precoding for interference cancellation and simultaneous diversity exploitation for a multi-antenna CDMA system.

## 2.2 Multi-Antenna Technique with CSI at the Transmitter

Different space-time communication schemes have been proposed in the past decade for exploiting the benefits of multi-antenna communications. The primary benefits involved are rate enhancement via spatial multiplexing, which creates parallel channels across the transmitter and receiver, and improved reliability achieved by reducing the bit-error-rate (BER) for a given SNR using diversity techniques. Our primary focus is on transmit diversity for MIMO multi-user communications. We first look at the single user point-to-point MIMO signalling techniques to motivate a suitable technique for our scheme.

It is interesting to note that if the same signal is transmitted from different antennas, no diversity benefit is obtained if the two distinct paths cannot be de-correlated at the receiver. Suppose that we are using two transmit antennas to communicate to a receiver with a single antenna over a flat Rayleigh fading channel. We denote the path coefficients on the two flat fading channel will by,  $h_1$  and  $h_2$ . Assume that BPSK transmission is being employed with  $b = \{\pm \sqrt{\mathcal{E}}_b\}$ . Let the same signal, say

$$x = \sqrt{\mathcal{E}_b/2},\tag{2.12}$$

be transmitted from both the antennae. The factor 2 is used to normalize the total transmit power

to  $\mathcal{E}_b$ . The received signal will be  $r = (h_1 + h_2)x + n$ , where  $n \in \mathcal{N}_c(0, \frac{N_0}{2})$  is a white Gaussian noise term. The decision statistic generated at the receiver will be  $y = \pm |h_1 + h_2|^2 \sqrt{\mathcal{E}_b/2} + (h_1 + h_2)^* n$ . The variance of the real part of the noise will be  $\frac{N_0|h_1+h_2|^2}{2}$ . Similar to the analysis done previously in Chapter 1, the average BER will be the average pairwise probability of error. The resulting BER for a particular SNR,  $\frac{\mathcal{E}_b}{N_0}$ , conditioned upon channel knowledge will be

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) = Q\left(\sqrt{\frac{\mathcal{E}_b}{\frac{N_0|h_1+h_2|^2}{2}}}\frac{|h_1+h_2|^2}{\sqrt{2}}\right)$$
$$= Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_0}}\frac{|h_1+h_2|}{\sqrt{2}}\right)$$
$$= Q\left(\sqrt{\frac{2|h|^2\mathcal{E}_b}{N_0}}\right), \qquad h = \frac{|h_1+h_2|}{\sqrt{2}}.$$
(2.13)

Since  $|h| = \frac{|h_1+h_2|}{\sqrt{2}}$  has the same distribution as  $|h_1|$  (which is the same as  $h_2$ ), the performance will be similar to that of a single path fading channel considered in Section 1.2.5. The BER being identical to that of a single antenna fading channel, its relation to the SNR in the high SNR region, i.e., as SNR  $\rightarrow \infty$  will be as given in (1.62)

$$P_e(\text{SNR}) \propto \text{SNR}^{-1}$$
.

There is no real benefit derived by using this technique, which in literature is also referred to as *simulcasting*. This led to the invention of the delay diversity scheme [77, 78], wherein the same symbol was transmitted, but with a single symbol period delay between transmissions from the two antennas. This enables path separation followed by maximum ratio combining type diversity combining techniques at the receiver. Other techniques came up, the significant ones being space-time trellis codes (STTC) [79], and space-time block codes (STBC) [38, 80]. STTC had high complexity but provided a tradeoff between rate enhancement, diversity and complexity, while providing a coding gain without any bandwidth expansion. STBC was an excellent technique based on orthogonal designs [79] and involved simple linear processing at the receiver. These techniques were essentially based upon channel knowledge at the receiver. In fact multiple antenna transmit diversity techniques can be classified into four primary categories:

 Transmitter has no channel knowledge, but receiver has complete knowledge of the channel [77, 83, 84], obtained via training or other means.

- 2. Transmitter has channel information while receiver does not. This is the case when TDD type systems [82] are used to make channel information available to the transmitter.
- 3. Both transmitter and receiver have channel knowledge. This corresponds to category 1 in conjunction with feedback [81] or category 2 with feedforward. This is an interesting case which has been studied in depth in [42] in terms of achievable diversity.
- 4. Blind schemes where channel information is available neither to the transmitter, nor to the receiver. Phase sweeping techniques used in conjunction with channel coding techniques are proposed in [85], while differential encoding and decoding techniques based upon unitary matrices are proposed in [98, 99].

In this thesis we are concerned with category 2, where only the transmitter has channel information. The STTC and STBC techniques are applicable to category 1, where only the receiver has channel information. When there is only one receive antenna and the transmitter has channel knowledge while the receiver has none, we can use maximum ratio weighting (MRW) [86], which is illustrated below.

Consider a BPSK system with  $M_T$  transmit antennas employed over a flat, Rayleigh fading channel. Let the channel coefficient from transmit antenna *i* be  $h^{(i)}$ . We assume that the transmit antennas are spaced sufficiently apart so that the effective paths are uncorrelated (and independent, since Gaussian). With MRW, the signal transmitted from antenna *i* is given by

$$x^{(i)} = \frac{h^{(i)*}}{\sqrt{\sum_{j=1}^{M_T} |h^{(j)}|^2}} b, \qquad (2.14)$$

where  $b \in \{\pm \sqrt{\mathcal{E}_b}\}$ , is the signal to be transmitted across the channel. Note that power normalization is implicit in the weighting process and there is no need to use any explicit factor, like that used in case of simulcasting (2.12). The received signal r will be a weighted sum of the signals transmitted from the different antennae,

$$r = \sum_{i=1}^{M_T} h^{(i)} x^{(i)} + n, \qquad (2.15)$$

where  $n \sim \mathcal{N}_c(0, \frac{N_0}{2})$  is the noise term. The discrete received signal also corresponds to the decision

statistic y used at the receiver. Thus,

$$y = \sum_{i=1}^{M_T} h^{(i)} \frac{h^{(i)*}}{\sqrt{\sum_{j=1}^{M_T} |h^{(j)}|^2}} b + n$$
  
=  $\pm \sqrt{\sum_i |h^{(i)}|^2 \mathcal{E}_b} + n.$  (2.16)

The BER expression can be derived using pairwise error probabilities in a manner similar to that in Section 1.1.1. The BER conditioned upon the channel for a particular SNR will be

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\right) = Q\left(\frac{\sqrt{\sum_i |h^{(i)}|^2 \mathcal{E}_b}}{\sqrt{N_0/2}}\right)$$
$$= Q\left(\sqrt{\frac{2\sum_i |h^{(i)}|^2 \mathcal{E}_b}{N_0}}\right),$$
(2.17)

which is similar in form to (1.73). Consequently, for MISO systems with perfect CSI at transmitter, the performance of MRW is identical to that of MRC with perfect CSI at the receiver. We have already seen in Chapter 1 that when channel knowledge is available at the receiver, MRC results into a maximum likelihood receiver. The ML receiver, which is also the best error minimizing receiver, since the input symbols are equally likely, maximizes the available diversity gain as we saw in Section 1.3.1. Thus, our key conclusion from this section is that for a single user MISO system, when perfect CSI is available at the transmitter, the available diversity gain is maximized if we employ MRW.

## 2.3 System Model

We consider a K-user synchronous CDMA downlink. The spreading codes are of length N and given by the matrix  $\mathbf{S} \in \mathbb{C}^{N \times K}$ . Each spreading waveform is normalized so that  $||\mathbf{s}_k|| = 1$ . In general the transmitted symbols can correspond to any modulation format. For simplicity though, we will stick to BPSK. We have every symbol of the vector  $\mathbf{b}$ , as  $b[k] \in \{\pm \sqrt{\mathcal{E}}_b\}$ . The transmitter is equipped with  $M_T$  antennas, while each receiver is equipped with 1 antenna.

The precoder employed at transmitter a, is represented by  $M^{(a)} \in \mathbb{C}^{K \times K}$ . We will assume **SMb**-precoding. **MSb**-precoding has been shown to perform better when code-adaptation is possible. This though, is difficult given the fact that our goal is to use ultra-low complexity

receivers. In this thesis, we will focus on deriving precoders when the spreading codes are fixed, and hence we will ignore the MSb case. The signal transmitted from antenna a is given as

$$\boldsymbol{x}^{(a)} = \alpha \boldsymbol{S} \boldsymbol{M}^{(a)} \boldsymbol{b}, \tag{2.18}$$

where  $\alpha$  is the power scaling factor as mentioned above. The average transmitted power will be

$$P_{precoding} = E_{\boldsymbol{b}} \left\{ \sum_{a=1}^{M_T} \boldsymbol{b}^H \boldsymbol{M}^{(a)H} \boldsymbol{S}^H \boldsymbol{S} \boldsymbol{M}^{(a)} \boldsymbol{b} \right\}.$$
 (2.19)

Since we are assuming that the total transmitted power is given by  $K\mathcal{E}_b$  as indicated in (2.9), we have

$$\alpha = \sqrt{\frac{K\mathcal{E}_{b}}{E_{b}\left\{\sum_{a=1}^{M_{T}} b^{H} M^{(a)H} S^{H} S M^{(a)} b\right\}}}$$
$$= \sqrt{\frac{K\mathcal{E}_{b}}{E_{b}\left\{\sum_{a=1}^{M_{T}} \operatorname{tr}\left\{bb^{H} M^{(a)H} S^{H} S M^{(a)}\right\}\right\}}}$$
$$= \sqrt{\frac{K}{\sum_{a=1}^{M_{T}} \operatorname{tr}\left\{M^{(a)H} R M^{(a)}\right\}}},$$
(2.20)

where we have used the fact that  $E_{\mathbf{b}}\{\mathbf{b}^H\mathbf{b}\} = \mathcal{E}_b\mathbf{I}_K$ . In the absence of a power constraint, we could simple transmit with infinite power. This not being practical, we use  $\alpha$  to impose the power constraint. In the following sections we will derive and characterize the performance of linear precoders  $\mathbf{M}^{(a)}$  operating over flat fading channels. The goal of our design is to chose precoders such that, conditioned upon the channel, the resulting decision statistics at each of the receivers is close, in the MMSE sense, to that due to MRW in a single user flat fading channel. Note that we make two important assumptions:

- 1. The channels for all users are independently and identically distributed (i.i.d.), and Rayleigh quasi-static fading.
- 2. The average received power, and hence the average SNR at each receiver is the same.

### 2.3.1 Maximum Ratio Weighting and Precoder Design

We saw in Section 1.4.1 that each receiver k match filters the discrete-time received signal with respect to the spreading sequence,  $s_k$ , that it has been assigned to form a decision statistic y[k]. We form the vector of these decision statistics represented as  $\boldsymbol{y}$  as

$$\boldsymbol{y} = \left[ \begin{array}{ccc} y[1] & y[2] & \cdots & y[K] \end{array} \right]^T.$$
(2.21)

We concluded from Section 2.2 that for a single user system, i.e., in the absence of other users, MRW will maximize the achievable diversity gain. We will refer to the decision statistic resulting from MRW in the absence of other users as single user MRW. Let  $h_k^{(a)}$ , be the channel from transmit antenna *a* to receiver *k*. In the absence of all other users, the decision statistic resulting at receiver *k* from single user MRW will be

$$y'[k] = b[k] \sqrt{\sum_{a=1}^{M_T} |h_k^{(a)}|^2} + n[k], \qquad (2.22)$$

where  $n[k] \sim \mathcal{N}_c(0, \frac{N_0}{2})$  is the Gaussian noise term resulting after match filtering with respect to  $s_k$  as indicated in (1.116). This statistic can also be written as

$$y'[k] = \sum_{a=1}^{M_T} y^{(a)'}[k], \qquad (2.23)$$

where

$$y^{(a)'}[k] = \frac{|h_k^{(a)}|^2}{\sqrt{\sum_{j=1}^{M_T} |h_k^{(j)}|^2}} b[k] + n^{(a)}[k]$$
(2.24)

is the statistic resulting from transmission from antenna *a* and  $n^{(a)}[k] \sim \mathcal{N}_c\left(0, \frac{N_0}{2M_T}\right)$  represents the corresponding noise part. Note that we can write

$$n[k] = \sum_{a=1}^{M_T} n^{(a)}[k].$$
(2.25)

Now consider that there are multiple users in the system and we are employing precoding. Our goal is to use precoding to maximize the achievable multipath diversity for each user, besides eliminating MUI. If the decision statistics in  $\boldsymbol{y}$  would each correspond to single user MRW, i.e., if y[k] = y'[k]was true for all  $k = 1, \dots, K$ , the multipath diversity gain for each user would be maximized. Consider the fact that we can write

$$y[k] = \sum_{a=1}^{M_T} y^{(a)}[k], \qquad (2.26)$$

where  $y^{(a)}[k]$  is the part of the decision statistic resulting from transmission using precoder  $M^{(a)}$ at transmit antenna *a*. If this decision statistic resulting from transmission due to each precoder would be similar to that given in (2.24), i.e.,

$$y^{(a)}[k] = y^{(a)'}[k]$$
(2.27)

then the diversity gain available to each user would be maximized. We can write

$$\boldsymbol{y}^{(a)} = \begin{bmatrix} y^{(a)}[1] & y^{(a)}[2] & \cdots & y^{(a)}[K] \end{bmatrix}^T$$
(2.28)

and

$$\boldsymbol{y}^{(a)'} = \left[ \begin{array}{ccc} y^{(a)'}[1] & y^{(a)'}[2] & \cdots & y^{(a)'}[K] \end{array} \right]^T.$$
(2.29)

We can choose our precoders such that  $y^{(a)}[k] = y^{(a)'}[k]$  would be true for some of the users. However, we want to maximize the diversity gain available to all the users, and hence we choose precoders accordingly. Thus, we choose precoder  $M^{(a)}$  so as to minimize the error between  $y^{(a)'}$  and  $y^{(a)}$ . To mathematically formulate this design criterion, we will form mean square error (MSE) [35] cost functions based upon the error between the desired single user MRW statistics given in  $y^{(a)'}$ and the actual resulting statistics  $y^{(a)}$ . We will choose the precoders such that this mean square error is minimized. Thus, we can say that we are deriving minimum mean square error (MMSE) precoders. We will often refer to the cost functions formed as MMSE cost functions to signify this fact. In the following section we will see the design of precoders when orthogonal spreading sequences are employed. We will say that optimal precoders achieve the same performance as single user MRW.

## 2.4 Precoding for Orthogonal Spreading

If the spreading codes are orthogonal, the cross-correlation matrix  $\mathbf{R} = \mathbf{S}^H \mathbf{S} = \mathbf{I}$ . Hence, we will effectively have K parallel non-interfering channels communicating with K users. Intuitively, we could simply use MRW by weighing each user's symbols with the weights corresponding to the paths from the transmitter to the particular user. We will see that this is true in the following.

The signal transmitted from antenna a will pass through the  $k^{th}$  user's channel  $h_k^{(a)}$ , and sum with the signals arriving over other paths at receiver k. Neglecting the noise components, the net

Precoding for Flat Fading Channels

received signal at receiver k will be

$$\boldsymbol{r}_{k} = \sum_{a=1}^{M_{T}} h_{k}^{(a)} \boldsymbol{x}^{(a)}.$$
(2.30)

The receiver will match filter this signal with respect to the spreading sequence  $s_k$ , to generate the decision statistic  $\tilde{y}[k]$ , given by

$$\tilde{y}[k] = \boldsymbol{s}_{k}^{H} \boldsymbol{r}_{k}$$

$$= \alpha \sum_{a=1}^{M_{T}} h_{k}^{(a)} \boldsymbol{s}_{k}^{H} \boldsymbol{S} \boldsymbol{M}^{(a)} \boldsymbol{b}.$$
(2.31)

Adding the noise term  $n[k] \sim \mathcal{N}_c(0, \frac{N_0}{2})$  generated at the output of the match filter, we write the resulting decision statistic y[k] as

$$y[k] = \alpha \sum_{a=1}^{M_T} h_k^{(a)} s_k^H S M^{(a)} b + n[k].$$
(2.32)

Since the spreading codes are orthogonal,

$$\boldsymbol{s}_k^H \boldsymbol{S} = \boldsymbol{e}_k^H. \tag{2.33}$$

Consequently, the decision statistic generated will be

$$y[k] = \alpha \sum_{a=1}^{M_T} h_k^{(a)} [\boldsymbol{M}^{(a)}]_{:,k} \boldsymbol{b} + n[k], \qquad (2.34)$$

Choosing a diagonal precoding matrix would result in no interference and the decision statistic will be

$$y[k] = \alpha \sum_{a=1}^{M_T} h_k^{(a)} [\boldsymbol{M}^{(a)}]_{k,k} b[k] + n[k], \qquad (2.35)$$

Clearly, choosing the optimal weights of MRW as the diagonal elements of the precoding matrices will give the desired performance, provided  $\alpha$  is independent of the channel coefficients. That is, we choose,

$$[\mathbf{M}^{(a)}]_{k,k} = \frac{h_k^{(a)*}}{\sqrt{\sum_{i=1}^{M_T} |h_k^{(i)}|^2}}.$$
(2.36)

The complete precoding matrix will be

$$\boldsymbol{M}^{(a)} = \operatorname{diag}\left(\frac{h_1^{(a)*}}{\sqrt{\sum_{i=1}^{M_T} |h_1^{(i)}|^2}}, \frac{h_2^{(a)*}}{\sqrt{\sum_{i=1}^{M_T} |h_2^{(i)}|^2}}, \cdots, \frac{h_K^{(a)*}}{\sqrt{\sum_{i=1}^{M_T} |h_K^{(i)}|^2}}\right), \qquad a = 1, \cdots, M_T.$$
(2.37)

60

The transmit power scaling factor  $\alpha$ , given by (2.20) will be

$$\alpha = \sqrt{\frac{K}{\sum_{a=1}^{M_T} \operatorname{tr}\{M^{(a)H} I_K M^{(a)}\}}}$$
$$= \sqrt{\frac{K}{\sum_{a=1}^{M_T} \sum_{k=1}^{K} \frac{h_k^{(a)*}}{\sqrt{\sum_{i=1}^{M_T} |h_k^{(i)}|^2}} \frac{h_k^{(a)}}{\sqrt{\sum_{i=1}^{M_T} |h_k^{(i)}|^2}}}$$
$$= 1.$$

Thus, the performance of this scheme will be the same as that of single user MRW since it results in exactly the same decision statistic as given in (2.16). This is not surprising, since, as we mentioned, orthogonal spreading results into the creation of K parallel channels, each without interference, and one for each user. In the following section, we will look at the design of precoders when the spreading codes are non-orthogonal, or specifically when,  $\mathbf{R} \neq \mathbf{I}_K$ . As we will see, substituting  $\mathbf{R} = \mathbf{I}_K$ , in the solution obtained will result into the precoder given by (2.37).

## 2.5 Precoding for Non-Orthogonal Spreading

We will now derive linear precoders when the spreading sequences are non-orthogonal. After reception, each receiver simply match filters the received signal with respect to the spreading sequence it has been assigned, in order to form a decision statistic. The system model is the same as that for precoding for orthogonal spreading. We begin with the decision statistic formed after match-filtering with respect to the sequence as given in (2.31). Stacking the decision statistics generated at each receiver, we form the vector of decision statistics  $\tilde{y}$  as

$$\tilde{\boldsymbol{y}} = \begin{bmatrix} \tilde{y}[1] & \tilde{y}[2] & \cdots & \tilde{y}[K] \end{bmatrix}^{T}$$
$$= \begin{bmatrix} \sum_{a=1}^{M_{T}} h_{1}^{(a)} \boldsymbol{s}_{1}^{H} \boldsymbol{S} \boldsymbol{M}^{(a)} \boldsymbol{b} \\ \sum_{a=1}^{M_{T}} h_{2}^{(a)} \boldsymbol{s}_{2}^{H} \boldsymbol{S} \boldsymbol{M}^{(a)} \boldsymbol{b} \\ \vdots \\ \sum_{a=1}^{M_{T}} h_{K}^{(a)} \boldsymbol{s}_{K}^{H} \boldsymbol{S} \boldsymbol{M}^{(a)} \boldsymbol{b} \end{bmatrix}$$
$$=\sum_{a=1}^{M_{T}} \underbrace{\begin{bmatrix} h_{1}^{(a)} & 0 & \cdots & 0 \\ 0 & h_{2}^{(a)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & h_{K}^{(a)} \end{bmatrix}}_{\boldsymbol{H}^{(a)}} \underbrace{\begin{bmatrix} \boldsymbol{s}_{1}^{H} \\ \boldsymbol{s}_{2}^{H} \\ \vdots \\ \boldsymbol{s}_{K}^{H} \end{bmatrix}}_{\boldsymbol{S}^{H}} \boldsymbol{S}\boldsymbol{M}^{(a)}\boldsymbol{b}$$

$$=\sum_{a=1}^{M_{T}} \boldsymbol{H}^{(a)} \boldsymbol{R}\boldsymbol{M}^{(a)}\boldsymbol{b}.$$
(2.39)

Note that we have neglected the noise terms and the power scaling factor above for simplicity. Introducing the Gaussian noise vector,  $\boldsymbol{n} = [n[1], n[2], \dots, n[K]]^T$ , where  $n[i] \sim \mathcal{N}_c(0, \frac{N_0}{2})$  is the noise component resulting after de-spreading at receiver i, we form the vector of decision statistics

$$\boldsymbol{y} = \sum_{a=1}^{M_T} \boldsymbol{H}^{(a)} \boldsymbol{R} \boldsymbol{M}^{(a)} \boldsymbol{b} + \boldsymbol{n}.$$
(2.40)

Note that the due to the non-orthogonality of the spreading sequences, the noise vector will not be white and will be distributed as  $\boldsymbol{n} \sim \mathcal{N}_c \left( \boldsymbol{0}, \frac{N_0}{2} \boldsymbol{R} \right)$ . However, as we will see, this does not affect the design of our precoders. Observe that the signal from a particular antenna does not pass through the paths from other antennas. This is evident from (2.40), which can be written as

$$\boldsymbol{y} = \sum_{a=1}^{M_T} \boldsymbol{y}^{(a)}, \qquad (2.41)$$

where

$$y^{(a)} = H^{(a)} R M^{(a)} b + n^{(a)}.$$
 (2.42)

is the vector of statistics resulting from transmission from antenna a, and  $\mathbf{n}^{(a)} \sim \mathcal{N}_c\left(0, \frac{N_0}{2M_T}\mathbf{R}\right)$  is the corresponding noise part such that

$$\boldsymbol{n} = \sum_{a=1}^{M_T} \boldsymbol{n}^{(a)}.$$
(2.43)

As mentioned earlier, we want the vector  $\boldsymbol{y}^{(a)}$ , to be as close as possible to the vector  $\boldsymbol{y}^{(a)'}$  given in (2.29), which can be written as

$$\boldsymbol{y}^{(a)'} = \boldsymbol{D}^{(a)}\boldsymbol{b},\tag{2.44}$$

where we define  $D^{(a)}$  as

$$\boldsymbol{D}^{(a)} \triangleq \operatorname{diag}\left(\frac{|h_1^{(a)}|^2}{\sqrt{\sum_{j=1}^{M_T} |h_1^{(j)}|^2}}, \frac{|h_2^{(a)}|^2}{\sqrt{\sum_{j=1}^{M_T} |h_2^{(j)}|^2}}, \cdots, \frac{|h_K^{(a)}|^2}{\sqrt{\sum_{j=1}^{M_T} |h_K^{(j)}|^2}}\right).$$
(2.45)

The choice of precoders should be such that the error e given as

$$||\boldsymbol{e}^{(a)}|| = ||\boldsymbol{y}^{(a)'} - \boldsymbol{y}^{(a)}||$$
(2.46)

is minimized. We approach this error minimization indirectly by minimizing the mean square error. To set up the optimization problem for obtaining  $M^{(a)}$ , we formulate the MSE cost function  $J^{(a)}$ as

$$J^{(a)} = E_{\boldsymbol{b},\boldsymbol{n}^{(a)}} \left\{ \left\| \boldsymbol{D}^{(a)}\boldsymbol{b} - \boldsymbol{H}^{(a)}\boldsymbol{R}\boldsymbol{M}^{(a)}\boldsymbol{b} - \boldsymbol{n}^{(a)} \right\|^{2} \right\}, \qquad (2.47)$$

Minimizing this cost function results into the optimal precoder  $M^{(a)}$ . The solution to the optimization problem is derived in Appendix C. The solution is given as

$$M^{(a)} = R^{-1} \left[ H^{(a)} \right]^{-1} D^{(a)}, \qquad a = 1, \cdots, M_T.$$
 (2.48)

The signal transmitted from antenna a can now be written as

$$\boldsymbol{x}^{(a)} = \alpha \boldsymbol{S} \boldsymbol{M} \boldsymbol{b} = \alpha \boldsymbol{S} \boldsymbol{R}^{-1} \left[ \boldsymbol{H}^{(a)} \right]^{-1} \boldsymbol{D}^{(a)} \boldsymbol{b}$$
(2.49)

Observe that for the case of orthogonal spreading, we have  $\mathbf{R} = \mathbf{I}_K$  and  $\alpha = 1$  and our precoder reduces to the form given in (2.37). Having obtained the optimal precoder we will now characterize its performance.

## 2.6 Performance Characterization of Precoding for Flat Fading Channels

In this section we will characterize the performance of the precoders given by (2.48). We will first characterize the performance for any general structure of the spreading sequences. Then we consider the special case where the spreading sequences have equal cross-correlations. We begin by using the decision statistic formed at a particular receiver to compute the probability of error conditioned on the channel. We average this conditional probability to obtain an analytic expression for the average BER, conditioned upon the scalar  $\alpha$ , which in turn depends on the cross-correlation matrix **R**. The decision statistics resulting from the use of the precoders in (2.48), will be

$$oldsymbol{y} = lpha \sum_{a=1}^{M_T} oldsymbol{H}^{(a)} oldsymbol{R} oldsymbol{M}^{(a)} oldsymbol{b} + oldsymbol{n}$$

Precoding for Flat Fading Channels

$$= \alpha \sum_{a=1}^{M_T} \boldsymbol{H}^{(a)} \boldsymbol{R} \boldsymbol{R}^{-1} \boldsymbol{H}^{(a)-1} \boldsymbol{D}^{(a)} \boldsymbol{b} + \boldsymbol{n}$$
  
$$= \alpha \sum_{a=1}^{M_T} \boldsymbol{D}^{(a)} \boldsymbol{b} + \boldsymbol{n}.$$
 (2.50)

We have assumed that all the channels are identically distributed and all receivers receive the same average power. Apart from the fact that different spreading sequences are used at each receiver, the detection process is identical at all the receivers. The decision statistic formed at all users will have an identical form. As is evident from (2.50), the decision statistic formed at receiver k will be

$$y[k] = \alpha b[k] \sqrt{\sum_{a=1}^{M_T} |h_k^{(a)}|^2} + n[k], \qquad (2.51)$$

where  $n[k] \in \mathcal{N}_c(0, \frac{N_0}{2})$ . This will be the decision statistic formed at all the receivers and hence, the average BER of each user will be the same as far as the noise level at each receiver is the same. Thus, it will suffice to characterize the BER of user k for a given SNR. We evaluate the BER at user k. We first compute the probability of error, conditioned upon the channel and the constant  $\alpha$ , expressed as  $p_e\left(\frac{\mathcal{E}_b}{N_0} \middle| \{h_k^{(a)}\}_a, \alpha\right)$ . For a particular SNR this probability will be

$$p_e\left(\frac{\mathcal{E}_b}{N_0}\Big|\{h_k^{(a)}\}_a,\alpha\right) = Q\left(\alpha\sqrt{\frac{2\sum_{a=1}^{M_T}|h_k^{(a)}|^2\mathcal{E}_b}{N_0}}\right).$$
(2.52)

As is evident from the above expression, to characterize the average performance we first need to evaluate the power scaling constant  $\alpha$ . For this we first find the denominator term given in (2.20). For this we use the fact that the matrices  $D^{(a)}$  and  $H^{(a)}$  are diagonal. Consider,

$$\sum_{a=1}^{M_T} \operatorname{tr} \left\{ \boldsymbol{M}^{(a)H} \boldsymbol{R} \boldsymbol{M}^{(a)} \right\} = \sum_{a=1}^{M_T} \operatorname{tr} \left\{ \boldsymbol{D}^{(a)*} [\boldsymbol{H}^{(a)*}]^{-1} \boldsymbol{R}^{-1} [\boldsymbol{H}^{(a)}]^{-1} \boldsymbol{D}^{(a)} \right\}$$
$$= \sum_{a=1}^{M_T} \sum_{k=1}^{K} [\boldsymbol{R}^{-1}]_{k,k} \frac{[\boldsymbol{D}^{(a)}]_{k,k}}{\sum_{i=1}^{M_T} |\boldsymbol{h}_k^{(i)}|^2},$$
$$\because \left[ \boldsymbol{D}^{(a)*} [\boldsymbol{H}^{(a)*}]^{-1} [\boldsymbol{H}^{(a)}]^{-1} \right]_{k,k} = \frac{|\boldsymbol{h}_k^{(a)}|^2}{\sum_{i=1}^{M_T} |\boldsymbol{h}_k^{(i)}|^2} \cdot \frac{1}{|\boldsymbol{h}_k^{(a)}|^2}$$
$$= \sum_{k=1}^{K} [\boldsymbol{R}^{-1}]_{k,k} \sum_{a=1}^{M_T} \frac{[\boldsymbol{D}^{(a)}]_{k,k}}{\sum_{i=1}^{M_T} |\boldsymbol{h}_k^{(i)}|^2}$$

$$= \sum_{k=1}^{K} [\mathbf{R}^{-1}]_{k,k},$$
  

$$\therefore \sum_{a=1}^{M_T} [\mathbf{D}^{(a)}]_{k,k} = \sum_{i=1}^{M_T} |h_k^{(i)}|^2$$

$$= \sum_{k=1}^{K} \frac{1}{\lambda_k},$$
(2.53)

where  $\lambda_k$ ,  $k = 1, \dots, K$ , are the eigenvalues of  $\mathbf{R}$  and we have used results from Appendix B, (B-6) and (B-5). Since  $\mathbf{R}$  is full rank, we are guaranteed that the eigenvalues are non-zero. Consequently, we have

$$\alpha = \sqrt{\frac{K}{\sum_{k=1}^{K} [\mathbf{R}^{-1}]_{k,k}}}$$

$$= \sqrt{\frac{K}{\sum_{k=1}^{K} \frac{1}{\lambda_k}}}.$$
(2.54)

Based on this expression we make two important observations:

1. The constant  $\alpha$  is only a function of the cross-correlation matrix  $\mathbf{R}$ , and is independent of the channel. Therefore, the probability of error conditioned upon the channel and on alpha as given in (2.52) will be the same as the probability of error conditioned upon the channel and on the matrix  $\mathbf{R}$ , i.e.,

$$p_e\left(\frac{\mathcal{E}_b}{N_0} \left| \left\{ h_k^{(a)} \right\}, \mathbf{R} \right) = p_e\left(\frac{\mathcal{E}_b}{N_0} \left| \left\{ h_k^{(a)} \right\}, \alpha \right).$$
(2.55)

As we will see a direct consequence of this is that full diversity gain is achievable, and the average probability of error will depend entirely on R.

2. The constant

$$\alpha \le 1. \tag{2.56}$$

The proof is given in Appendix B.1. As indicated there, the equality will hold only if  $\mathbf{R} = \mathbf{I}$ , i.e., if the spreading codes are orthogonal.

The average probability of error, will now be a function of  $\boldsymbol{R}$ . We express this probability as  $P_e\left(\frac{\varepsilon_b}{N_0}|\boldsymbol{R}\right)$ . This average probability, which is conditioned upon  $\boldsymbol{R}$  can be computed using (2.55)

and (2.52) as

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\Big|\boldsymbol{R}\right) = E_{\left\{h_k^{(a)}\right\}}\left\{Q\left(\alpha\sqrt{\frac{2\sum_{a=1}^{M_T}|h_k^{(a)}|^2\mathcal{E}_b}{N_0}}\right)\right\}.$$
(2.57)

We compute a bound on this probability using the bound on the Q-function given in (A-28). We substitute  $r = \sum_{a=1}^{M_T} |h_k^{(a)}|^2$ ,  $n = 2m = 2M_T$ , and  $a^2 = \frac{2\mathcal{E}_b}{N_0}\alpha^2$  in (A-28) to get the upper bound on SNR as

$$P_e\left(\frac{\mathcal{E}_b}{N_0}\Big|\boldsymbol{R}\right) \leq \frac{1}{\left(1+\alpha^2 \frac{2\mathcal{E}_b}{N_0} \sigma_h^2\right)^{M_T}}.$$
(2.58)

Using the solution given for the expectation of a Q-function with a first order Rayleigh random variable as its argument given in (A-30), we find the exact value of the expectation in (2.57) as

$$P_{e}\left(\frac{\mathcal{E}_{b}}{N_{0}}\Big|\boldsymbol{R}\right) = \frac{1}{2(M_{T}-1)!} \left[1 - \sqrt{\frac{\alpha^{2}\frac{2\mathcal{E}_{b}}{N_{0}}\sigma_{h}^{2}}{\alpha^{2}\frac{2\mathcal{E}_{b}}{N_{0}}\sigma_{h}^{2} + 1}}\right] \sum_{k=0}^{M_{T}-1} \binom{2k}{k} \left(\frac{1}{4\left(\alpha^{2}\frac{2\mathcal{E}_{b}}{N_{0}}\sigma_{h}^{2} + 1\right)}\right)^{k}, \quad (2.59)$$

where we have used the same values of r, n and a as we used for computing the bound in (2.58). Note that (2.58) will be an upper bound for (2.59) as is indicated in (A-35). Since our goal is to check diversity, we will simply work with the upper bound specified by (2.58), importantly observing that  $\alpha$  is independent of the ratio,  $\frac{\mathcal{E}_b}{N_0}$ . From the bound in (2.58), noting the fact that  $\alpha$ will be constant of a given set of spreading codes, and that  $\sigma_h^2$  is also constant, we can say that at high SNRs

$$P_e\left(\mathrm{SNR} \middle| \mathbf{R}\right) \propto \mathrm{SNR}^{-M_T},$$
 (2.60)

and thus from (1.63) we can say that full diversity gain is achieved. We observe that the direct consequence of  $\alpha$  is a reduction in the effective SNR. This arises from the fact that  $\alpha \leq 1$ . Note that this will not reduce the slope of the BER curve, but only shifts it to the right. We call this right shift a loss in SNR. The higher the value of  $\alpha$ , the less will be the loss in SNR. The value of  $\alpha$  varies from 0 to 1, with  $\alpha = 1$  corresponding to the case where the spreading sequences are orthogonal as we saw in (2.38). Figure 2.1 shows the bound given in (2.58) for different numbers of antennas, for  $\alpha = 0.4$  and  $\alpha = 0.1$ . For lower values of  $\alpha \approx 0.1$ , the loss is large.

To gain more insight we now impose some structure on the spreading cross-correlation matrix  $\boldsymbol{R}$ , and hence on the spreading matrix  $\boldsymbol{S}$ , beyond the fact that it is symmetric (by definition). We

assume that all the non-diagonal entries are equal, meaning that all the spreading codes have the same cross-correlation with all other codes, i.e.

$$\langle \boldsymbol{s}_{k}, \boldsymbol{s}_{l} \rangle = \begin{cases} \rho & k \neq l \\ 1 & k = l, \end{cases}$$

$$(2.61)$$

where the second equality stems from the normalization of the spreading sequences. Thus, we have

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho & \rho & \cdots & \rho & 1 \end{bmatrix}.$$
(2.62)

As shown in Appendix C.1, we will have

$$\sum_{k=1}^{K} [\mathbf{R}^{-1}]_{k,k} = K \left( \frac{1}{1-\rho} - \frac{\rho}{1+(K-2)\rho - (K-1)\rho^2} \right).$$
(2.63)

Thus from (2.54), we get

$$\alpha = \left(\frac{1}{1-\rho} - \frac{\rho}{1+(K-2)\rho - (K-1)\rho^2}\right)^{-1/2}.$$
(2.64)

A frequent measure for the performance of a multi-user communication system is obtained by considering the performance in the presence of infinite users. We define

$$\alpha_{\infty} \triangleq \lim_{K \to \infty} \alpha, \tag{2.65}$$

evaluating which for our special case gives

$$\alpha_{\infty} = \left(\frac{1}{1-\rho}\right)^{-1/2}.$$
(2.66)

Note that  $\rho$  is a function of K as well as the spreading factor N. Keeping  $\rho$  constant while increasing the number of users would require increasing the length of spreading codes, thereby reducing the chip interval and increasing the bandwidth. However, besides characterizing the performance with infinite users, the value,  $\alpha_{\infty}$  also gives an indication of the worst case performance since it is the lowest possible value taken by  $\alpha$ .



Figure 2.1: Figure illustrates that full diversity, as indicated by the slope of the BER curve, is achieved using precoding with perfect CSI at transmitter for flat fading channels. There will be a loss in SNR resulting from  $\alpha \neq 1$ , which corresponds to the case of non-orthogonal spreading. We have used  $\alpha = 0.4$  and  $\alpha = 0.1$  and  $M_T = 1, 2, 4$  antennas.

For a given value of  $\rho$ , Figure (2.2) illustrates the performance of precoding for varying number of antennas, users and different values of equal cross-correlations. The performance achieved with 4 or 16 users in the system with using 2 or 4 transmit antennas is considered. Non-orthogonal spreading with cross-correlation  $\rho = 0.4$ ,  $\rho = 0.9$  as well as orthogonal spreading has been considered. From the figure observe that the loss in SNR for high values of cross-correlations is relatively high while for medium values it is relatively small. For cross-correlation values like  $\rho = 0.2$ , the loss is relatively insignificant. Figure 2.3 illustrates the variation in  $\alpha$  with the number of users and  $\rho$ , as given in (2.64) and (2.66). The variation has been shown for 2, 4, 8, 16, 32 and an infinite number of users. Observe that  $\alpha$  tends to  $\alpha_{\infty}$  very quickly with K. We observed in Figure 2.2 that for high values of cross-correlation, there is significant loss in the SNR while for medium values there is a relatively negligible loss. Previously, in Figure 2.1, we observed that the loss is negligible for medium values of the power scaling coefficient, while it is relatively high for small values. Figure 2.3 shows that  $\alpha$  has a steep fall as cross-correlations vary from 0.7 to 1. Observe from the bound in (2.58) and the expression in (2.59), the loss in SNR will be  $10 \log(\alpha^2)$  dB. Figure 2.4 shows the variation in  $10 \log(\alpha^2)$  with  $\rho$  for 2, 4, 8, 16, 32 and an infinite number of users. Figure (2.4) clearly shows that as the cross-correlations values go from  $\rho = 0.5$  to  $\rho \approx 1$ , the SNR loss goes from around 3 dB to around 20 dB, while the loss for cross-correlations from  $\rho \approx 0$  to  $\rho \approx 0.5$  lies in the range from 0dB to 3dB. The rise in loss is extremely steep for the values in the interval  $\rho \approx 0.75$  to  $\rho \approx 1$ .



Figure 2.2: Figure illustrates the variation in diversity gain with  $\rho$  when perfect CSI is available at the transmitter. There will be a loss in SNR resulting from  $\rho \neq 0$ , which corresponds to the case of orthogonal spreading. We have used  $\rho = 0.0, 0.4, 0.9$  and  $M_T = 2, 4$  transmit antennas for a system with K = 4, 16 users.



Figure 2.3: Figure illustrates the variation of  $\alpha$  with  $\rho$  for  $K = 2, 4, 8, 16, 32, \infty$  users.



Figure 2.4: Figure illustrates the variation in the loss in SNR  $=10 \log_{10} \alpha^2$  with  $\rho$  for  $K = 2, 4, 8, 16, 32, \infty$  users.

### 2.7 Conclusion

In this chapter we derived and analyzed the performance of precoders for flat fading channels. The precoder structure consists essentially of a decorrelator  $(\mathbf{R}^{-1})$  applied at the transmitter before spreading, and a MIMO filter with single user MRW weights. Full diversity is achievable as seen by the upper bound given in (2.58) and indicated in (2.60). Interference is perfectly cancelled, but this comes at the cost of a loss in the SNR performance, by a factor proportional to the eigenvalues of the spreading sequence cross-correlation matrix  $\mathbf{R}$ . For the special case when the spreading codes have equal cross-correlation, we observe that the loss is relatively high when the cross-correlations are greater than 0.7, while for lesser values of cross-correlation the loss is not that significant. During the course of this research, some other research considering a similar system model but different precoding techniques appeared in [87, 88]. These were based on lattice reduction strategies, techniques proposed for MIMO communications in [89] and used DFE and Tomlinson-Harashima type non-linear precoders to achieve full diversity. Our techniques are linear and relatively simple and achieve full diversity. Note that in this chapter we have assumed perfect channel knowledge, which is an exception rather than the rule. In Chapter 3 we design precoders assuming partial channel knowledge available at the transmitter.

# Precoding with Partial Channel Information

In the previous chapter we have designed precoders based upon availability of perfect channel state information (CSI) at the transmitter. In this chapter, we will derive precoders under the assumption that the available CSI is imperfect, or rather partial. As we will see, there are various ways of modelling partial CSI. We will assume that the available information is statistically correlated to the actual channel. The resulting precoding solutions are general, in the sense that they are completely independent of the structure of the statistical correlation between the available information and the actual channel. The precoders thus obtained, can be computed for any given correlation structure between the actual channel and the available partial CSI. As an example we will assume that the available channel estimates are perfect, but old and belong to the past channel state rather than the channel state during which communication takes place. In Chapter 1, we observed that a wireless channel can be modelled as a complex Gaussian process. We exploit the temporal correlation between the channel coefficients, samples of this process taken at distinct time instants, to compute the precoders. We will simulate the performance of the precoders thus obtained, and compare it with the performance of precoding when perfect CSI is available at the transmitter. Again, no CSI is assumed available to the decentralized receivers. Part of the work presented here is previously published in [3, 108].

### 3.1 Introduction

Modelling partial CSI is a problem which can be approached in various ways. In [90] it was assumed that for a single-user system with  $M_T$  transmit antennas and  $M_R$  receive antennas, the first few eigenvectors of the channel cross-correlation matrix  $\mathbf{H}^H \mathbf{H} \in \mathbb{C}^{M_R \times M_R}$  are known to both the transmitter and the receiver. The channel matrix  $\mathbf{H} \in \mathbb{C}^{M_R \times M_T}$  is such that,  $[\mathbf{H}]_{i,j} = h_{ij}$ , where  $h_{ij}$  represents the channel from transmit antenna j to receive antenna i. An informationtheoretic study of space-time precoding for a single user MIMO channel, assuming partial CSI availability at the transmitter and perfect CSI at the receiver was conducted in [91]. There, the authors assumed the presence of channel feedback from the receiver to the transmitter in the form of first order (mean), and second order (variance) channel statistics. For channel models like the Rician or Nakagami-m fading models, knowledge of additional parameters may be considered as partial CSI. Another approach to modelling partial CSI is to assume knowledge of receiver signal to noise ratio (SNR) [92]. Interestingly, all this research was focussed on single user MIMO techniques along with the assumption that the receiver also has channel knowledge.

Channel estimation techniques are employed for estimating the channel at the transmitter and/or receiver. Channel estimation techniques are primarily classified as either training based [93, 94] or blind [95, 96]. In training based channel estimation, the fundamental idea is to provide the channel estimator with the channel response to a symbol sequence known *a priori* to the estimator. The estimator then uses the provided information to estimate the channel rather than the symbols in the sequence. Blind techniques exploit orthogonality between the noise and signal sub-spaces to estimate the channel [69]. Our system model assumes channel estimation at the transmitter. In a Time-Division-Duplex (TDD) system, the mobile receivers can either transmit training sequences over the uplink, or blind techniques can be used. Blind channel estimation techniques used for designing adaptive precoders were studied in [3]. Since the estimation is inherently performed in the presence of noise, the estimates will never be perfect. However, the actual channel will be statistically correlated to this estimated value provided the SNR is sufficiently high. Even if the channel estimate is perfect, the channel is time-varying and due to mobility of either the receiver or any other objects in the environment surrounding the receiver, the actual channel response during transmission will definitely be different from that which was estimated. One way of modelling channel.

partial CSI is to assume that the known channel is perfect but corresponds to the past channel state. This approach was followed in [97], where the focus was on single user MIMO channels. The authors there derived an optimal transformation to be performed on a pre-determined space-time code, such that the resulting code maximizes the diversity and capacity utilization of the MIMO

The approaches mentioned above are all for single user MIMO systems and assumed a special form of partial CSI availability at the transmitter and partial or perfect CSI at the receiver. Our focus however, is on deriving precoders for a broadcast MISO channel where CSI is available to the transmitter. The partial CSI models considered in [90, 91, 92, 97] are special cases of the general scenario that the known channel information is statistically correlated to the actual channel state during which communication takes place. In this research, instead of considering some special form of partial CSI, we will obtain a precoding solution under the general assumption that the CSI available is statistically correlated to the actual channel and not impose any specific structure on the correlation. The design criteria are conditioned upon the available partial CSI instead of the actual CSI. We will use conditional channel correlation matrices, i.e., correlation matrices that are functions of the actual channel conditioned upon the available CSI, to derive the general solution for the precoders. In Section 3.2, we will derive the precoders for the general case. Similar to the approaches for single user MIMO systems, we will target diversity maximization, but we will lay equal focus on interference cancellation. Special cases like those mentioned above can be computed from this general solution by imposing the correlation between the partial CSI and the actual channel on these conditional channel correlation matrices. In Section 3.3, we will consider the special case considered in [97], that the available CSI is perfect but corresponds to old channel estimates, and compute the precoders for this special case. We will assume that the channel closely follows the Jake's model [2]. In Section 3.4, we will see the achievable BER performance of precoding when partial CSI is available as per the special case considered.

### 3.2 Precoders with Partial Channel Information

The system model will be identical to that in Chapter 2. The transmitted signal is the same as given by (2.18). The signal received, in the absence of noise, at any particular receiver will be

identical to that given in (2.30). The vector of decision statistics  $\boldsymbol{y}$ , will be as given in (2.40). The key difference stems from the formulation of the cost functions used to obtain the precoders. In Chapter 2, we assumed that the channel was perfectly known. In deriving the precoders we specifically computed the expectations in the cost function (2.47) assuming that the channel matrix  $\boldsymbol{H}^{(a)}$  was perfectly known. With partial CSI we cannot assume  $\boldsymbol{H}^{(a)}$  to be a constant when evaluation the expected mean square error given by (2.47). Here we will need to compute an additional expectation over the actual channel conditioned upon the partial CSI. To model partial CSI, we define the matrix  $\hat{\boldsymbol{H}}^{(a)}$  as a known quantity, that is statistically correlated to the actual channel matrix  $\boldsymbol{H}^{(a)}$ , for all  $a = 1, \dots, M_T$ , given in equation (2.39). In (2.47), we formed the MMSE cost functions,  $J^{(a)}, a = 1, \dots, M_T$ , by taking expectations over noise and the transmitted data symbols. With partial CSI, as mentioned above, the key step is to take the expectation, given in (2.47), over the actual channels  $\{\boldsymbol{H}^{(a)}\}$ , conditioned upon the available information,  $\{\hat{\boldsymbol{H}}^{(a)}\}$ . The expectation over the bits and the noise needs to be taken as usual. The resulting cost function is

$$J^{(a)} = E_{\boldsymbol{b},\boldsymbol{n}^{(a)},\left\{\boldsymbol{H}^{(i)}\right\}} \left\{ \left\| \boldsymbol{D}^{(a)}\boldsymbol{b} - \boldsymbol{H}^{(a)}\boldsymbol{R}\boldsymbol{M}^{(a)}\boldsymbol{b} + \boldsymbol{n}^{(a)} \right\|^{2} \left\| \left\{ \hat{\boldsymbol{H}}^{(i)} \right\} \right\}.$$
(3.1)

Using elements from vector calculus, the precoders are derived in [100]. An alternate form of the derivation, resulting in the same set of precoders is given in Appendix D. The resulting precoders are functions of the following conditional channel correlation matrices:

$$\boldsymbol{C}_{HH}^{(a)} \triangleq E_{\boldsymbol{H}^{(a)}} \left\{ \boldsymbol{H}^{(a)H} \boldsymbol{H}^{(a)} \middle| \hat{\boldsymbol{H}}^{(a)} \right\}$$
(3.2)

$$\boldsymbol{C}_{HD}^{(a)} \triangleq E_{\left\{\boldsymbol{H}^{(i)}\right\}} \left\{\boldsymbol{H}^{(a)H}\boldsymbol{D}^{(a)} \middle| \left\{\hat{\boldsymbol{H}}^{(i)}\right\}\right\}.$$
(3.3)

As shown in Appendix D, the precoding matrix can be expressed as

$$\boldsymbol{M}^{(a)} = \boldsymbol{R}^{-1} \left[ \boldsymbol{C}_{HH}^{(a)} \right]^{-1} \boldsymbol{C}_{HD}^{(a)}, \qquad a = 1, \cdots, M_T.$$
(3.4)

Observe that these precoders have a form similar to the precoder for flat fading channels with perfect CSI given in (2.48). While deriving these precoders, we have not imposed any structure on the partial CSI modelled using the matrix  $\hat{\boldsymbol{H}}^{(a)}$ . The following section illustrates, the computation of these precoding matrices when we impose a special structure on the partial CSI.

# 3.3 Special Case: Partial Channel Information as Old Channel Estimates

In this section we will consider the special case that the available channel information is perfect, but corresponds to the old channel. Let the complex time-variant channel between user k's receiving antenna and transmit antenna a be  $h_k^{(a)}(t)$ . Let t be the time instant during which communication will take place and  $t - \Delta t$ , be the instant for which perfect channel estimates are available. Then, we denote

$$h_k^{(a)} = h_k^{(a)}(t) (3.5)$$

$$\hat{h}_{k}^{(a)} = h_{k}^{(a)}(t - \Delta t).$$
 (3.6)

Note that  $h_k^{(a)}$  corresponds to the  $k^{th}$  diagonal element of the diagonal channel matrix  $\boldsymbol{H}^{(a)}$ . The partial CSI matrix can be represented as a diagonal matrix given as

$$\hat{\boldsymbol{H}}^{(a)} = \begin{bmatrix} \hat{h}_{1}^{(a)} & 0 & \cdots & 0 \\ 0 & \hat{h}_{2}^{(a)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \hat{h}_{K}^{(a)} \end{bmatrix}.$$
(3.7)

Each channel corresponds to a complex Gaussian process. Hence, each real dimension of the actual channel  $h_k^{(a)}$ , over which communication takes place, has a conditional Gaussian distribution, conditioned upon the corresponding real dimensions of the old channel, as given in Appendix A.2, and given below in (A-10).

We characterize the channel as Rayleigh fading as per the Jake's model [2]. This model is suitable for, and is widely used for, modelling mobile wireless channels. In Chapter 1, we defined the channel autocorrelation function in (1.44). In an *isotropic scattering* environment [101], where the rays arrive from all directions, the angle of arrival has a uniform distribution. According to the Jake's Model, with isotropic scattering the autocorrelation function of the time-variant channel transfer function is given as

$$\phi_h(\Delta t) \triangleq \frac{1}{2\sigma_h^2} E\{h(t)h^*(t - \Delta t)\}$$
  
=  $J_0(2\pi f_d \Delta t),$  (3.8)

where

 $J_0(\cdot)$  : zero-th order Bessel function of first kind

 $f_d$  : maximum Doppler frequency

As indicated by (3.4), to compute the precoders we need to evaluate

$$\begin{bmatrix} \boldsymbol{C}_{HH}^{(a)} \end{bmatrix}_{k,k} = \begin{bmatrix} E \left\{ \boldsymbol{H}^{(a)H} \boldsymbol{H}^{(a)} \middle| \hat{\boldsymbol{H}}^{(a)} \right\} \end{bmatrix}_{k,k}$$
  
=  $E \left\{ |h_k^{(a)}|^2 \middle| \hat{h}_k^{(a)} \right\},$  (3.9)

and

$$\begin{bmatrix} \boldsymbol{C}_{HD}^{(a)} \end{bmatrix}_{k,k} = \begin{bmatrix} E \left\{ \boldsymbol{H}^{(a)H} \boldsymbol{D}^{(a)} \middle| \{ \hat{\boldsymbol{H}}^{(j)} \} \right\} \end{bmatrix}_{k,k} \\ = E \left\{ h_k^{(a)*} \frac{|h_k^{(a)}|^2}{\sqrt{\sum_{i=1}^a |h_k^{(i)}|^2}} \middle| \{ \hat{h}_k^{(i)} \} \right\}.$$
(3.10)

For simplification we use the following notation

$$\hat{h}_{k}^{(a)} = \hat{h}_{r} + j\hat{h}_{i}$$
 (3.11)

$$h_k^{(a)} = h_r + jh_i,$$
 (3.12)

where we have dropped the user index k, and the antenna index a for convenience. Here,  $\hat{h}_r, h_r$ and  $\hat{h}_i, h_i$  each represent samples of two independent, stationary Gaussian random processes with mean 0 and variance  $\sigma_h^2$ . The conditional distribution of  $h_r$  given  $\hat{h}_r$ , and  $h_i$  given  $\hat{h}_i$ , is as given in Appendix A.2

$$p(h|\hat{h}) = (2\pi\sigma_{h|\hat{h}}^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_{h|\hat{h}}^2}(h-\mu_{h|\hat{h}})^2\right\},\tag{3.13}$$

where we have dropped the sub-scripts r and i for convenience and

$$\mu_{h|\hat{h}} = \rho_{h|\hat{h}}\hat{h} \tag{3.14}$$

$$\sigma_{h|\hat{h}}^2 = \sigma_h^2 (1 - \rho_{h|\hat{h}}^2)$$
(3.15)

$$\rho_{h|\hat{h}} = J_0(2\pi f_d \triangle t). \tag{3.16}$$

From (3.14) and (3.15) and using  $\rho_{h|\hat{h}} = \frac{\phi_c(\triangle t)}{\sigma_h^2}$ , we arrive at

$$E\left\{|h_{k}^{(a)}|^{2}\left|\hat{h}_{k}^{(a)}\right\} = J_{0}^{2}(2\pi f_{d} \triangle t)(\hat{h}_{r}^{2} + \hat{h}_{i}^{2}) + \sigma_{h}^{2}(1 - J_{0}^{2}(2\pi + f_{d} \triangle t)) + \sigma_{h}^{2}(1 - J_{0}^{2}(2\pi + f_{d} \triangle t)) \\ = J_{0}^{2}(2\pi f_{d} \triangle t)|\hat{h}_{k}^{(a)}|^{2} + 2\sigma_{h}^{2}(1 - J_{0}^{2}(2\pi f_{d} \triangle t)).$$

$$(3.17)$$

We can evaluate (3.10) empirically via a Monte-Carlo simulation for every channel realization. The following section shows the simulation results obtained from using the resulting precoders. But first we will make an important observation.

Consider the precoders given by (3.4). If the channel information is very good, i.e., highly correlated with the actual channel, then the precoder will tend to that given in (2.48). This is clear, since with high correlation  $\rho_{h|\hat{h}} \rightarrow 1$ , and the conditional variance, given by (3.15), will tend to 0. As a result,

$$\boldsymbol{C}_{HH}^{(a)} = E_{\boldsymbol{H}^{(a)}} \{ \boldsymbol{H}^{(a)H} \boldsymbol{H}^{(a)} \big| \hat{\boldsymbol{H}}^{(a)} \} \stackrel{\rho_{h|\hat{h}} \to 1}{\longleftrightarrow} E_{\boldsymbol{H}^{(a)}} \{ \boldsymbol{H}^{(a)H} \boldsymbol{H}^{(a)} \big| \boldsymbol{H}^{(a)} \} = \boldsymbol{H}^{(a)H} \boldsymbol{H}^{(a)}, \quad (3.18)$$

and similarly

$$\boldsymbol{C}_{HD}^{(a)} \stackrel{\rho_{h|\hat{h}} \to 1}{\longleftrightarrow} \boldsymbol{H}^{(a)H} \boldsymbol{D}^{(a)}.$$
(3.19)

Substituting these limits in (3.4), we get the form given in (2.48). Next, we will simulate the BER performance of these precoders and observe the loss as compare to perfect CSI, that results from the mismatch in the channel estimate used. We will observe that interference is completely eliminated. This is evident from the perfect presence of the pre-decorrelator structure in the precoder given in (3.4). However, full diversity is not achieved unless good partial channel information is available.

# 3.4 Performance of Precoding with Partial CSI as Old Channel Estimates

An analysis similar to that done for precoding with perfect channel information would require:

• Computation of the matrix

$$\boldsymbol{H}^{(a)} \left[ \boldsymbol{C}_{HH}^{(a)} \right]^{-1} \boldsymbol{C}_{HD}^{(a)} \tag{3.20}$$

• Computation of the power scaling factor  $\alpha$ .

For analyzing the special case, we first need to compute (3.10). Then we need to average the resulting BER conditioned upon the channel over the actual channel variations. This would be extremely complicated. Even if we find an expression for (3.10), the complexity of averaging the BER, conditioned upon the channel, would be very high. Since the analysis is extremely complicated, we resort to simulation in order to characterize the system performance. We will simulate the BER performance using precoders derived above for the system where old and perfect channel estimates are available to the transmitter.

An important aspect we need to consider when characterizing the performance of precoding with partial CSI is performance for a given quality of partial channel information. The correlation between the available old channel estimates, and the actual communication channel indicates the quality of partial channel information. For our correlation model (3.8), this correlation will depend not only on the delay  $\Delta t$ , but also on the Doppler frequency  $f_d$ . Figure 3.1 illustrates the autocorrelation function given by (3.8) for different Doppler frequencies and different delays.



Figure 3.1: Figure illustrates the correlation between the past and perfectly known channel and the actual channel as a function of different Doppler frequencies  $f_d = 20, 50, 80$  Hz.

Instead of considering different Doppler frequencies and different delays, we consider a single parameter  $f_d \Delta t$ , also referred to as the normalized time delay [21] or the normalized Doppler frequency. This product is a suitable representative for the quality of partial channel information. Figure 3.2 shows a plot of the autocorrelation function vs this  $f_d \Delta t$ . Observe from the figure that for values  $f_d \Delta t \ll 0.05$ , the correlation is very high and it falls sharply from that point onwards. We will say that the channel is fast fading if the normalized Doppler frequency,  $f_d \Delta t < 0.1$  and slow fading if  $f_d \Delta t < 0.001$ . We will now see the simulation setup and the process of generating a correlated Gaussian vector for our channel simulation.



Figure 3.2: Figure illustrates the correlation between the past and perfectly known channel and the actual channel as a function of the normalized Doppler frequency  $f_d \Delta t$ .

#### 3.4.1 Simulation Setup

The assumption of partial channel knowledge leads to problems arising from imperfect phase knowledge. We have assumed perfect knowledge of past coefficients implying perfect phase information of the elements of  $\hat{H}^{(a)}$ . This information though, cannot be used to perfectly estimate the phase of the elements of  $H^{(a)}$ . The channel introduces phase distortion and the transmitted signal may undergo a 180° phase change when passing through the channel. In the absence of perfect phase knowledge, the receiver cannot differentiate between a +1 and a -1, which are 180° apart, since it cannot say if there has been any phase distortion. Any phase distortion will result in errors even in the absence of background noise, and hence cause an error floor in the BER curve. With perfect phase knowledge at the transmitter, these phase changes are compensated at the transmitter itself. This is achieved by the precoders used when perfect CSI is available to the transmitter. However, with partial CSI at the transmitter, perfect phase information is not available and the transmission would result into a phase mismatch. Hence, we resort to the use of differentially encoded phase shift keying (DEPSK) which does not require phase information at the receiver or transmitter. We will use differential encoding, such that the transmitted signal changes its phase received signal to make a decision between a 1 and a -1. Unlike BPSK, absolute phase information is not needed. In our DPSK receiver, the detector uses the product of the signals received during say interval  $\tau$  and  $\tau + 1$ , to estimate the signal transmitted during interval  $\tau + 1$ . If the received signals are  $x[\tau]$  and  $x[\tau + 1]$ , then the estimate for the bit transmitted during the interval  $\tau + 1$  is generated as

$$\hat{b}[\tau+1] = \operatorname{sign} \Re \{ -x[\tau+1]x^*[\tau] \},$$
(3.21)

where we have assumed that a phase change in the transmitted signal corresponds to the transmission of a +1. We can view a DPSK receiver as deriving phase knowledge from the received signals. An important requirement for our simulation is the generation of a correlated fading channel which is described next.

#### Generation of Correlated Fading

Wideband measurements made on a real channel give the channel impulse response. A large number of such measurements could be used to collect samples of the channel impulse response which could then be used for determining the performance of precoders. This not being the case, we resort to Monte-Carlo based techniques for generating a fading channel for our simulations. Besides, these techniques also enable performance characterization for different types of channel models with different parameters. We will restrict ourselves to the Jake's model. The generation of a correlated fading channel has received lot of attention and focus has primarily been on reducing the number of computations required. Recent research, has focused on generating the Nakagami-m envelopes [103], and on generating correlated fading envelopes [104, 105]. We assume that the users are distant enough to have uncorrelated Rayleigh fading channels. We know that the different user channels correspond to independent complex Gaussian random processes with identical statistics. Each of the 2 real processes corresponding to a particular channel is a zero-mean, stationary random processes with identical statistics. Since we are assuming the Jake's correlation model for scattering channels, we need to generate two identical but, independent real colored Gaussian random processes with the autocorrelation given by

$$\phi_x(\Delta t) = \sigma_h^2 J_0(2\pi f_d \Delta t), \qquad (3.22)$$

and combine them in quadrature to simulate particular user's channel. We use c(t) to denote a correlated Gaussian random process. We want to generate a channel vector c, corresponding to samples of this correlated Gaussian random process. It will have a cross-covariance matrix  $\boldsymbol{G} \in \mathbb{R}^{P \times P}$ , such that

$$[\boldsymbol{G}]_{i,j} = \frac{1}{\sigma_h^2} J_0(2\pi f_d \triangle t(i-j))$$
(3.23)

One technique for generating a Rayleigh fading envelope by is given in [12]. The technique uses the Fourier transform relationship between the Doppler power spectrum and the autocorrelation function as given in (1.48). Outputs of two independent Gaussian noise sources, one for the inphase component and the other for the quadrature component, are multiplied with samples of the Doppler spectrum. After taking the inverse fast fourier transform (IFFT), these in-phase and quadrature components are combined to simulate a Rayleigh fading channel. This however, involves taking the IFFT and is a computationally intensive frequency-domain technique. A common timedomain technique for generating a correlated Gaussian vector is to generate a Gaussian vector with i.i.d. entries and then pass it through a low-pass filter [21]. Two such filters are used to generate the in-phase and quadrature components. The order of the filter in such applications is critical. For example a simple Markov process [106], where a particular channel realization will depend only on the previous channel realization, can be obtained by using a filter of order 1. Using the filter approach requires the implementation of a high order filter requiring larger run times during simulations [21]. Since we are in essence generating colored noise vector from an independent one, we could use the eigenvalue decomposition of the Hermitian and positive definite covariance matrix  $G = U\Lambda U^H$ , [18, 35], to generate the required channel from an uncorrelated Gaussian vector. If q, is a Gaussian random vector with zero-mean, i.i.d, unit variance entries, i.e.,  $\boldsymbol{q} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{P})$ , then

$$\boldsymbol{c} = \boldsymbol{U} \Lambda^{1/2} \boldsymbol{q} \tag{3.24}$$

will have the covariance matrix

$$E\{\boldsymbol{c}\boldsymbol{c}^{H}\} = \boldsymbol{U}\Lambda^{1/2}E\{\boldsymbol{q}\boldsymbol{q}^{H}\}\Lambda^{1/2}\boldsymbol{U}^{H} = \boldsymbol{G}.$$
(3.25)

The eigenvectors of G are the columns of the matrix U, with corresponding eigenvalues given as the elements of the diagonal matrix  $\Lambda$ . It turns out that most of the eigenvalues are close to zero, when the matrix G is large in size [19]. Hence, to reduce the computations, we can use only eigenvalues (and corresponding eigenvectors) that are greater than some value, say  $\epsilon$ . We have used this technique in our simulations to generate the in-phase and quadrature components for the channel of each of the K users. In our channel generation process, we assumed  $\epsilon = 10^{-5}$ .

Table 3.1 gives the details of the system we have simulated. We assume the spreading sequences to be such that the cross-correlations between any two users is the same and given by  $\rho$ . Besides these, we assume that the channel is constant for a few symbol durations. Specifically, a block size of 16 is chosen for the simulations.

Parameter	Symbol	Value
No. of Users	K	10
No. of transmit antennas	$M_T$	2
Spreading Codes	Codes with equal cross-correlations	0.4
	ho	0.7
Normalized Doppler Frequency		0.1
	$f_d  riangle t$	0.05
		0.001

Table 3.1: System model for simulation

#### 3.4.2 Results

We simulated precoding for the system as given in Table 3.1 using channels generated as described above. Figures 3.3 and 3.4 show the BER performance when the equal cross-correlation between the spreading codes are  $\rho = 0.4$  and  $\rho = 0.7$ , respectively. The case where  $f_d \Delta t = \infty$ , corresponds to the case when the available old channel information is independent of the actual channel. It can be viewed as if no channel information is available. No diversity gain is seen but interference cancellation is achieved. No phase information is required since DPSK is employed. For a fast fading channel corresponding to  $f_d \Delta t = 0.1$ , again no diversity gain is available, though the loss in SNR is slightly lesser than that when no channel information is available. For the case where  $f_d \Delta t = 0.05$ , the available diversity gain is slightly better than that for a single antenna system. For a slow fading channel corresponding to  $f_d \Delta t = 0.001$ , the partial channel information is highly correlated to the actual channel and almost full diversity gain is available. The performance is very close to that when perfect channel information is available at the transmitter, but has a higher SNR loss.



Figure 3.3: Figure shows the simulation result for precoding when  $M_T = 2$  transmit antennas are used for K = 10 users and  $\rho = 0.4$ . Different values of  $f_d \triangle t$  are used. Performance of MRW with perfect channel information at the transmitter for a single user system with  $M_T = 1, 2$  transmit antennas is shown for comparison.



Figure 3.4: Figure shows the simulation result for precoding with partial channel information when  $M_T = 2$  transmit antennas are used for K = 10 users and  $\rho = 0.7$ . Different values of  $f_d \Delta t$  are used. Performance of MRW with perfect channel information for a single user system with  $M_T = 1, 2$  transmit antennas is shown for comparison.

From Figure 3.2, observe that for  $f_d \Delta t = 0.05$ , the correlation function has a value of 0.95. Some diversity gain is available when  $f_d \Delta t < 0.05$ . This implies that the available information should have a correlation greater than 0.95 for our precoders to achieve full diversity. Irrespective of how channel estimation is implemented at the transmitter, it is necessary that the downlink receivers should send signals on the uplink. If training based techniques are used, then the training data should be sent within a maximum time-interval such that the estimated channel can be used for precoding. This maximum time-interval between the transmission of training data will correspond to the delay between estimates. For a given Doppler frequency  $f_d$ , this minimum timeinterval should be less than  $.05/f_d$  so that some diversity can be achieved using the partial channel information obtained via training. This imposes a constraint on the rate with which communication should be established with the uplink. Even if blind techniques are used, a minimum data-rate constrain will be imposed on the uplink. Thus, we can say that for implementing precoding in a practical scenario, communication will have to be established with the uplink at a minimum rate. Table 3.2 indicates the possible values of delay which correspond to a value of correlation greater than 0.95, for given values of the relative velocity between the transmitter and receiver and the frequency band employed.

Table 3.2: Maximum delay  $\Delta t$ , between actual channel and available estimate, corresponding to  $f_d \Delta t < 0.05$ , for different velocities and frequency bands.

Frequency Band	Velocity	Maximum Delay	Minimum Uplink data Rate
(GHz)	(mi/hr)	(ms)	(bps)
1	40	2.2	454
	60	1.4	715
2	40	1.1	910
	60	0.72	1390
3	40	0.72	1390
	60	0.48	2084

This table indicates the maximum delay with which the channel estimates should be updated. The corresponding minimum rate with which uplink communication should take place is also indicated in the table. In the 2 GHz band, assuming the vehicle is moving at 60 mi/hr, we see that the required rate on the uplink is around 1.5 kbps or more per second. This is not very high for data or voice communications, where the minimum data rate is 8 kbps. Future networks, aimed at higher data rates are centered in the bands between 2 and 3 GHz. The required uplink rate, corresponding to the constraint of  $f_d \Delta t < 0.05$ , would be around 2 kbps in this band when moving at a velocity of 60 miles/hr. Future standards guarantee minimum uplink data rates that are more than 50 times this data rate. Thus, with good channel estimation techniques precoders with partial channel information could exploit diversity, thereby enhancing the energy efficiency of low-power receivers.

### 3.5 Conclusion

In this chapter we derived precoders for the general case where partial channel information, that is statistically related to the actual channel, is available at the transmitter. These general expressions for the precoders could be used when any form of channel information, correlated to the actual channel, is available. We computed and simulated precoders for the special case where the partial channel information is available in the form of old but perfect channel estimates. We simulated precoders assuming the Jake's model. Results indicate that these precoders can achieve almost full diversity provided the quality of the channel information is high. This corresponds to the condition that the product of the Doppler frequency and the delay between the estimates,  $f_d \Delta t < 0.05$ . This being the case for slow fading channels, full diversity gain is available for slow fading channels. No diversity gain is available when the channel is fast fading. To implement precoding in a practical scenario, a minimum rate constrain is imposed on the uplink. For mobiles moving at high speeds, and hence, having high Doppler frequencies, uplink data rates exceeding a couple of kbps are required. The fact that future generation systems are specified to operate at much higher rates motivates the use of precoding for such systems.

# Precoding for Frequency Selective Channels

Precoders for multi-input single-output (MISO) CDMA systems with flat fading channels, assuming partial or perfect CSI at the transmitter, have been developed in the previous chapters. In Chapter 1, we saw that when the multipath delay spread  $T_d$ , is greater than the symbol period  $T_s$ , the resulting channel is frequency selective and may result into intersymbol interference (ISI) unless zero-padding is employed. In this chapter we consider the design and performance of linear transmitter precoding techniques for the downlink of MISO CDMA systems over frequency selective channels with and without intersymbol interference (ISI). Our design is motivated by a technique known as pre-Rake filtering [8]. Part of the work presented here is published in [100, 109].

### 4.1 Introduction

We know from Section 1.3.1 that maximum ratio combining (MRC) results in maximum likelihood receiver performance when diversity combining is possible at the receiver. However, from (1.70) we observe that the receiver needs to know  $r_i$ , the signal received with fading  $h_i$ , as well as  $h_i$  to perform MRC. In Chapter 1 we saw that a frequency-selective channel can be modelled as a multi-tap filter. We can say that each tap corresponds to a distinct path from the transmitter to the receiver. If the receiver can resolve the multiple paths, i.e., isolate the signals received due to each tap in the multi-tap model for a frequency selective channel, and have some information about the instantaneous signal to noise ratio (SNR) on each path, diversity combining can be employed. The use of a Rake receiver for this purpose was first proposed in [59]. The Rake receiver gets its name from a garden Rake which has a set of metal fingers at its end. In CDMA communication over frequency selective channels, a Rake receiver enables path separation and, consequently, diversity exploitation. We refer to diversity so obtained as frequency diversity since it is obtained by exploiting the frequency selectivity of the channel. A Rake receiver uses the autocorrelation properties of a spreading sequence to resolve multipath components from the received signal. This can be followed by optimal combining of the components, which in most cases is MRC provided  $h_i$ is perfectly known. The following briefly elaborates the working of a Rake receiver.

Let us assume a single user, single antenna system. Hence, no user or antenna index will be used in this discussion. The transmitted signal  $x \in \mathbb{C}^{N \times 1}$  will be given as

$$\boldsymbol{x} = \boldsymbol{s}\boldsymbol{b},\tag{4.1}$$

where s is the spreading sequence employed and  $b \in \{\pm \sqrt{\mathcal{E}}_b\}$ . As we saw in Chapter 1, the each tap of the multi-tap filter corresponds to an independent complex Gaussian process, and is hence a complex Gaussian random variable. We have assumed the channel to be Rayleigh fading, hence each of these taps will be zero-mean complex Gaussian random variables. Assuming a channel of order L-1, i.e., a channel with a L-tap finite impulse response (FIR), the channel impulse response  $h(\tau)$  is given using  $T_c = 1/B$ , where  $T_c$  is the duration of a single chip in the spreading waveform, in (1.53) as

$$h(\tau) = \sum_{l=0}^{L-1} h[l]\delta(\tau - lT_c).$$
(4.2)

We can say that h[l] is the  $l^{th}$  path from the transmit antenna to the receive antenna of the user. Note that we have assumed a single chip-delay between any two consecutive paths. The channel being frequency selective, the transmitted signal of duration  $NT_c$  will spread in time and appear at the receiver as a signal of duration  $(L + N)T_c$ . As shown in Appendix E, the receiver match filters the continuous-time received signal r(t) with respect to the chipping pulse shape and collects N + L match filter outputs to form the discrete-time received signal  $\mathbf{r} \in C^{(N+L)\times 1}$ , which can be written as

$$\mathbf{r} = \sum_{l=0}^{L-1} h[l]\mathbf{s}[l]\mathbf{b} + \mathbf{n}, \qquad (4.3)$$

where  $\boldsymbol{s}[l] \in \mathcal{C}^{N+L\times 1}$ , as shown in (E-12), is the vector  $\boldsymbol{s}$  shifted down by l with zeros padded from the top and L-l zeros padded from below, and  $\boldsymbol{n} \in \mathcal{N}(0, \frac{N_0}{2(L+N)}\boldsymbol{I}_{N+L})$  is the corresponding noise term. A Rake receiver has a bank of L filters. The  $l^{th}$  filter in the bank is matched to the  $l^{th}$ shifted version of the spreading sequence, i.e., matched to  $\boldsymbol{s}[l]$ . An ideal spreading sequence has white power spectral density, or equivalently, its autocorrelation is the Kroneckor delta  $\delta[\tau]$ , i.e.,

$$\boldsymbol{s}[l]^{H}\boldsymbol{s}[m] = \delta[l-m]. \tag{4.4}$$

Assuming a spreading sequence with ideal autocorrelation properties and perfect synchronization between the transmitter and receiver, the output of filter l will be the signal h[l]b. Thus, the output of each filter will be the original information bearing signal, received over a distinct path from the transmitter to the receiver. These outputs can now be optimally combined using MRC to minimize the BER and improve the system performance. To enable MRC, where channel coefficients need to be used, training based techniques can be used to estimate h[l], provided the channel is slowly fading. We are not concerned with this part and will assume that the channel coefficients are known, in our case, at the transmitter.

In [8], a technique called pre-Rake diversity combining was introduced to maximize the diversity gain without receiver CSI for frequency selective channels. Channel information is assumed available at the transmitter instead. The transmitter is now equipped with a filter called pre-Rake filter, which is chosen such that the resulting system gives the same performance as that with a Rakereceiver employing MRC. The idea is to simply pass the signal to be transmitted through the pre-Rake filter. The pre-Rake filter response,  $c(\tau)$  is corresponds to the time-inverted channel filter with tap coefficients c[l] given by

$$c[l] = \frac{h^*[L-1-l]}{\sqrt{\sum_{i=0}^{L-1} |h[i]|^2}}, \qquad l = 0, \cdots, L-1.$$
(4.5)

We can write

$$c(\tau) = \sum_{l=0}^{L-1} c[l]\delta(\tau - lT_c)$$

$$= \sum_{l=0}^{L-1} \frac{h^*[L-1-l]}{\sqrt{\sum_{i=0}^{L-1} |h[i]|^2}} \delta(\tau - lT_c).$$
(4.6)

The denominator corresponds to a scaling factor used to ensure that the total transmit power is normalized to  $E\{x^Hx\}$ . To illustrate the structure of a pre-Rake filter, consider a 2-path channel

with the multi-tap channel response given as

$$h(\tau) = h[0]\delta(\tau) + h[1]\delta(\tau - T_c)$$
(4.7)

Then the corresponding pre-Rake filter response is given as

$$c(\tau) = \frac{1}{\sqrt{|h[1]|^2 + |h[2]|^2}} \left(h^*[1]\delta(\tau) + h^*[0]\delta(\tau - T_c)\right)$$
(4.8)

The signal at the input of the pre-Rake will pass through both the filters  $h(\tau)$  and  $c(\tau)$ . The effective filter response between the input to the pre-Rake and signal received at the receiver can be written as the convolution of  $c(\tau)$  and  $h(\tau)$  as

$$c(\tau) \star h(\tau)$$

$$= \frac{1}{\sqrt{|h[1]|^2 + |h[2]|^2}} \left(h^*[1]h[0]\delta(\tau) + (h^*[0]h[0] + h^*[1]h[1])\delta(\tau - T_c) + h^*[0]h[1]\delta(\tau - 2T_c)\right),$$
(4.9)

where  $\star$  represents the convolution operator.



Figure 4.1: Figure shows the filter response of a 2-tap channel, the corresponding pre-Rake filter response and the effective channel output. The scaling factor  $\gamma = \frac{1}{\sqrt{|h[1]|^2 + |h[2]|^2}}$ . The taps are spaced  $T_c$  seconds apart.

Figure 4.1, where  $\gamma = \frac{1}{\sqrt{|h[1]|^2 + |h[2]|^2}}$ , shows a typical 2-path channel, the corresponding pre-Rake filter response, and the effective filter as seen between the input to the pre-Rake filter and the receiver. The second tap of the effective filter given as

$$\gamma \left( |h[0]|^2 + |h[1]|^2 \right) = \sqrt{|h[0]|^2 + |h[1]|^2}$$
(4.10)

and the output weighted by this tap will correspond to the output of MRW over paths h[0] and h[1]. In the absence of the cross-terms  $h^*[0]h[1]$  and  $h^*[1]h[0]$  full diversity would be achieved.

The performance of this technique is limited by interference, a direct consequence of frequency selectivity or the multi-tap nature of the channel impulse response. One type of interference, which we have already seen is intersymbol interference (ISI) where adjacent symbols interfere with each other. This, however, can be eliminated by using zero-padding. In spite of zero padding to avoid ISI the system is still interference limited even in a single user scenario. To see the different forms of interference besides ISI, possible in a frequency selective channel employing a pre-Rake filter, let us use the following notation. Let  $s_k$  be the spreading sequence of user k. Let  $s_k[n] \in C^{N \times 1}$ , which we will call as a shifted version of the sequence  $s_k$  be defined as

$$\boldsymbol{s}_{k}[n] = \begin{cases} \begin{bmatrix} \boldsymbol{0}_{n \times 1} \\ [\boldsymbol{s}_{k}]_{1:N-n} \end{bmatrix} & n > 0 \\ \boldsymbol{s}_{k} & n = 0 \\ \begin{bmatrix} \boldsymbol{s}_{k} \end{bmatrix}_{N-n+1:N} \\ \boldsymbol{0}_{n \times 1} \end{bmatrix} & n < 0 \end{cases}$$
(4.11)

When pre-Rake transmission is used, the following properties are desired of the spreading sequences:

1. The transmitted spreading sequence of the desired user  $s_k$  is orthogonal to the transmitted spreading sequence of all other users. This can be mathematically represented as

$$\langle \mathbf{s}_j, \mathbf{s}_k \rangle = 0, \qquad k \neq j.$$
 (4.12)

This corresponds to the case of orthogonal spreading we discussed in Chapter 2.

2. The shifted sequences of all the other users are orthogonal to the transmitted spreading sequence of the desired user, i.e.,

$$< s_{j}[n], s_{k} >= 0, \qquad k \neq j, n \neq 0.$$
 (4.13)

3. Shifted versions (with non-zero shifts) of the desired user's spreading sequence are orthogonal to the transmitted spreading sequence of the desired user, i.e.,

$$\langle s_k[n], s_k \rangle = 0, \qquad n \neq 0.$$
 (4.14)

For use in this chapter we define the following interference terms:

- Multi-Access Interference (MAI): When the transmitted spreading sequences used are nonorthogonal, i.e., when property 1 is violated we refer to the resulting interference as MAI. This is the same as the interference we eliminated in Chapters 2 and 3.
- 2. InterChip Interference (ICI): When properties 2 or 3 are not satisfied, similar to ISI, adjacent chips interfere with each other. In this thesis we refer to this as ICI.
- 3. Multi-User Interference (MUI): When properties 1 or 2 are not satisfied we refer to the resulting interference as MUI. Note that MAI itself is a form of MUI.
- 4. Self-Interference: When property 3 above is not satisfied we refer to the resulting interference as self-interference.
- 5. Intersymbol Interference (ISI): When zero padding is avoided to enhance the rate, adjacent symbols interfere with each other. We refer to this as ISI. Note that interference arises due to the symbols of the desired user as well as the symbols of other users which are transmitted instead of the zero-pad when the properties 1-3 are not satisfied. We will refer to this form of interference arising from the absence of zero-padding as ISI.

It is not possible to design sequences that satisfy all the properties mentioned above and thus interference is unavoidable. Even if the transmitted sequences are chosen to be orthogonal to eliminate MAI, there will be interference due to ICI which results when property 2 is not satisfied (i.e. from MUI without MAI), and when property 3 is not satisfied (i.e., from self-interference). We will often combine MUI minus the MAI and the self-interference and refer to it as ICI.

Similar techniques, like pre-Rake filtering were proposed for space-time architectures in [107]. In both [8] and [107], MUI is either ignored or modelled as AWGN and no attempt has been made to minimize the interference. Motivated by the pre-Rake mechanism, in this chapter we develop precoders which, besides exploiting multi-antenna diversity, use the knowledge of the MUI to eliminate it. Again, our receivers are constrained to simple match filters while the transmitter is assumed to have complete or partial channel state information (CSI). Our contribution in this chapter is the design of linear interference cancelling precoders, which not only cancel interference, but also exploit the available multi-antenna and frequency diversity to improve the SNR performance. When using  $M_T$  transmit antennas with L paths between each transmit antenna and each receiver, the total number of distinct paths, and thus as per Section 1.3.1, the maximum achievable diversity gain is  $M_T L$ . We design precoders such that the resulting decision statistics at the receivers closely mimic those due to maximum ratio weighting for a system with  $M_T L$  distinct paths. We assume that the channel coefficients are perfectly known at the transmitter. Unlike the flat-fading case, where each antenna is equipped with only one precoder, in this case, depending on the order of the channel filter, each antenna is equipped with more than one precoder. In the flat fading case, precoders at each antenna were designed independently. Here, we can design a precoder individually. However, as we will see, jointly designing all L precoders at each antenna results in better BER performance.

We will approach the precoder design problem for a frequency selective channel in two ways. In the first method, ISI is eliminated by zero-padding between consecutive symbol vectors. A symbol vector is a vector having one information symbol of each user. It is similar to the vector  $\boldsymbol{b}$  used in the previous discussions. In the other approach, we form frames by stacking more than one symbol vector. The precoders are now derived to operate upon an entire frame of symbol vectors instead of a single symbol vector. The tradeoff involved for this complexity is an increase in rate and interference, because elimination of zero padding between symbol vectors enhances the rate but simultaneously gives rise to ISI. As mentioned in Chapter 1, we will often refer to the resulting channel where zero padding is used as a frequency selective channel without ISI and that when zero-padding is not used as a frequency selective channel with ISI.

In Section 4.2 we re-introduce our MISO CDMA system model, incorporating suitable changes introduced by the frequency selective nature of the channel, and the pre-Rake technique. Frequency selective channels without ISI are considered in Section 4.3, while ISI channels are considered in Section 4.4. The sub-optimality resulting from designing all precoders independently is shown for channels without ISI by simulation. Similar to the approach in previous chapters, we formulate MMSE cost functions, the solution to which gives the optimal precoders. We present our results in Section 4.5 and finally conclude in Section 4.6.

### 4.2 System Model

We consider a K-user synchronous multipath channel, appropriate for the CDMA downlink, where precoding is most relevant. The transmitter is equipped with  $M_T$  antennas while each receiver is equipped with a single antenna. For simplicity, we will assume that there are L paths from each transmit antenna to each receiver. The spreading codes are assumed to be non-orthogonal and the spreading gain is fixed to be N. We assume that the channel is sufficiently slow fading so that the perfect channel knowledge can be used for designing the precoders. We also derive precoders when partial CSI, statistically correlated to the actual channel, is available to the transmitter. For simplicity, we assume BPSK modulation and unit chip duration  $T_c$  as the path delay between successive paths. The channel between transmit antenna a and receiver k is assumed to be  $\sum_{l=0}^{L-1} h_k^{(a)}[l]\delta(t - lT_c)$ , where  $h_k^{(a)}[l] \sim \mathcal{N}_c(0, \sigma_h^2)$  corresponds to the  $l^{th}$  path from the  $a^{th}$  transmit antenna to the  $k^{th}$  receiver. The vector  $\mathbf{b}$ , with the  $i^{th}$  element corresponding to a single symbol of user i, will be referred to as a symbol vector throughout this discussion.

First we will look into the details of pre-Rake transmission when multiple antennas are used to communicate to K users. We consider the transmission of a single symbol b[k] per user. Let  $\mathbf{x}^{(a)}$  be the discrete-time signal transmitted from antenna a using pre-Rake filtering. There are K pre-Rake filters for each user at each antenna and the signal transmitted from each antenna is the sum of the outputs of the pre-Rake filters used at that antenna. Each pre-Rake filter will be of length L. The input pre-Rake filter k will be the signal  $\mathbf{s}_k b[k]$ , the spread signal of the corresponding user kwhich is of length N. The output of each pre-Rake filter, and hence the signal transmitted from antenna a, will be of length N+L-1. The corresponding continuous-time signal will be of duration  $(N + L - 1)T_c$ . The signal transmitted from antenna a during the  $i + 1^{th}$  time slot is represented as  $[\mathbf{x}^{(a)}]_i$  and be written as

$$[\boldsymbol{x}^{(a)}]_{i} = \sum_{i=1}^{K} \sum_{l=0}^{L-1} h_{k}^{(a)*} [L-1-l][\boldsymbol{p}_{k}]_{i-l}, \qquad i = 0 \dots N + L - 1,$$
(4.15)

where  $p_k/\gamma_k$  is the signal corresponding to user k, i.e.,

$$[\mathbf{p}_{k}]_{0:N-1} = \gamma_{k} \mathbf{s}_{k} b[k]$$

$$[\mathbf{p}_{k}]_{i} = 0, \qquad i > N-1,$$
(4.16)

and the power scaling coefficient  $\gamma_k = \frac{1}{\sqrt{\sum_{a=1}^{M_T} \sum_{l=0}^{L-1} |h_k^{(a)}[l]|^2}}$ , so that the average transmit power per

user is normalized to  $\mathcal{E}_b$ . Each time slot is of duration  $T_c$ . Figure 4.2 illustrates the transmission of  $p_k$  for the pre-Rake scheme.

$$[\boldsymbol{p}_k]_i = \gamma_k[\boldsymbol{s}_k]_i b[k], \qquad i = 0 \dots N - 1$$

Transmission Interval

Figure 4.2: Figure shows the signal at the output of the pre-Rake filter for a particular user k.  $p_K/\gamma_k$  is the signal transmitted to be transmitted for user k. We assume that there are L = 3 resolvable paths. l = 0 is the shortest path and l = L - 1 = 2 is the longest path.

The transmitted signal is a sum of such signals for each user. The receiver is assumed to be chip-synchronous with the transmitter, and starts reception during the  $L^{th}$  time slot and continues for the next N time slots. During the reception interval indicated, the signals arrive coherently at the antenna of receiver k. We say that coherent summation of the signals  $[\mathbf{p}_k]_i$  takes place at receiver k during the reception interval. The noise-free signal  $\mathbf{r}_k^{(a)}$  received at receiver k during the indicated reception interval due to the transmission from transmit antenna a , will be such that the  $i^{th}$  element  $[\mathbf{r}_k^{(a)}]_i$  can be written as

$$\begin{bmatrix} \boldsymbol{r}_{k}^{(a)} \end{bmatrix}_{i}$$

$$= \sum_{l=0}^{2} |h_{k}^{(a)}[l]|^{2}[\boldsymbol{p}_{k}]_{i} + \underbrace{h_{k}^{(a)}[0]h_{k}^{(a)*}[1][\boldsymbol{p}_{k}]_{i+1} + h_{k}^{(a)}[0]h_{k}^{(a)*}[2][\boldsymbol{p}_{k}]_{i+2} + h_{k}^{(a)}[1]h_{k}^{(a)*}[2][\boldsymbol{p}_{k}]_{i+1}}_{\text{Desired Signal}}$$

$$+ \sum_{l=0}^{2} h_{k}^{(a)}[l] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[l][\boldsymbol{p}_{j}]_{i} + h_{k}^{(a)}[0] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[1][\boldsymbol{p}_{j}]_{i+1}}_{j\neq k}$$

$$+ h_{k}^{(a)}[2] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[0][\boldsymbol{p}_{j}]_{i+2} + h_{k}^{(a)}[1] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[2][\boldsymbol{p}_{j}]_{i+1}$$

$$+ h_{k}^{(a)}[2] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[0][\boldsymbol{p}_{j}]_{i+2} + h_{k}^{(a)}[1] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[2][\boldsymbol{p}_{j}]_{i+1}$$

$$+ h_{k}^{(a)}[2] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[0][\boldsymbol{p}_{j}]_{i+2} + h_{k}^{(a)}[1] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[2][\boldsymbol{p}_{j}]_{i+1}$$

$$+ h_{k}^{(a)}[2] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[0][\boldsymbol{p}_{j}]_{i+2} + h_{k}^{(a)}[1] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[2][\boldsymbol{p}_{j}]_{i+1}$$

$$+ h_{k}^{(a)}[2] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[0][\boldsymbol{p}_{j}]_{i+2} + h_{k}^{(a)}[1] \sum_{\substack{j=1\\ j\neq k}}^{K} h_{j}^{(a)*}[2][\boldsymbol{p}_{j}]_{i+1}$$

Observe how, due to the pre-Rake mechanism, coherent summation of signals reaching the receiver through different paths takes place. Coherently added (desired signal) versions of  $[p_i^{(a)}]_i$  are available at each of the receivers,  $j = 1, \dots, K$ , during the reception interval. In all cases, the interference terms are present. We will focus on cancelling the self interference as well as the MUI terms resulting at all the receivers. Note that the receiver needs to determine the exact time-slot during which to begin reception. As a consequence of the coherent summation represented by the desired signal in (4.17), most of the power is received starting from the duration corresponding to the  $L^{th}$  time slot. Hence, the receiver needs to simply detect if the received power level is above a particular threshold and start reception from that time slot. We assume that the receiver detects the time-slot to begin reception correctly.

In the pre-Rake mechanism, the pre-Rake filter used is of length L. We can say that L weighted copies of the same signal are transmitted from each antenna. Similar to pre-Rake diversity combining [8], with precoding, we will transmit L copies of every signal from each transmit antenna. Each transmit antenna is equipped with L precoders, one for each path, and each activated at a different time instant to achieve the pre-Rake effect. We will represent the  $l^{th}$  precoder from transmit antenna a as  $M^{(a)}[l]$ . The precoder for the longest path L-1, is activated during the first time slot, whereas the precoder for the shortest path is activated during the  $L^{th}$  time slot. In fact, the pre-Rake case can be viewed as a special case of precoding. If ideal sequences as mentioned in Section 4.1 were to be used, there would be no ICI. Consequently, the self-interference and MUI terms in (4.17) would disappear, and the diagonal precoding matrix

$$\left[\boldsymbol{M}^{(a)}[l]\right]_{k,k} = \frac{h_k^{(a)*}[l]}{\sqrt{\sum_{a=1}^{M_T} \sum_{j=0}^{L-1} |h_k^{(a)}[j]|^2}}.$$
(4.18)

would result in optimal diversity performance at each receiver provided we eliminate ISI by zeropadding. However, it is not possible to design such sequences. This serves as the primary motivation for our designing precoders here. The design of our precoders will be such that

- Interference resulting from non-orthogonal spreading (MAI) is eliminated.
- Interference resulting from ICI is eliminated.
- Multiple antenna diversity gain is maximized.
- Frequency diversity gain is maximized.
- When zero-padding is avoided, the resulting ISI should be eliminated.

In the following section we will look at precoders for channels without ISI.

## 4.3 Precoding for Frequency Selective Channels without Inter-Symbol Interference

In this section we consider zero padding between successive symbol vectors  $\boldsymbol{b}$ , to eliminate ISI. To design the precoders we can take two approaches. One would be to optimize the signal resulting at all receivers from transmission due to a particular precoder  $\boldsymbol{M}^{(a)}[l]$ . In this case all precoders are designed individually. Alternately, we can try to optimize the received signal resulting at all receivers due to transmission from all precoders from a particular antenna. In this case precoders to be used at the same transmit antenna are designed jointly. The first method is considered in Section 4.3.1, while the second method is considered in Section 4.3.2.

#### 4.3.1 Individually Optimal Precoders

We begin by writing down the vector of decision statistics, formed due to transmission from a particular precoder, and then derive the precoder by forming and solving an appropriate MMSE cost function. As specified in the system model, each precoder is activated during a different time slot. Precoder  $M^{(a)}[l]$  is activated at the beginning of the  $L - l^{th}$  time slot. Thus precoder  $M^{(a)}[L-1]$  is activated during the first time slot and precoder  $M^{(a)}[0]$  is activated during the  $L^{th}$  time slot. This can be modelled by using vertically shifted versions of the spreading matrix S. Let  $\tilde{S}[l] \in \mathbb{C}^{(N+L-1)\times K}$ , be an extension of the spreading matrix S corresponding to the activation of the precoder  $M^{(a)}[l]$ . Here  $\tilde{S}[l]$  can be written as

$$\tilde{\boldsymbol{S}}[l] = \begin{bmatrix} \boldsymbol{0}_{K \times (L-1-l)} & \boldsymbol{S}^T & \boldsymbol{0}_{K \times l} \end{bmatrix}^T.$$
(4.19)

As an example, we will have

$$\tilde{\boldsymbol{S}}[L-1] = \begin{bmatrix} \boldsymbol{S}^T & \boldsymbol{0}_{K \times L-1} \end{bmatrix}^T, \qquad (4.20)$$

and

$$\tilde{\boldsymbol{S}}[0] = \begin{bmatrix} \boldsymbol{0}_{K \times L-1} & \boldsymbol{S}^T \end{bmatrix}^T.$$
(4.21)

Using this notation, the spreading and successive activation of different precoders can be accounted for by writing down the signal transmitted from antenna a as,  $\boldsymbol{x}^{(a)} \in \mathbb{C}^{(N+L-1)\times 1}$  as,

$$\boldsymbol{x}^{(a)} = \alpha \sum_{l=0}^{L-1} \tilde{\boldsymbol{S}}[l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}$$

$$= \alpha \breve{\boldsymbol{S}} \boldsymbol{\mathcal{M}}^{(a)} \boldsymbol{b}, \qquad a = 1, \dots, M_T.$$
(4.22)

where,

$$\breve{\boldsymbol{S}} = \begin{bmatrix} \tilde{\boldsymbol{S}}[0] & \tilde{\boldsymbol{S}}[1] & \cdots & \tilde{\boldsymbol{S}}[L-1] \end{bmatrix}_{(N+L-1)\times KL},$$
(4.23)

$$\mathcal{M}^{(a)} = \begin{bmatrix} \mathbf{M}^{(a)}[0] \\ \vdots \\ \mathbf{M}^{(a)}[L-1] \end{bmatrix}_{KL \times K}$$
(4.24)

The scaling factor  $\alpha$  is used to constrain the total transmit power and **b**, as always, is the symbol vector of the K symbols of the K users. Similar to the computation of  $\alpha$  in (2.20), here we have

$$\alpha = \sqrt{\frac{K}{\sum_{a=1}^{M_T} \operatorname{tr}\left\{\boldsymbol{\mathcal{M}}^{(a)H} \check{\boldsymbol{S}}^H \check{\boldsymbol{S}} \boldsymbol{\mathcal{M}}^{(a)}\right\}}}.$$
(4.25)

Using the fact that

$$\breve{\boldsymbol{S}}^{H}\breve{\boldsymbol{S}} = \sum_{l=0}^{L-1} \tilde{\boldsymbol{S}}^{H}[l]\tilde{\boldsymbol{S}}[l] 
= L\boldsymbol{R} \qquad \because \tilde{\boldsymbol{S}}^{H}[l]\tilde{\boldsymbol{S}}[l] = \boldsymbol{R},$$
(4.26)

which results from the zero-padded structure of  $\tilde{\boldsymbol{S}}^{H}[l]$  in (4.19), we get

$$\alpha = \sqrt{\frac{K}{L \sum_{a=1}^{M_T} \operatorname{tr} \left\{ \mathcal{M}^{(a)H} \mathcal{R} \mathcal{M}^{(a)} \right\}}}.$$
(4.27)

Each of the K receivers start reception and match filtering from the chip duration corresponding to the  $L^{th}$  time slot and continue for the next N time slots. We also assume that the receiver is chip synchronous with the transmitter and always starts match filtering from the correct time slot. Let us look at the signal received at user k due to transmission from precoder  $M^{(a)}[l]$ . The signal transmitted from precoder  $M^{(a)}[l]$  will pass through all the paths and during the reception interval it will be incident from multiple paths with different delays. The signal arriving from path l contributes to the desired signal part of the net received signal at user k. The part received from
98

paths j < l and j > l will result into interference in the absence of ideal spreading. Neglecting noise and the power scaling factor  $\alpha$ , the signal received by user k through all paths, due to transmission from  $M^{(a)}[l]$  is represented as  $r_k^{(a)}[l]$  and can be written as shown in Appendix E.1 as,

$$\boldsymbol{r}_{k}^{(a)}[l] = \sum_{j=0}^{L-1} h_{k}^{(a)}[j]\boldsymbol{S}[j-l]\boldsymbol{M}^{(a)}[l]\boldsymbol{b}$$
  
=  $\underbrace{h_{k}^{(a)}[l]\boldsymbol{S}\boldsymbol{M}^{(a)}[l]\boldsymbol{b}}_{\text{Desired Signal+MAI}} + \underbrace{\sum_{\substack{j=0\\i\neq l}}^{L-1} h_{k}^{(a)}[j]\boldsymbol{S}[j-l]\boldsymbol{M}^{(a)}[l]\boldsymbol{b}}_{\text{ICI}}, \quad \because \boldsymbol{S}[0] = \boldsymbol{S},$  (4.28)

where  $S[n] \in \mathbb{C}^{N \times K}$  is defined in (E-21). We observe that S[n] is generated from the spreading code matrix with the codes shifted down by n, zeros being inserted from the top when n > 0. When n < 0, zeros are inserted from the bottom and the codes are shifted up. Note that unlike the definition of  $\tilde{S}[l]$  given in (4.19), where zeros are appended from both sides, in S[n] zeros are forced from one side and the size of the matrix S[n] is the same of the spreading matrix. The ICI is now clearly seen. The net signal received by user k will be the sum of the signals arriving from different paths. We can write the net received noise-free signal  $r_k$  as

$$\mathbf{r}_{k} = \sum_{a=1}^{M_{T}} \sum_{l=0}^{L-1} \mathbf{r}_{k}^{(a)}[l].$$
(4.29)

At each receiver k, a decision statistic is formed by match filtering with respect to the spreading sequence  $s_k$ . Similar to the decomposition of the noise free decision statistic into statistics resulting due to transmission from a single precoder in (2.41), the noise-free decision statistic  $y'_k$  formed at user k can be expressed as

$$y'_{k} = s^{H}_{k} \boldsymbol{r}_{k}$$

$$= \sum_{a=1}^{M_{T}} \sum_{l=0}^{L-1} y^{(a)'}_{k}[l],$$
(4.30)

where

$$y_{k}^{(a)'}[l] = \mathbf{s}_{k}^{H} \mathbf{r}_{k}^{(a)}[l]$$
  
=  $\sum_{i=0}^{L-1} h_{k}^{(a)}[i] \mathbf{s}_{1}^{H} \mathbf{S}[i-l] \mathbf{M}^{(a)}[l] \mathbf{b}.$  (4.31)

Using the notation

$$\boldsymbol{h}_{k}^{(a)} = \left[ \begin{array}{ccc} h_{k}^{(a)}[0] & h_{k}^{(a)}[1] & \cdots & h_{k}^{(a)}[L-1] \end{array} \right]_{L \times 1}^{T},$$
(4.32)

and

$$\boldsymbol{R}_{k}[l] = \begin{bmatrix} \boldsymbol{s}_{k}^{H}\boldsymbol{S}[-l] \\ \boldsymbol{s}_{k}^{H}\boldsymbol{S}[1-l] \\ \vdots \\ \boldsymbol{s}_{k}^{H}\boldsymbol{S}[L-1-l] \end{bmatrix}_{L \times K}, \qquad (4.33)$$

we can now write

$$y_k^{(a)'}[l] = h_k^{(a)T} R_k[l] M^{(a)}[l] b,$$
 (4.34)

which is the noise-free decision statistic at receiver k due to precoder  $M^{(a)}[l]$ . Stacking all the decision statistics due to transmission from precoder  $M^{(a)}[l]$  at all the receivers, and incorporating the corresponding Gaussian noise terms  $n^{(a)}[l] = [n^{(a)}[1], \dots, n^{(a)}[K]]$ , which has a distribution given as  $\mathcal{N}_c(\mathbf{0}, \frac{N_0}{2M_T L} \mathbf{R})$ , we form the vector of decision statistics resulting due to transmission from the precoder  $M^{(a)}[l]$  as  $y^{(a)}[l]$  which can be written as

$$\boldsymbol{y}^{(a)}[l] = \begin{bmatrix} y_1^{(a)'}[1] & y_2^{(a)'}[2] & \cdots & y_K^{(a)'}[K] \end{bmatrix}_{L\times 1}^T + \boldsymbol{n}^{(a)}[l] \\ = \begin{bmatrix} \boldsymbol{h}_1^{(a)T} \boldsymbol{R}_1[l] \\ \boldsymbol{h}_2^{(a)T} \boldsymbol{R}_2[l] \\ \vdots \\ \boldsymbol{h}_K^{(a)T} \boldsymbol{R}_K[l] \end{bmatrix} \boldsymbol{M}^{(a)}[l] \boldsymbol{b} + \boldsymbol{n}^{(a)}[l]$$

$$\boldsymbol{\mathcal{H}}^{(a)}[l] \in \mathbb{C}^{K \times K} \\ = \boldsymbol{\mathcal{H}}^{(a)}[l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b} + \boldsymbol{n}^{(a)}[l].$$

$$(4.35)$$

Here, the terms  $\boldsymbol{n}^{(a)}[l]$  are such that

$$\sum_{a=1}^{M_T} \sum_{l=0}^{L-1} \boldsymbol{n}^{(a)}[l] = \boldsymbol{n}, \qquad (4.36)$$

where  $\boldsymbol{n} \sim \mathcal{N}_c(\boldsymbol{0}, \frac{N_0}{2}\boldsymbol{R})$  contains the Gaussian noise terms formed at each of the K receivers after match filtering the net received signal.

Let us define  $\mathcal{D}^{(a)}[l] \in \mathbb{C}^{K \times K}$  as

$$\boldsymbol{\mathcal{D}}^{(a)}[l] \triangleq \operatorname{diag}\left(\frac{|h_1^{(a)}[l]|^2}{\sqrt{\sum_{t=1}^{M_T} \sum_{j=0}^{L-1} |h_1^{(t)}[j]|^2}}, \cdots, \frac{|h_K^{(a)}[l]|^2}{\sqrt{\sum_{t=1}^{M_T} \sum_{j=0}^{L-1} |h_K^{(t)}[j]|^2}}\right).$$
(4.37)

There are  $M_T$  transmit antennas and L paths. Hence, there are a total of  $M_T L$  diversity channels. The system can be thought of as having  $M_T L$  flat fading channels between the transmitter and the receiver. Consequently, if MRW was employed over these  $M_T L$  channels, the vector of decision statistics given by  $\mathcal{D}^{(a)}[l]\boldsymbol{b}$ , would result from MRW transmission over path l from transmit antenna a, when the system is noise-free and interference-free. Hence, we form MMSE cost functions to choose precoders such that the vector of decision decisions statistic  $\boldsymbol{y}^{(a)}[l]$  formed due to reception over the  $l^{th}$  path from antenna a is close in the MMSE sense to  $\mathcal{D}^{(a)}[l]\boldsymbol{b}$ . We form MMSE cost functions  $J^{(a)}[l]$ , corresponding to the error between the effective decision statistic  $\boldsymbol{y}^{(a)}[l]$  as given in (4.35) and the noise-free and interference-free statistic  $\mathcal{D}^{(a)}[l]\boldsymbol{b}$  as

$$J^{(a)}[l] = E_{\boldsymbol{b},\boldsymbol{n}^{(a)}[l]} \left\{ \left\| \boldsymbol{\mathcal{D}}^{(a)}[l]\boldsymbol{b} - \boldsymbol{\mathcal{H}}^{(a)}[l]\boldsymbol{M}^{(a)}[l]\boldsymbol{b} + \boldsymbol{n}^{(a)}[l] \right\|^{2} \right\}.$$
(4.38)

Solving this cost function using  $J = J^{(a)}[l]$ ,  $\mathbf{M} = \mathbf{M}^{(a)}[l]$ ,  $\mathbf{D} = \mathbf{\mathcal{D}}^{(a)}[l]$ ,  $\mathbf{H} = \mathbf{\mathcal{H}}^{(a)}[l]$ ,  $\mathbf{R} = \mathbf{I}$  and  $\mathbf{n} = \mathbf{n}^{(a)}[l]$  in (C-2) we now get

$$\boldsymbol{M}^{(a)}[l] = \left[\boldsymbol{\mathcal{H}}^{(a)}[l]\right]^{-1} \boldsymbol{\mathcal{D}}^{(a)}[l].$$
(4.39)

Since the matrix  $\mathcal{H}^{(a)}[l]$  is square and the channel coefficients are uncorrelated and Gaussian (hence, independent), we see that the inverse exists since  $\mathcal{H}^{(a)}[l]$  is full rank with probability 1. The performance of this scheme is shown as the solid blue plot in Figure 4.3. It corresponds to the case where the peak to average transmitted power ratio, represented as  $\beta$  and explained below, equals 1. As we see from the figure all diversity is lost.

We have been using the factor  $\alpha$  as defined in (4.27) to scale the transmitted signal in order to transmit at a constant total power  $K\mathcal{E}_b$ . Now we allow for variations in the instantaneous transmitted power and exploit the channel variations. We assume that when the channels are relatively good, the transmit power can be low and when they are bad we can transmit at high power. As a result, while the performance for good channels is not significantly affected, that for bad channels improves significantly due to higher transmission power. The instantaneous power is allowed to vary with the channel but the average transmitted power is kept constant. We define  $\beta$  as the peak to average power ratio. So far we have been assuming a fixed transmission power level corresponding to  $\beta = 1$ . Now the peak power is limited to  $\beta K\mathcal{E}_b$ . Let  $\rho = \sum_{(a)} \sum_l tr\{\mathbf{M}^{(a)H}[l]\mathbf{R}\mathbf{M}^{(a)}[l]\}$ . Then, in the absence of power scaling,  $\rho\mathcal{E}_b$  would be the total average transmitted power using precoders  $\mathbf{M}^{(a)}[l]$ . From (4.39), due to the presence of  $[\mathbf{\mathcal{H}}^{(a)}[l]]^{-1}$  in the precoder, when the channels of all users would be relatively poor,  $\rho$  would be high and when the channels would be relatively good  $\rho$  would be low. Using this fact, the transmitted signal now is given by

$$\boldsymbol{x}^{(a)} = \zeta \sum_{i=0}^{L-1} \boldsymbol{S} \boldsymbol{M}^{(a)}[l] \boldsymbol{b}, \qquad (4.40)$$

where

$$\zeta = \begin{cases} \sqrt{\frac{zK}{E}}, & \varrho < \beta K \\ \sqrt{\frac{\beta K}{\varrho}}, & \varrho \ge \beta K. \end{cases}$$
(4.41)

Here the factor z is given as

$$z = \frac{1 - \beta p_0}{1 - p_0},\tag{4.42}$$

where  $p_0 = P(\rho \ge \beta K)$  and it can be derived from the assumption that the average transmitted power is  $K\mathcal{E}_b$ . The derivation is given in the Appendix F.1. Note that the condition  $\rho < \beta K$ corresponds to the users having relatively good channels, while the condition  $\rho > \beta K$  corresponds to the users having relatively poor channels. Figure 4.3 illustrates the simulated performance of the precoders given in (4.39) with different peak to average power ratios  $\beta$ .



Figure 4.3: Figure illustrates performance achieved with precoders derived individually for each path for a channel with no ISI. *m*-sequences of length N = 15 were used with  $M_T = 2$  transmit antennas, L = 2 paths, and K = 10 users. The BER achieved by using peak to average power ratios  $\beta = 1, 2, 10, 20, 40, \infty$  is shown. The slopes corresponding to the diversity performance of a single user system with L = 1, 2, 4 diversity gain are plotted for comparison.

Observe that when  $\beta = \infty$  we do get full diversity. This case corresponds to no power scaling and is practically infeasible as it leads to infinite transmission power scenarios. For values of  $\beta = 10, 20, 40$ , there is a significant reduction in the diversity gain at high SNRs. The reason for the loss of diversity is that, while the signal transmitted from a particular antenna passes through all the paths from that antenna, we have designed our precoders to optimize only over a single path. Signals transmitted from distinct antennas travel through distinct, independent paths and do not interfere with each other. But signals transmitted from the same antenna pass through the same paths, interfering with each other. Hence, the precoders to be used at a particular antenna need to be jointly designed. Signals transmitted using jointly derived precoders jointly contribute to the energy transmitted across all the paths they pass through. We look at the derivation of these in the following section.

### 4.3.2 Jointly Optimal Precoders

In this section we will jointly design the precoders to be used at a particular antenna. Here we will view the received signal and the corresponding decision statistic resulting from all the transmissions from a particular antenna. We begin by expanding upon (4.28) so as to consider the reception due to transmissions from all the precoders at antenna a. The noise-free signal received by receiver k from transmission at antenna a, represented as  $\mathbf{r}_{k}^{(a)}$ , can be written as

$$\begin{aligned} \boldsymbol{r}_{k}^{(a)} &= \sum_{l=0}^{L-1} \boldsymbol{r}_{k}^{(a)}[l] \\ &= \sum_{l=0}^{L-1} \sum_{j=0}^{L-1} h_{k}^{(a)}[j] \boldsymbol{S}[j-l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b} \\ &= \sum_{\substack{l=0\\ l=0}}^{L-1} h_{k}^{(a)}[l] \boldsymbol{S} \boldsymbol{M}^{(a)}[l] \boldsymbol{b} + \sum_{\substack{l=0\\ i\neq l}}^{L-1} \sum_{\substack{j=0\\ i\neq l}}^{L-1} h_{k}^{(a)}[j] \boldsymbol{S}[j-l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}. \end{aligned}$$
(4.43)

The net received noise-free signal at user k and the noise-free decision statistic formed at user k after match filtering are the same as that given by (4.29), (4.30) and (4.31). The noise-free match

filter output  $y_k^{(a)'}$ , formed at receiver k due to transmission from antenna a can be written as

$$y_k^{(a)'} = \sum_{l=0}^{L-1} y_k^{(a)}[l]$$
(4.44)

$$= \sum_{j=0}^{L-1} \sum_{l=0}^{L-1} h_k^{(a)}[j] \boldsymbol{s}_k^H \boldsymbol{S}[j-l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}.$$
(4.45)

Stacking the decision statistics resulting from match filtering the signal received from antenna a at all the receivers, we form the vector  $\boldsymbol{y}^{(a)} = [y_1^{(a)'}, y_2^{(a)'}, \cdots, y_K^{(a)'}]^T + \boldsymbol{n}^{(a)}$ , where  $\boldsymbol{n}^{(a)} = [n^{(a)}[1], \cdots, n^{(a)}[K]]$ , corresponds to the vector of Gaussian noise terms with distribution given as  $\mathcal{N}_c\left(\mathbf{0}, \frac{N_0}{2M_T}\boldsymbol{R}\right)$ . This vector can be expressed as

$$\boldsymbol{y}^{(a)} = \underbrace{\begin{bmatrix} \boldsymbol{H}^{(a)}[0] & \dots & \boldsymbol{H}^{(a)}[L-1] \end{bmatrix}}_{\boldsymbol{\mathcal{H}}^{(a)}} \cdot \underbrace{\begin{bmatrix} \boldsymbol{R}_{0} & \cdots & \boldsymbol{R}_{-(L-1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{L-1} & \cdots & \boldsymbol{R}_{0} \end{bmatrix}}_{\boldsymbol{\mathcal{R}}} \boldsymbol{\mathcal{M}}^{(a)}\boldsymbol{b} + \boldsymbol{n}^{(a)}$$
(4.46)

 $=\mathcal{H}^{(a)}\mathcal{RM}^{(a)}b+n^{(a)},$ 

where  $\boldsymbol{\mathcal{H}}^{(a)} \in \mathbb{C}^{K \times KL}$ ,  $\boldsymbol{\mathcal{R}}^{(a)} \in \mathbb{C}^{KL \times KL}$ , and

$$\boldsymbol{H}^{(a)}[l] = \operatorname{diag}\{h_1^{(a)}[l], h_2^{(a)}[l], \dots, h_K^{(a)}[l]\}$$
(4.47)

$$\mathbf{R}_{j-l} = \mathbf{S}^H \mathbf{S}[j-l]. \tag{4.48}$$

The steps to get to the form in (4.46) are given in the Appendix F. Here, the terms  $\boldsymbol{n}^{(a)}$  are such that

$$\sum_{a=1}^{M_T} \sum_{l=0}^{L-1} \boldsymbol{n}^{(a)} = \boldsymbol{n}, \qquad (4.49)$$

where  $\boldsymbol{n} \sim \mathcal{N}_c\left(\boldsymbol{0}, \frac{N_0}{2M_T}\boldsymbol{R}\right)$  contains the Gaussian noise terms formed at each of the K receivers after match filtering the net received signal.

As usual, we now form an appropriate cost function and minimize it to find the optimal precoders. Since the reception from transmit antenna a is over L paths, we want to choose precoders such that the resulting decision statistics given by (4.46), closely mimic those of single user maximum ratio weighting (MRW) over the L paths from antenna a to each of the receivers. Using  $\mathcal{D}^{(a)} \in \mathbb{C}^{K \times K}$ , defined as

$$\boldsymbol{\mathcal{D}}^{(a)} \triangleq \operatorname{diag}\left(\frac{\sum_{j=0}^{L-1} |h_1^{(a)}[j]|^2}{\sqrt{\sum_{t=1}^{M_T} \sum_{j=0}^{L-1} |h_1^{(t)}[j]|^2}}, \cdots, \frac{\sum_{j=0}^{L-1} |h_K^{(a)}[j]|^2}{\sqrt{\sum_{t=1}^{M_T} \sum_{j=0}^{L-1} |h_K^{(t)}[j]|^2}}\right),$$
(4.50)

we form the MMSE cost function  $J^{(a)}$  as

$$J^{(a)} = E_{\boldsymbol{b},\boldsymbol{n}} \left\{ \left\| \boldsymbol{\mathcal{D}}^{(a)}\boldsymbol{b} - \boldsymbol{\mathcal{H}}^{(a)}\boldsymbol{\mathcal{R}}\boldsymbol{\mathcal{M}}^{(a)}\boldsymbol{b} + \boldsymbol{n}^{(a)} \right\|^2 \right\}.$$
(4.51)

As indicated, the expectation is with respect to symbol vector  $\boldsymbol{b}$ , and the noise component  $\boldsymbol{n}$ . Minimizing this cost function with respect to  $\mathcal{M}^{(a)}$  gives the precoders that will be used at transmit antenna  $\boldsymbol{a}$ . Solving this cost function using  $J = J^{(a)}$ ,  $\boldsymbol{M} = \mathcal{M}^{(a)}$ ,  $\boldsymbol{D} = \mathcal{D}^{(a)}$ ,  $\boldsymbol{H} = \mathcal{H}^{(a)}$ ,  $\boldsymbol{R} = \mathcal{R}$ and  $\boldsymbol{n} = \boldsymbol{n}^{(a)}$  in (C-2), we can show that the optimal  $\mathcal{M}^{(a)}$  and the corresponding precoders to be used at transmit antenna  $\boldsymbol{a}$  are given by

$$\mathcal{M}^{(a)} = \left[ \mathcal{H}^{(a)} \mathcal{R} \right]^{\dagger} \mathcal{D}^{(a)}$$
(4.52)

$$\boldsymbol{M}^{(a)}[l] = [\boldsymbol{\mathcal{M}}^{(a)}]_{Kl+1:K(l+1),:}, \qquad (4.53)$$

where (4.53) is obtained by simply re-writing (4.24) and  $[\cdot]^{\dagger}$  represents the pseudoinverse operator [35]. We use the following definition for the pseudoinverse

$$\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{H} \left[ \boldsymbol{A} \boldsymbol{A}^{H} \right]^{-1}. \tag{4.54}$$

The performance of precoders with perfect CSI at the transmitter appears in Section 4.5.

### Precoding for Frequency Selective Channels without ISI with Partial Channel State Information

We use the same approach as in Chapter 3 to model the partial CSI as statistically correlated to the actual channel but do not impose any structure on the correlation and obtain the precoders to be used over frequency selective channels without ISI. The partial CSI is now represented as the set of matrices  $\{\hat{\mathcal{H}}^{(i)}\}$ . When partial CSI is available, the cost function formed will be the same as that given in (4.51), but an additional expectation is computed over the channel matrix. Thus the cost function with partial CSI  $J_p^{(a)}$  will be

$$J_{p}^{(a)} = E_{\boldsymbol{b},\boldsymbol{n}\{\boldsymbol{\mathcal{H}}^{(i)}\}} \left\{ \left\| \boldsymbol{\mathcal{D}}^{(a)}\boldsymbol{b} - \boldsymbol{\mathcal{H}}^{(a)}\boldsymbol{\mathcal{R}}\boldsymbol{\mathcal{M}}^{(a)}\boldsymbol{b} + \boldsymbol{n}^{(a)} \right\|^{2} \left| \{\hat{\boldsymbol{\mathcal{H}}}^{(i)}\} \right\}.$$
(4.55)

which is the same as that given in (D-2) with  $J^{(a)} = J_p^{(a)}$ ,  $M^{(a)} = \mathcal{M}^{(a)}$ ,  $D^{(a)} = \mathcal{D}^{(a)}$ ,  $H^{(a)} = \mathcal{H}^{(a)}$ ,  $\{\hat{H}^{(i)}\} = \{\hat{\mathcal{H}}^{(i)}\}$ ,  $R = \mathcal{R}$  and  $\tilde{n} = n^{(a)}$ . Similar to the approach followed in Chapter 3 for flat fading channels, we can obtain the precoders as functions of conditional channel correlation matrices

given as

$$\boldsymbol{C}_{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}^{(a)} \triangleq E_{\boldsymbol{\mathcal{H}}^{(a)}} \left\{ \boldsymbol{\mathcal{H}}^{(a)H} \boldsymbol{\mathcal{H}}^{(a)} \middle| \hat{\boldsymbol{\mathcal{H}}}^{(a)} \right\}$$
(4.56)

$$\boldsymbol{C}_{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{D}}}^{(a)} \triangleq E_{\{\boldsymbol{\mathcal{H}}^{(i)}\}} \left\{ \boldsymbol{\mathcal{H}}^{(a)H}\boldsymbol{\mathcal{D}}^{(a)} \middle| \hat{\boldsymbol{\mathcal{H}}}^{(a)}, \hat{\boldsymbol{\mathcal{D}}}^{(a)} \right\}.$$
(4.57)

We obtain the optimal  $\mathcal{M}^{(a)}$  as per the solution to (D-2) given in Appendix D as

$$\mathcal{M}^{(a)} = \left[ C_{\mathcal{H}\mathcal{H}}^{(a)} \mathcal{R} \right]^{\dagger} C_{\mathcal{H}\mathcal{D}}^{(a)}.$$
(4.58)

So far, zero-padding has been used to eliminate intersymbol interference. Section 4.4 addresses the design of precoders with the alternative approach in which we stack symbol vectors into a frame, inserting zeros between frames, thereby increasing the rate. Here, precoders are designed for optimizing over a frame and are burdened with the additional task of eliminating ISI.

## 4.4 Precoding for Frequency Selective Channels with Intersymbol Interference

In this section we consider precoding for achieving higher rates by eliminating the zero-padding between symbol vectors **b**. Zero-padding is done to eliminate ISI. When the total number of paths, i.e., the number of taps in the equivalent channel filter is L, and the time delay between successive paths is  $T_c$ , the total number of zero pads required are  $\lceil \frac{L}{N} \rceil$ . When the maximum number of paths assumed is N, ISI occurs among adjacent symbols and with zero-padding the rate reduces to half, since each symbol vector is followed by a zero-pad. On the other hand, by stacking symbol vectors into frames we can have almost full rate provided the frame length, measured in terms of symbol vectors, is sufficiently large. The tradeoff involved though, will be increased complexity and importantly, the introduction of ISI. In this section we will derive precoders which try to optimize the performance over an entire frame. Like precoders for channels without ISI, these also employ a pre-Rake type structure to exploit diversity. Based on the results from Section 4.3, we choose to jointly optimize the precoders to be used at a particular antenna.

The system model is similar to that specified in Section 4.2. However, since we are forming a frame of symbol vectors, we will now have to index them. We denote  $b_i \in \mathbb{C}^{K \times 1}$  as the symbol vector formed by stacking the  $i^{th}$  bit of each of the K users. We stack p symbol vectors to form

a frame denoted by  $\tilde{\boldsymbol{b}} = \begin{bmatrix} \boldsymbol{b}_1^T & \boldsymbol{b}_2^T & \dots & \boldsymbol{b}_p^T \end{bmatrix}^T$ . Zeros are now padded between frames instead of symbol vectors. We design precoders to optimize performance over this frame.

The choice of precoders to be used to generate the signal to be transmitted is interesting and strongly depends on the fast or slow fading nature of the channel. Since we are transmitting multiple symbol vectors, we could choose to use a different precoder for each symbol vector  $b_i$ . This would be required in case the channel varies during a frame interval and this variation is known *a priori*. However, if the channel is constant for the frame interval, then each symbol vector experiences the same structure of interference and hence, the same set of precoders should be used. For the former case perfect *a priori* knowledge is not possible and we can use the partial channel knowledge model as considered in Chapter 3. Our goal is to characterize the performance of linear precoding to see if any diversity gain and interference cancellation is achieved by using simple linear techniques. We can assume that the channel is constant for a frame duration, i.e., the channel is quasi-static fading, and is perfectly known to the transmitter. This is a special case and will in fact act as an upper bound on the performance achieved when the channel varies during the transmission of a frame. The slow, quasi-static fading assumption will imply the use of identical precoders for all symbol vectors. We begin by looking at the transmitted signal and the corresponding received signal.

The pre-Rake signal transmitted from antenna a for the entire frame is represented by  $\boldsymbol{x}^{(a)}$  and given as

$$\boldsymbol{x}^{(a)} = \alpha \sum_{i=1}^{p} \hat{\boldsymbol{S}}_{i} \boldsymbol{\mathcal{M}}^{(a)} \boldsymbol{b}_{i}, \qquad a = 1, \dots, M_{T},$$
(4.59)

where  $\mathcal{M}^{(a)}$  is the same as that given in (4.24), while

$$\hat{\boldsymbol{S}}_{i} = \begin{bmatrix} \tilde{\boldsymbol{S}}_{i}[0] & \tilde{\boldsymbol{S}}_{i}[1] & \cdots & \tilde{\boldsymbol{S}}_{i}[L-1] \end{bmatrix}.$$
(4.60)

Again, like in (4.19)  $\tilde{\mathbf{S}}_i[l] \in \mathbb{C}^{(pN+L-1)\times K}$ , is a simple extension of the spreading matrix  $\mathbf{S}$ , so as to correspond to the activation point of the corresponding precoder  $\mathbf{M}^{(a)}[l]$  during the transmission of symbol vector  $\mathbf{b}_i$ . However, due to the framing operation, it will have a size that is different than that given in (4.19). We can write it down as

$$\tilde{\boldsymbol{S}}_{i}[l] = \begin{bmatrix} \boldsymbol{0}_{K \times (i-1)N + (L-1-l)} & \boldsymbol{S}^{T} & \boldsymbol{0}_{K \times l + (i+1)N} \end{bmatrix}^{T}.$$
(4.61)

Similar to the case without ISI, each receiver starts match filtering from the time slot corresponding to the  $L^{th}$  time slot. In this case though, it continues to receive for the next pN time slots. To

write down the received signal model, for simplicity let us assume that the total number of paths is L < N. Extension to larger delays would be straightforward as indicated at the end of this section. The largest possible delay is  $NT_c = T_s$ , where  $T_s$  is the symbol duration. Thus, ISI terms appear only from the adjacent symbol vectors, either due to channel memory, or due to the overlap occurring in transmissions of adjacent symbol vectors to achieve the pre-Rake combining effect. Observe that the ISI encountered by symbol vector  $\mathbf{b}_i$  is only from the adjacent symbol vectors  $\mathbf{b}_{i-1}$  and  $\mathbf{b}_{i+1}$ . In the absence of noise, the signal received by receiver k due to transmission from antenna a,  $\mathbf{r}_{k_i}^{(a)} \in \mathbb{C}^{N \times 1}$ , during the reception interval corresponding to the transmission of symbol vector  $\mathbf{b}_i$ , which here will be L + (i-1)N : L + iN - 1,

$$\mathbf{r}_{k_{i}}^{(a)} = \underbrace{\sum_{l=0}^{L-1} h_{k}^{(a)}[l] \mathbf{S} \mathbf{M}^{(a)}[l] \mathbf{b}_{i}}_{\text{Desired Signal+MAI}} + \underbrace{\sum_{j=0}^{L-1} \sum_{\substack{j\neq l\\l=0}}^{L-1} h_{k}^{(a)}[j] \mathbf{S}[j-l] \mathbf{M}^{(a)}[l] \mathbf{b}_{i}}_{\text{ICI}} + \underbrace{\sum_{j=1}^{L-1} \sum_{l=0}^{L-2} h_{k}^{(a)}[j] \mathbf{S}[-N+j-l] \mathbf{M}^{(a)}[l] \mathbf{b}_{i-1}}_{\text{ISI}} + \underbrace{\sum_{j=0}^{L-2} \sum_{l=1}^{L-1} h_{k}^{(a)}[j] \mathbf{S}[N+j-l] \mathbf{M}^{(a)}[l] \mathbf{b}_{i+1}}_{\text{ISI}},$$

$$(4.62)$$

where S[l] is again a vertically shifted version of the spreading matrix as indicated in Section 4.3.1 and in (E-21). The first and second terms in (4.62) are the same as those in (4.43), and correspond to the case without ISI. The third and fourth terms above represent the ISI. Note the limits on j and l in these terms. These arise from the fact that the ISI due to the previous symbol vector arises from channel memory, and that due to the next symbol vector arises from the pre-Rake transmission. Interference due to signals transmitted for previous symbol vector  $\mathbf{b}_{i-1}$  will appear from all but the shortest path  $h_k^{(a)}[0]$ , and hence, the lower limit on j in the third term. The signal transmitted for the previous symbol vector from the precoder  $\mathbf{M}^{(a)}[L-1]$ , will no longer remain in channel memory during the corresponding reception interval for  $\mathbf{b}_i$ . Consequently, no interference will appear from the precoder corresponding to the longest path and hence the upper limit on lin the third term. Interference from the signals transmitted for the next symbol vector  $\mathbf{b}_{i+1}$ , will appear from all paths but for the longest path  $h_k^{(a)}[L-1]$ , and hence, the upper limit on j in the fourth term. Due to the pre-Rake mechanism, each receiver is synchronized with the transmission from the precoder for the shortest path,  $\mathbf{M}^{(a)}[0]$ . Hence, it will not cause any interference and implies the limit on l in the fourth term.

The net noise-free received signal during the reception for  $\mathbf{b}_i$ ,  $\mathbf{r}_{k_i} = \sum_{a=1}^{M_T} \mathbf{r}_{k_i}^{(a)}$ , is match filtered with respect to the spreading sequence  $\mathbf{s}_k$ . In a manner similar to (4.30), (4.30) and (4.45), the noise-free decision statistic for the  $k^{th}$  bit of  $\mathbf{b}_i$ ,  $y_k[i]$ , formed by match filtering the signal  $\mathbf{r}_{k_i}$  with respect to the spreading sequence  $\mathbf{s}_k$  can be, can be decomposed in terms of statistics formed due to transmission from antennas  $a = 1, \dots, M_T$ . This noise-free decision statistic  $y_k^{(a)'}[i]$ , formed at user k due to the transmission from antenna a is given as

$$y_{k}^{(a)'}[i] = \sum_{j=1}^{L-1} \sum_{l=0}^{L-2} h_{k}^{(a)}[j] \boldsymbol{s}_{k}^{H} \boldsymbol{S}[-N+j-l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}_{i-1} + \sum_{j=0}^{L-1} \sum_{l=0}^{L-1} h_{k}^{(a)}[j] \boldsymbol{s}_{k}^{H} \boldsymbol{S}[j-l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}_{i} + \sum_{j=0}^{L-2} \sum_{l=1}^{L-1} h_{k}^{(a)}[j] \boldsymbol{s}_{k}^{H} \boldsymbol{S}[N+j-l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}_{i+1}.$$

$$(4.63)$$

Similar to  $\boldsymbol{y}^{(a)'}$  in (4.46), we now form the vector  $\boldsymbol{y}_i^{(a)'} = [y_1^{(a)'}[i], y_2^{(a)'}[i], \cdots, y_K^{(a)'}[i]]^T$ , where we have stacked the noise-free decision statistics for the  $i^{th}$  symbol generated at the K receivers. This will be given by

$$\boldsymbol{y}_{i}^{(a)'} = \begin{bmatrix} \boldsymbol{\mathcal{H}}_{-1}^{(a)} \boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)} \boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)} \boldsymbol{\mathcal{R}}_{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(a)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{M}}^{(a)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{M}}^{(a)} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{i-1} \\ \boldsymbol{b}_{i} \\ \boldsymbol{b}_{i+1} \end{bmatrix}, \quad (4.64)$$

where  $\mathcal{H}_0^{(a)}$  and  $\mathcal{R}_0$  are the same as  $\mathcal{H}^{(a)}$  and  $\mathcal{R}$  given in (4.46), respectively, and are used to represent the ISI free terms. The other unspecified matrices in the (4.64) are defined as follows:

$$\mathcal{H}_{-1}^{(a)} \triangleq \begin{bmatrix} \mathbf{0} & \mathbf{H}^{(a)}[1] & \cdots & \mathbf{H}^{(a)}[L-1] \end{bmatrix}_{K \times KL}$$
(4.65)

$$\mathcal{H}_{1}^{(a)} \triangleq \begin{bmatrix} \mathbf{H}^{(a)}[0] & \cdots & \mathbf{H}^{(a)}[L-2] & \mathbf{0} \end{bmatrix}_{K \times KL}$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} \end{bmatrix}$$
(4.66)

$$\boldsymbol{\mathcal{R}}_{-1} \triangleq \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{R}_{-N+1} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \mathbf{0} & \mathbf{0} \end{bmatrix}$$
(4.67)

$$\boldsymbol{\mathcal{R}}_{1} \triangleq \begin{bmatrix} \mathbf{0} & \mathbf{R}_{-N+L-1} & \cdots & \mathbf{R}_{-N+1} & \mathbf{0} \end{bmatrix}_{KL \times KL} \\ \begin{bmatrix} \mathbf{0} & \mathbf{R}_{N-1} & \cdots & \mathbf{R}_{N-L+1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \mathbf{0} & \mathbf{R}_{N-1} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}_{KL \times KL}$$
(4.68)

where  $\boldsymbol{H}^{(a)}[\cdot]$  and  $\boldsymbol{R}_j$  are the same as defined above in (4.47) and (4.48), respectively. They are present because of the ISI terms. The interference terms from the previous symbol appear through

the L-1 paths apart from the shortest path. Matrices  $\mathcal{R}_{-1}$  and  $\mathcal{H}_{-1}^{(a)}$  are used to represent the resulting interference. Column *i* from the matrix  $\mathcal{R}_{-1}$  is used to represent the ISI from transmission due to precoder  $M^{(a)}[i]$ . The elements of rows j > i for a particular column *i* are used to represent the interference resulting from the transmission of signals transmitted from precoder  $M^{(a)}[i]$ , over path *j*. As mentioned earlier, no interference occurs from the signal transmitted from the  $L^{th}$ precoder for the previous symbol. This is represented by the all zero column in the matrix  $\mathcal{R}_{-1}$ . On the other hand, the terms  $\mathcal{R}_1$  and  $\mathcal{H}_1^{(a)}$  are used to represent the ISI terms from the next symbol received along all paths, but for the longest one. Again, the *i*<sup>th</sup> column of  $\mathcal{R}_1$  is used to represent the interference due to transmission from precoder  $M^{(a)}[i]$ . The elements of rows j > ifor a particular column *i* are used to represent the interference resulting from the transmission of signal transmitted from precoder  $M^{(a)}[i]$ , over path *j*. The all zero column of  $\mathcal{R}_1$  indicates the absence of any interference from the signal transmitted from precoder  $M^{(a)}[0]$  for the next symbol. The formulation of these matrices is similar to that for (4.46). We perform the operations as shown in Appendix F on each of the summation terms in (4.63).

Since we are transmitting over frames, we wish to design precoders which optimize over the entire frame. Hence, we form a cost function over the entire frame. For this we stack the decision statistics  $\boldsymbol{y}_i^{(a)} = \boldsymbol{y}_i^{(a)'} + \boldsymbol{n}_i^{(a)}, \ \forall i = 1 : p$ , where  $\boldsymbol{n}_i^{(a)}$  represents the noise term similar to  $\boldsymbol{n}^{(a)}$  in (4.46), formed at all the receivers after match filtering during the pN reception slots to form,  $\tilde{\boldsymbol{y}}^{(a)} = \begin{bmatrix} \boldsymbol{y}_0^{(a)} & \boldsymbol{y}_1^{(a)} & \dots & \boldsymbol{y}_{p-1}^{(a)} \end{bmatrix}^T$ , which can be expressed as

$$\tilde{\boldsymbol{y}}^{(a)} = \tilde{\boldsymbol{\mathcal{H}}}^{(a)} \tilde{\boldsymbol{\mathcal{M}}}^{(a)} \tilde{\boldsymbol{b}} + \boldsymbol{n}^{(a)}, \qquad (4.69)$$

where we use the following definitions,

$$\tilde{\boldsymbol{\mathcal{H}}}^{(a)} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)}\boldsymbol{\mathcal{R}}_{1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{\mathcal{H}}_{-1}^{(a)}\boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)}\boldsymbol{\mathcal{R}}_{1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \dots & \boldsymbol{\mathcal{H}}_{-1}^{(a)}\boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)}\boldsymbol{\mathcal{R}}_{1} \\ \boldsymbol{0} & \dots & \dots & \boldsymbol{0} & \boldsymbol{\mathcal{H}}_{-1}^{(a)}\boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} \end{bmatrix}_{pK \times pKL}, (4.70)$$

$$\tilde{\boldsymbol{\mathcal{M}}}^{(a)} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(a)} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{M}}^{(a)} \end{bmatrix}_{pKL \times pKL}$$
(4.71)

The noise term  $\tilde{\boldsymbol{n}}^{(a)} = [\boldsymbol{n}_1^{(a)}, \cdots, \boldsymbol{n}_p^{(a)}]$ , where each component  $\boldsymbol{n}_i^{(a)}, i = 1, \cdots, p$ , has a distribution similar to that of the noise term in (4.46), and each term  $\boldsymbol{n}_i^{(a)}$  is independent of all other terms  $\boldsymbol{n}_j^{(a)}, i \neq j$ . To form the cost function, we define

$$\tilde{\boldsymbol{\mathcal{D}}}^{(a)} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{D}}^{(a)} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{D}}^{(a)} \end{bmatrix}_{pK \times pK}, \qquad (4.72)$$

where  $\mathcal{D}^{(a)}$  is as defined in (4.50). This definition motivated by the fact that we want the vector of decision statistics  $\boldsymbol{y}_i^{(a)}$  formed for each symbol vector  $\boldsymbol{b}_i$  to be of the form  $\mathcal{D}^{(a)}\boldsymbol{b}_i$  in order to achieve identical performance over the entire frame. Next we form the MMSE cost function  $J^{(a)}$  as

$$J^{(a)} = E_{\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}^{(a)}} \left\{ \left\| \tilde{\boldsymbol{\mathcal{D}}}^{(a)} \tilde{\boldsymbol{b}} - \tilde{\boldsymbol{\mathcal{H}}}^{(a)} \tilde{\boldsymbol{\mathcal{M}}}^{(a)} \tilde{\boldsymbol{b}} + \tilde{\boldsymbol{n}}^{(a)} \right\|^2 \right\}.$$
(4.73)

The solution for this cost function is

$$\mathcal{M}^{(a)} = \begin{bmatrix} \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{H}_0^{(a)} \mathcal{R}_0 \\ + \frac{(p-1)}{p} \mathcal{R}_{-1}^H \mathcal{H}_{-1}^{(a)H} \mathcal{H}_{-1}^{(a)} \mathcal{R}_{-1} \\ + \frac{(p-1)}{p} \mathcal{R}_1^H \mathcal{H}_1^{(a)H} \mathcal{H}_1^{(a)} \mathcal{R}_1 \end{bmatrix}^{\dagger} \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{D}^{(a)}.$$
(4.74)

The solution is derived in Appendix F.2. The optimal precoders can be obtained from  $\mathcal{M}^{(a)}$  using (4.53).

Note that in deriving the solution given in (4.74) we assumed that the maximum number of paths or the order of the channel to be L < N. Even if we assume more paths the solution will have a similar form. Observe that the off diagonal blocks of the matrix  $\tilde{\mathcal{H}}^{(a)}$  in (4.70) correspond to the ISI terms. Specifically, the off diagonal blocks of the  $i^{th}$  column represent the ISI occurring from transmission of the symbol vector  $\mathbf{b}_i$ . The elements on top of the diagonal block model the ISI resulting from the pre-Rake mechanism while those below model the ISI due to channel memory. With L > N let  $I = \lceil \frac{L}{N} \rceil < (p-1)/2$ . Then I blocks on either side of the diagonal of the  $i^{th}$  column will model the interference due to the symbol vector  $\mathbf{b}_i$ . The precoders can then be derived. The solution will have a form identical to that obtained in (4.74), with the only difference being that

110

there will be more summation terms in the term with the pseudoinverse. Simulation results with L < N, using the precoders with perfect CSI at transmitter given in (4.74) are presented in the following section.

### Precoding for Intersymbol Interference Channels with Partial Channel Information

Similar to the approach followed in Chapter 3 and Section 4.3.2, we model partial CSI as knowledge of some quantity that is statistically correlated to the channel. Again to obtain the precoders we need to compute the expectation in the cost function given by (4.73) over the actual channel. We form the cost function  $J_p^{(a)}$  for the case of partial CSI as

$$J_{p}^{(a)} = E_{\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{n}}, \tilde{\boldsymbol{\mathcal{H}}}^{(i)}} \left\{ \left\| \tilde{\boldsymbol{\mathcal{D}}}^{(a)} \tilde{\boldsymbol{b}} - \tilde{\boldsymbol{\mathcal{H}}}^{(a)} \tilde{\boldsymbol{\mathcal{M}}}^{(a)} \tilde{\boldsymbol{b}} + \tilde{\boldsymbol{n}} \right\|^{2} \left\| \{ \check{\boldsymbol{\mathcal{H}}}^{(i)} \} \right\}.$$
(4.75)

where the  $\{\check{\mathcal{H}}^{(i)}\}$  represents the partial CSI statistically correlated to  $\{\check{\mathcal{H}}^{(i)}\}$  Again the general solution can be given in terms of conditional channel correlation matrices given as

$$\boldsymbol{C}_{HH_{0}}^{(a)} \triangleq E_{\boldsymbol{\mathcal{H}}_{0}^{(a)}} \left\{ \boldsymbol{\mathcal{H}}_{0}^{(a)H} \boldsymbol{\mathcal{H}}_{0}^{(a)} \middle| \hat{\boldsymbol{\mathcal{H}}}_{0}^{(a)} \right\}$$
(4.76)

$$\boldsymbol{C}_{HH_{1}}^{(a)} \triangleq E_{\boldsymbol{\mathcal{H}}_{1}^{(a)}} \left\{ \boldsymbol{\mathcal{H}}_{1}^{(a)H} \boldsymbol{\mathcal{H}}_{1}^{(a)} \middle| \hat{\boldsymbol{\mathcal{H}}}_{1}^{(a)} \right\}$$
(4.77)

$$C_{HH_{-1}}^{(a)} \triangleq E_{\mathcal{H}_{-1}^{(a)}} \left\{ \mathcal{H}_{-1}^{(a)H} \mathcal{H}_{-1}^{(a)} \middle| \hat{\mathcal{H}}_{-1}^{(a)} \right\}$$

$$(4.78)$$

$$\boldsymbol{C}_{H_0D}^{(a)} \triangleq E_{\{\boldsymbol{\mathcal{H}}_0^{(i)}\}} \left\{ \boldsymbol{\mathcal{H}}_0^{(a)H} \boldsymbol{\mathcal{D}}^{(a)} \middle| \{ \hat{\boldsymbol{\mathcal{H}}}_0^{(i)} \} \right\}.$$
(4.79)

where  $\{\hat{\mathcal{H}}_{0}^{(i)}\}$  is the partial CSI statistically correlated to  $\{\mathcal{H}_{0}^{(i)}\}$  and the matrices  $\hat{\mathcal{H}}_{0}^{(i)}$ ,  $\hat{\mathcal{H}}_{1}^{(i)}$  and  $\hat{\mathcal{H}}_{-1}^{(i)}$  are statistically correlated to  $\mathcal{H}_{0}^{(i)}$ ,  $\mathcal{H}_{1}^{(i)}$  and  $\mathcal{H}_{-1}^{(i)}$  respectively. The precoding super-matrix  $\mathcal{M}^{(a)}$  as derived in Appendix F.2 will be given a as

$$\boldsymbol{\mathcal{M}}^{(a)} = \begin{pmatrix} \boldsymbol{\mathcal{R}}_{0}^{H} \boldsymbol{C}_{HH_{0}}^{(a)} \boldsymbol{\mathcal{R}}_{0} \\ -\frac{(p-1)}{p} \boldsymbol{\mathcal{R}}_{-1}^{H} \boldsymbol{C}_{HH_{-1}}^{(a)} \boldsymbol{\mathcal{R}}_{-1} \\ -\frac{(p-1)}{p} \boldsymbol{\mathcal{R}}_{1}^{H} \boldsymbol{C}_{HH_{1}}^{(a)} \boldsymbol{\mathcal{R}}_{1} \end{pmatrix}^{\mathsf{T}} \boldsymbol{\mathcal{R}}_{0}^{H} \boldsymbol{C}_{H_{0}D}^{(a)}.$$
(4.80)

The precoders to be used can be obtained using (4.53).

### 4.5 Performance of Precoding for Frequency Selective Channels

An analysis similar to that for precoding for flat fading channels with perfect channel information will require the derivation for the power scaling factor  $\alpha$ . The additional matrices characterizing the ICI and ISI, lead to highly complex terms, even under very stringent assumptions. Hence, similar to the case of partial channel information we resort to simulation to characterize the BER performance of the precoders derived. In this section we see the simulation performance when using the precoders derived for the cases mentioned in Section 4.3 and 4.4 above. We assume perfect channel knowledge at the transmitter with  $M_T = \{1, 2, 3\}$  transmit antennas and  $L = \{2, 3, 4, 6\}$ paths between the transmit antennas and the receivers. We use *m*-sequences, characterized by good cross-correlation properties for all the cases. We also present results where the performance is averaged over random codes for the case where there is no ISI. All simulations assume K = 10users and a spreading gain of N = 15. Figures 4.4 and 4.5 correspond to the case where there is no ISI. Note that we do not use pure random codes but rather use codes chosen from a set of random codes with the property that the maximum cross-correlation between any two non-identical codes is  $\rho_{ij} < 0.9$ . This constraint is used only to avoid rank problems arising from the use of highly correlated sequences. The figures indicate that full diversity gain, defined by the slope of the biterror-rate curve, is achieved. Interference is completely eliminated at the cost of a right shift in the bit-error rate performance curves.

Figures 4.6 and 4.7 indicate the performance achieved when precoding is used for frequency selective channels with ISI. For channels affected by ISI the performance degrades as the number of paths increases. No diversity gain is achieved for the case where 1 transmit antenna is used and the number of paths is 6. To justify the degradation for ISI channels we look at the noise-free decision statistic given in (4.64)

$$\boldsymbol{y}_{i}^{(a)'} = \begin{bmatrix} \boldsymbol{\mathcal{H}}_{-1}^{(a)} \boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)} \boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)} \boldsymbol{\mathcal{R}}_{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{M}}^{(a)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\mathcal{M}}^{(a)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\mathcal{M}}^{(a)} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{i-1} \\ \boldsymbol{b}_{i} \\ \boldsymbol{b}_{i+1} \end{bmatrix}, \quad (4.81)$$
$$= \boldsymbol{\mathcal{H}}_{-1}^{(a)} \boldsymbol{\mathcal{R}}_{-1} \boldsymbol{\mathcal{M}}^{(a)} \boldsymbol{b}_{i-1} + \boldsymbol{\mathcal{H}}_{0}^{(a)} \boldsymbol{\mathcal{R}}_{0} \boldsymbol{\mathcal{M}}^{(a)} \boldsymbol{b}_{i} + \boldsymbol{\mathcal{H}}_{1}^{(a)} \boldsymbol{\mathcal{R}}_{1} \boldsymbol{\mathcal{M}}^{(a)} \boldsymbol{b}_{i+1}$$

resulting from the use of the precoders as given in (4.74)

$$\boldsymbol{\mathcal{M}}^{(a)} = \begin{bmatrix} \boldsymbol{\mathcal{R}}_{0}^{H} \boldsymbol{\mathcal{H}}_{0}^{(a)H} \boldsymbol{\mathcal{H}}_{0}^{(a)} \boldsymbol{\mathcal{R}}_{0} \\ + \frac{(p-1)}{p} \boldsymbol{\mathcal{R}}_{-1}^{H} \boldsymbol{\mathcal{H}}_{-1}^{(a)H} \boldsymbol{\mathcal{H}}_{-1}^{(a)} \boldsymbol{\mathcal{R}}_{-1} \\ + \frac{(p-1)}{p} \boldsymbol{\mathcal{R}}_{1}^{H} \boldsymbol{\mathcal{H}}_{1}^{(a)H} \boldsymbol{\mathcal{H}}_{1}^{(a)} \boldsymbol{\mathcal{R}}_{1} \end{bmatrix}^{\dagger} \boldsymbol{\mathcal{R}}_{0}^{H} \boldsymbol{\mathcal{H}}_{0}^{(a)H} \boldsymbol{\mathcal{D}}^{(a)}$$

Observe from (4.68) and (4.68) that the matrices  $\mathcal{R}_1$  and  $\mathcal{R}_{-1}$  are sparse when the number of interfering paths is small compared to the length of the spreading sequence. As a result, when the

number of paths is less, the pseudoinverse

$$\begin{bmatrix} \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{H}_0^{(a)} \mathcal{R}_0 \\ + \frac{(p-1)}{p} \mathcal{R}_{-1}^H \mathcal{H}_{-1}^{(a)H} \mathcal{H}_{-1}^{(a)} \mathcal{R}_{-1} \\ + \frac{(p-1)}{p} \mathcal{R}_1^H \mathcal{H}_1^{(a)H} \mathcal{H}_1^{(a)} \mathcal{R}_1 \end{bmatrix}^{\dagger} \approx \begin{bmatrix} \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{H}_0^{(a)} \mathcal{R}_0 \end{bmatrix}^{\dagger}.$$

Using this approximation to the precoder the noise-free decisions statistics formed can be obtained from (4.81) as

$$\mathcal{H}_{0}^{(a)}\mathcal{R}_{0}\mathcal{M}^{(a)}\boldsymbol{b}_{i} \approx \mathcal{H}_{0}^{(a)}\mathcal{R}_{0}\left[\mathcal{R}_{0}^{H}\mathcal{H}_{0}^{(a)H}\mathcal{H}_{0}^{(a)}\mathcal{R}_{0}\right]^{\dagger}\mathcal{R}_{0}^{H}\mathcal{H}_{0}^{(a)H}\mathcal{D}^{(a)}\boldsymbol{b}_{i} \qquad (4.82)$$

$$\approx \mathcal{D}^{(a)} \boldsymbol{b}_i,$$
 (4.83)

where we go from step (4.82) to using (4.83) using the eigenvalue decomposition of  $\mathcal{H}_0^{(a)}\mathcal{R}_0$ . As we observe the channel is completely diagonalized since  $\mathcal{D}^{(a)}$  is diagonal in nature and interference is eliminated. However, this is not the case when the number of interfering paths is large and the matrices  $\mathcal{R}_1$  and  $\mathcal{R}_{-1}$  are not sparse anymore. As a result, as the number of paths increases interference shows up, limiting the system performance. However, it is important to note that with fewer paths, interference does not affect the performance and BERs as low as  $10^{-8}$  are easily achievable with almost full diversity as indicated by Figures 4.6 and 4.7.



Figure 4.4: Performance of Precoding for Frequency Selective Channels without ISI for random sequences such that  $|\rho_{ij}| < 0.9$  and *m*-sequences for  $M_T = 2, 1$  transmit antennas, L = 2, 4 paths, N = 15 and K = 10 users. The performance of a single user system exploiting full 4 path diversity is plotted for comparison.



Figure 4.5: Performance of Precoding for Frequency Selective Channels without ISI for random sequences such that  $|\rho_{ij}| < 0.9$  and *m*-sequences for  $M_T = 1, 2, 3$  transmit antennas, L = 2, 3, 6 paths, N = 15 and K = 10 users. The performance of a single user system exploiting full 6 path diversity is plotted for comparison.



Figure 4.6: Performance of Precoding for Frequency Selective Channels with ISI and using *m*-sequences for  $M_T = 2, 1$  transmit antennas, L = 2, 4 paths, N = 15 and K = 10 users. The slopes corresponding to diversity gains achieved by a single user system with a total of L = 3, 4 paths are plotted for comparison.



Figure 4.7: Performance of Precoding for Frequency Selective Channels with ISI and using msequences for  $M_T = 1, 2, 3$  transmit antennas, L = 2, 3, 6 paths, N = 15 and K = 10 users. The slopes corresponding to diversity gains achieved by a single user system with a total of L = 1, 5, 6paths are plotted for comparison.

### 4.6 Conclusion

We have derived precoders for MISO CDMA systems over frequency selective fading channels. With perfect CSI, our precoders for channels without ISI achieve full diversity gain while completely suppressing interference due to non-orthogonal spreading and interchip interference. The interference cancellation comes at the cost of a loss in the SNR performance. Precoders derived for channels with ISI can also achieve full diversity gains, but not in all scenarios even with perfect CSI. In particular, for the ISI case, diversity gain is reduced as the number of paths from the transmit antennas to the decentralized receive antennas increases. Some preliminary analysis indicates that as the number of paths increases, interference is not completely eliminated. However, the results obtained for precoding with 2 and 3 paths encourages the use of precoding in such scenarios. In fact when there are fewer paths, the total throughput can be optimized by using our precoders to achieve diversity and eliminating zero-padding to enhance the rate. Besides, non-linear techniques like pre-equalization for ISI elimination, not considered here, can also be used to eliminate interference while reducing the loss in both diversity as well as SNR for such channels.

# Summary, Conclusion and Future Research

In this thesis we have studied the performance of linear precoding techniques for multi-user multi-antenna communications. We assumed the downlink of a DS-CDMA system with nonorthogonal spreading. We constrained our receivers to be simple match filters and and so they have ultra-low complexity, and we have imposed a total transmit power constraint on the transmitter. We assumed perfect as well as partial channel information at the transmitter and derived interference cancelling precoders which simultaneously exploited multi-antenna diversity to improve the effective bit-error-rate (BER) performance of the system. Precoders were derived for flat fading channels and frequency selective channels with and without intersymbol interference (ISI). The precoders derived for flat fading channels and frequency selective channels without ISI completely eliminate interference and achieve full diversity gain when perfect CSI is available to the transmitter. Simulation results indicate that for frequency selective channels with ISI, extremely low BERs can be achieved with full diversity when the effective channel can be modelled as a 2 or 3 tap filter. Partial channel information was defined as a statistical quantity correlated to the channel and no structure imposed on this correlation to obtain general precoding solutions. Precoders derived for flat fading channels, when partial CSI is available in the form of old channel estimates, eliminate interference and exploit diversity provided the available CSI is highly correlated to the actual channel.

From this research we can conclude that linear precoding techniques can be used to reduce the

receiver complexity at the cost of enhanced complexity at the transmitter. Precoding solutions can be used to not only eliminate interference, but also to maximize the diversity benefit promised by the use of multiple antennas. For flat fading and frequency selective fading channel without ISI, simple linear techniques can eliminate interference while maximizing the diversity gain. The performance of precoding with partial CSI motivated the use of precoding in future wireless systems.

Future research can focus on the use of non-linear precoding techniques specifically for frequency selective channels with ISI. We have obtained a general solution for precoding with partial CSI at the transmitter. Specific cases could be applied. One practical example would be to consider the case where the available estimates are noisy and the channel is quasi-static fading. The effect of channel estimation was not considered and needs to be researched. A laboratory implementation of precoders derived could be performed as was done in [7]. The precoding schemes considered in this thesis assumed the users to have equal rate, or equivalently, SNR constraints and to have spreading codes of equal lengths. Precoding without these constraints could be considered. We had constrained our system model to be such that the receiver has a single antenna. Precoding techniques such that receivers have multiple antennas but no channel knowledge need to be investigated. Most of the results in this thesis were based on simulations. Advanced random matrix methods could be used to obtain analytical expressions for the BER achieved using precoding for frequency selective channels and for precoding with partial CSI. Precoding schemes for multi-user MIMO communications when CSI is available neither at the transmitter, nor at the receiver should be considered. These could use the recent research on differential space-time codes to exploit diversity.

# References

- [1] M. Loomis, "Improvement in telegraphing," U.S. Patent, July 1872.
- [2] W. Jakes, Microwave Mobile Communications, Wiley, 1974.
- [3] D. Reynolds, A. Høst Madsen, and X. Wang, "Adaptive transmitter precoding for time division duplex CDMA in fading multipath channels: Strategy and analysis," *EURASIP Jour. Applied Sig. Proc.*, vol. 2002, no. 12, pp. 1325–1334, Dec. 2002.
- [4] S. Georgoulis and D. Cruickshank, "Pre-equalization, transmitter precoding and joint transmission techniques for time division duplex CDMA," in Proc. 2001 Conf. on 3G Mobile Communication Technologies, London, UK, Mar. 2001, pp. 257–261.
- [5] N. Al-Dhahir, "Transmitter optimization for noisy ISI channels in the presence of crosstalk," *IEEE Trans. Sig. Proc.*, vol. 48, no. 3, pp. 907–911, Mar. 2000.
- [6] B. Rainbolt and S. Miller, "CDMA transmitter filtering for cellular overlay systems," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 290–297, Feb. 2000.
- [7] M. Brandt-Pearce and A. Dharap, "Transmitter-based multiuser interference rejection for the down-link of a wireless CDMA system in a multipath environment," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 407–417, Mar. 2000.
- [8] R. Esmailzadeh, E. Sourour, and M. Nakagawa, "Prerake diversity combining in time-division duplex CDMA mobile communications," *IEEE Trans. Vehicular Tech.*, vol. 48, no. 3, pp. 795–801, May 1999.
- B. Vojčić and W. Jang, "Transmitter precoding in synchronous multiuser communications," IEEE Trans. Commun., vol. 46, no. 10, pp. 1346–1355, October 1998.
- [10] G. Durgin, Space-Time Wireless Channels, Prentice Hall, Upper Saddle River, NJ, USA., 2002.
- [11] B. Baker and E. Copson, The Mathematical Theory of Huygen's Principle, Clarendon Press, Oxford, England, 1950.
- [12] T. Rappaport, Wireless Communications: Principles & Practice, 2nd ed., Prentice Hall, Upper Saddle River, NJ, USA., 2001.
- [13] A. Lewis, "Wordweb: English thesaurus and dictionary for Windows," http://wordweb.info/free/.

- [14] B. Sklar, "Rayleigh fading channels in mobile digital communication systems part I: Characterization," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 136–146, Sept. 1997.
- [15] B. Sklar, "Rayleigh fading channels in mobile digital communication systems part II: Mitigation," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 148–155, Sept. 1997.
- [16] M Valenti, Communication Theory, Course Notes, WVU, Spring 2003.
- [17] D. Reynolds, Wireless Communication Systems, Course Notes, WVU, Fall 2002.
- [18] D. Reynolds, Statistical Signal Processing, Course Notes, WVU, Spring 2004.
- [19] M. Valenti, Stochastic Systems Theory, Course Notes, WVU, Fall 2002.
- [20] D. Tse and P. Vishwanath, Fundamentals of Wireless Communication, draft, 2004.
- [21] G. Stüber, *Principles of Mobile Communication*, Kluwer Academic Publishers, Norwell, MA, USA., 1996.
- [22] J. Proakis, Digital Communications, 3rd ed., McGraw-Hill, New York, NY, USA., 1995.
- [23] Y. Okumura, E. Ohmuri, T. Kawano and K. Fukuda, "Field strength and its variability in VHF land mobile radio services," *Rev. of the ECL*, vol. 16, pp. 825–873, 1968.
- [24] COST 231 TD(91)109, "1800 MHz mobile net planning based on 900MHz measurements," 1991.
- [25] W. Lee, Mobile Communications Engineering, McGraw-Hill, 1985.
- [26] D. Cox, R. Murray and A. Norris, "800Mhz Attenuation measured in and around sub-urban houses," AT&T-Bell Labs Technical Jour., July-Aug 1984.
- [27] R. Hogg and A. Craig, Introduction to Mathematical Statistics, 5th ed. Prentice Hall, Upper Saddle River, NJ, 1995.
- [28] H. Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton, NJ, 1946.
- [29] A. Carlson, P. Crilly, and J. Rutledge, Communication Systems: An Introduction to Signals and Noise in Electronic Communication, 4th ed., McGraw-Hill, New York, NY, USA., 2002.
- [30] S. Verdú, Multiuser Detection, Cambridge, UK: Cambridge University Press, 1998.
- [31] M. Nakagami, "The m-distribution: A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. C. Hoffmann, Ed., New York, 1960, Pergamon.
- [32] H. Suzuki, "A statistical model for urban multipath propogation," *IEEE Trans. Commun.*, vol. 25, pp. 673–680, July 1977.
- [33] G. Kennedy and B. Davis, *Electronic Communication Systems*, McGraw-Hill: New York, 1992.
- [34] B. Sklar, Digital Communications Fundamentals and Applications, 2nd ed., Prentice Hall, Upper Saddle River, NJ, USA., 2001.

- [35] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, USA., 2000.
- [36] H. Poor, An Introduction to Signal Detection and Estimation, Springer-Verlag, 2nd edition, 1994.
- [37] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T-Bell Labs Internal Tech. Memo., June 1995.
- [38] S. Alamouti, "A simple transmit diversity technique for wireless communications," IEEE J. Select. Areas Commun., vol. 16, no. 8, pp. 1451 – 1458, Oct. 1998.
- [39] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. Journal*, vol. 1, pp. 41–59, Autumn 1996.
- [40] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49.
- [41] H. El Gamal, G. Caire and M. Damen, "Lattice coding and decoding achieve the optimal diversity-vs-multiplexing tradeoff of mimo channels," *IEEE Trans. Inform. Theory*, To appear 2004.
- [42] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis and H. Sampath, "Optimal designs for space-time precoders and decoders," *IEEE Trans. Sig. Proc.*, vol. 50, no. 5, pp. 1051–1064, May 2002.
- [43] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, NY, USA., 1991.
- [44] R. Fischer, C. Windpassinger, A. Lampe, J. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," in *Proc. of International Zurich Seminar on Broadband Communications*, Zurich, Switzerland, Feb. 2002.
- [45] C. Windpassinger, T. Vencel and R. Fischer, "Precoding and loading for BLAST-like systems," in *Proc. of IEEE International Conference on Communications*, Anchorage, AK, USA, May 2003.
- [46] M. Tomlinson, "A new automatic equalizer employing modulo arithmetic," *Electr. Lett.*, vol. 7, pp. 138–139, 1971.
- [47] H. Harashima and H. Miyakawa, "Matched transmission techniques for channel with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780, 1972.
- [48] ARIB, "Japan's proposal for candidate radio transmission technology on IMT-2000: W-CDMA," June, 1998.
- [49] http://www.arib.or.jp/IMT-2000.
- [50] H. Holma and A. Toskala, Eds., WCDMA for UMTS: Radio Access for Third Generation Mobile Communications, Wiley, New York, NY, USA., 2001.
- [51] http://www.umtsworld.com.

- [52] TIA TR45.5, "The CDMA2000 ITU-R RTT candidate submission," June, 1998.
- [53] http://www.3gpp2.org.
- [54] http://www.tdscdma-forum.org.
- [55] http://grouper.ieee.org/groups/802/11/
- [56] http://www.bluetooth.com/
- [57] P. Jung, P. Baier and A. Steil, "Advantages of CDMA and spread spectrum over FDMA and TDMA in cellular mobile radio applications," *IEEE Trans. Vehicular Tech.*, vol. 42, no. 3, pp. 357–364, Aug. 1998.
- [58] A. Viterbi, CDMA: Principles of Spread Spectrum Communications, Addison-Wesley, Reading, MA, USA., 1995.
- [59] R. Price and P. Green, "A communication technique for multipath channels," vol. 46, pp. 555–570, Mar 1958.
- [60] A. Oppenheim, A. Willsky with S. Nawab, Signals and Systems, 2nd ed. Prentice Hall, Upper Saddle River, NJ, 1996.
- [61] M. Petrich, "On the number of orthogonal signals which can be placed in a WT-product," SIAM Journal, pp. 936–940, Dec. 1963.
- [62] V. V'yugin, "Algorithmic complexity and stochastic properties of finite binary sequences," 2004, vol. 42, pp. 294–317, Oxford University Press, UK.
- [63] M. Rupf and J. Massey, "Optimum sequence multisets for synchronous code-division multipleaccess," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1261–1266, July 1994.
- [64] P. Viswanath, V. Anantharam, and D. Tse, "Optimal sequences, power control and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1968–1983, Sep. 1999.
- [65] T. Guess, "Optimal sequences for CDMA with decision-feedback receivers," IEEE Trans. Inform. Theory, vol. 49, no. 4, pp. 886–900, April 2003.
- [66] http://mathworld.wolfram.com/PrimitivePolynomial.html.
- [67] S. Miller, "Notes CDMA course: Texas A&M University," .
- [68] M. Simon, J. Omura, R. Scholtz and B. Levitt, Spread Spectrum Communications Handbook, McGraw-Hill: New York, 1994.
- [69] X. Wang and H. Poor, Wireless Communication Systems: Advanced Techniques for Signal Reception, Prentice Hall, Upper Saddle River, NJ, 2004.
- [70] D. George, R. Bowen and J. Storey, "An adaptive decision feedback equalizer," *IEEE Trans. Commun. Tech.*, vol. 19, pp. 281–293, 1971.
- [71] J. Salz, "Optimum mean-square decision feedback equalizer," Bell Labs. Tech. Journal, vol. 52, pp. 1341–1373, Oct. 1973.

- [72] M. Costa, "Writing on dirty paper," IEEE Trans. Inform. Theory, vol. 29, no. 3, pp. 439–441, May 1983.
- [73] P. Vishwanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplin-downlink duality," *IEEE Trans. Inform. Theory*, vol. 49., no. 8, pp. 1912–1921, Aug. 2003.
- [74] J. Cioffi and W. Yu, "Trellis precoding for the broadcast channel," in Proc. 2001 GLOBE-COM, San Antonio, TX, Nov. 2001, pp. 1344–1348.
- [75] P. Viswanath, D. Tse and L. Zheng, "Diversity-multiplexing tradeoff in multiple access channels," Available on-line at http://www.eecs.berkeley.edu/~dtse/TVZ03.pdf.
- [76] G. Durgin and T. Rappaport, "Theory of multipath shape factors for small-scale fading wireless channels," *IEEE Trans. Antennas Propagat.*, vol. 48, no. 5, pp. 682–693, May 2000.
- [77] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. 1993 International Conference on Communications* (ICC'93), Geneva, Switz., 1993, pp. 1630–1634.
- [78] N. Seshadri and J. Winters, "Two signaling schemes for improving the error performance of frequency-division duplex (FDD) transmission systems using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, no. 1, 1994.
- [79] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [80] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.
- [81] J. Winters, "Switched diversity with feedback for DPSK mobile radio systems," *IEEE Trans. Vehicular Tech.*, vol. 32, no. 1, pp. 134–150, Feb. 1983.
- [82] G. Povey and M. Nakagawa, "A review of time division duplex-CDMA techniques," in Proc. IEEE 5th International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'98), Sun City, South Africa, Sep. 1998, pp. 630–633.
- [83] N. Seshadri and J. Winters, "Two signaling schemes for improving the error performance of frequency-division duplex (FDD) transmission systems using transmitter antenna diversity," in *Proc. 1993 IEEE Vehicular Technology Conference (VTC'93)*, Secaucus, NJ, May 1993, pp. 508–511.
- [84] J. Guey, M. Fitz, M. Bell, and W. Kuo, "Signal design for transmitter diversity wireless communications over rayleigh fading channels," in *Proc. IEEE 1996 Vehicular Technology Conference (VTC '96)*, 1996, pp. 136–140.
- [85] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Vehicular Tech.*, vol. 41, no. 2, pp. 170–176, May. 1992.

- [86] B. Hochwald, T. Marzetta, and C. Papadias, "A transmitter diversity scheme for wideband CDMA systems based on space-time spreading," *IEEE J. Select. Areas Commun.*, vol. 19, no. 1, pp. 48–60, Jan. 2001.
- [87] R. Fischer and C. Windpassinger, "Improved mimo precoding for decentralized receivers resembling concepts from lattice reduction," in *Proc. 2003 GLOBECOM*, San Fransisco, CA, Dec. 2003.
- [88] R. Fischer, C. Windpassinger and J. Huber, "Lattice-reduction-aided broadcast precoding," in Proc. 5th International ITG Conference on Source and Channel Coding (SCC), Erlangen, Gernamy, Jan. 2004, pp. 403–408.
- [89] H. Yao and G. Wornell, "Lattice-reduction-aided detectors for mimo communication systems," in Proc. 2002 GLOBECOM, Taipei, Taiwan, Nov. 2002.
- [90] J. Roh and B. Rao, "Multiple antenna channels with partial channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 677–688, Mar. 2004.
- [91] E. Visotsky and U.Madhow, "Space-time transmit precoding with imperfect feedback," IEEE Trans. Inform. Theory, vol. 47, pp. 2632–2639, Sep. 2001.
- [92] S. Bhashyam, A. Sabharwal, B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 785–798, May 2002.
- [93] G. Caire and U. Mitra, "Structured multiuser channel estimation for block-synchronous DS-CDMA," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1605–1617, Sep 2001.
- [94] E. Ertin, U. Mitra and S. Siwamgostham, "Maximum-likelihood-based multipath channel estimation for code-division multiple-access systems," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 290–302, Feb. 2001.
- [95] S. Bensley and B. Aazhang, "Subspace-based channel estimation for code-division multipleaccess communication systems," *IEEE Trans. Commun.*, vol. COM-44, no. 8, pp. 1009–1020, Aug. 1996.
- [96] H. Liu and G. Xu, "A subspace method for signal waveform estimation in synchronous CDMA systems," *IEEE Trans. Commun.*, vol. COM-44, no. 10, pp. 1346–1354, Oct. 1996.
- [97] G. Jongren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal spacetime block coding," *IEEE Trans. Inform. Theory*, vol. 48, pp. 611–627, mAR. 2002.
- [98] B. Hughes, "Differential space-time modulation," IEEE Trans. Inform. Theory, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.
- [99] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," IEEE Trans. Commun., vol. 48, no. 12, pp. 2041–2052, Dec. 2000.
- [100] D. Reynolds, X. Wang and K. Modi, "Interference suppression and diversity exploitation for multi-antenna CDMA with ultra-low complexity receivers," submitted to *IEEE Trans. Sig. Proc.*, 2003. Pre-print available on-line at http://csee.wvu.edu/~reynolds/multiprecoding.pdf.
- [101] R. Clarke, "A statistical theory of mobile radio reception," Bell Labs. Tech. Journal, vol. 47, pp. 957–1000, 1968.

- [102] L. Scharf, Statistical Signal Processing: Detection estimation and Time Series Analysis, Addison Wesley, 1 edition, 1991.
- [103] Q. Zhang, "A decomposition technique for efficient generation of correlated nakagami fading channels," *IEEE J. Selected Areas Commun.*, vol. 18, no. 11, pp. 2385–2392, Nov. 2000.
- [104] R. Ertel and J. Reed, "Generation of two equal power correlated rayleigh fading envelopes," *IEEE Commun. Lett.*, vol. 2, pp. 276–278, Oct. 1998.
- [105] B. Natarajan, C. Nassar and V. Chandrashekhar, "Generation of correlated fading envelopes for spread spectrum applications," *IEEE Commun. Lett.*, vol. 4, no. 1, pp. 9–11, Jan 2000.
- [106] K. Miller, Complex Stochastic Processes, Addison-Wesley, 1974.
- [107] S. Guncavdi and A. Duel-Hallen, "A novel space-time pre-Rake transmitter diversity method for W-CDMA using long range prediction," in *Proc. 2001 Conference on Information Sciences* and Systems, Princeton, NJ, Mar. 2001, pp. 32–37.
- [108] D. Reynolds and K. Modi, "Precoding with partial channel information for CDMA with ultralow complexity receivers," in *Proc. 41st Annual Allerton Conference on Comm., Control and Comput.*, Monticello, IL, USA, Oct. 2003.
- [109] K. Modi, D. Reynolds and X. Wang, "Linear Precoding for Multi-Antenna CDMA over Multipath ISI Channels," Proc. 2004 Conference on Information Sciences and Systems, Princeton, NJ, Mar. 2004, pp. 1237–1241.
- [110] C. li, and X. Wang, "Performance comparisons of MIMO techniques with application to WCDMA systems," *EURASIP Jour. Applied Sig. Proc.*, vol. 2004, no. 5, pp. 649–661, May 2004.
- [111] J. Craig, "A new, simple and exact result result for calculating the probability of error for two-dimensional signal constellations," in *Proc. of MILCOM*, Boston, MA, 1991, pp. 25.5.1–25.5.5.
- [112] M. Simon, Probability Distributions Involving Gaussian Random Variables, A Handbook for Engineers and Scientists, Kluwer Academic Press, Norwell, MA, USA., 2002.

# A Gaussian Random Variables and Functions of Gaussian Random Variables

The density function of a *normalized* Gaussian random variable, say Y, which has mean  $\mu = 0$ and variance  $\sigma^2 = 1$ , is given by

$$p_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right). \tag{A-2}$$

A scalar random variable say, W, generated by applying a scalar transformation to Y, i.e.,  $W = \sigma_w Y$ ,  $\sigma_w \in \Re$  will have a density function given by

$$p_W(w) = \frac{1}{\sigma_w} p_Y\left(\frac{w}{\sigma_w}\right) \tag{A-3}$$

$$p_W(w) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{w^2}{2\sigma_w^2}\right).$$
 (A-4)

It will have mean 0 but the variance will now be  $\sigma_w^2$ . A new scalar random variable, say X, generated by shifting the random variable W, by  $\mu_x$ , i.e.,  $X = \mu_x + W = \mu_x + \sigma_x W$ , with  $\sigma_x = \sigma_w$ , will have a density function given by

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right).$$
 (A-5)

The cumulative distribution function (C.D.F.) of this density is given by

$$P_X(x) = P(X \le x)$$

$$= 1 - \int_x^\infty p_X(x') dx'$$

$$= 1 - \int_{\frac{x-\mu_x}{\sigma_x}}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

$$= 1 - Q\left(\frac{x-\mu_x}{\sigma_x}\right)$$

$$= Q\left(\frac{\mu_x - x}{\sigma_x}\right).$$
(A-6)

where the definition of the Q-function is explicit from the steps and given below in (A-15) and we have used the property given by (A-16). The characteristic function of a Gaussian p.d.f is given by

$$\psi(w) = E_x \left\{ \exp(jwx) \right\}$$
  
=  $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} + jwx\right) dx$  (A-7)  
=  $\exp\left(jw\mu_x - \frac{w^2\sigma_x^2}{2}\right).$ 

### A.1 Multidimensional Gaussian Distribution

If the vector  $\boldsymbol{x} \in \boldsymbol{X} \in \Re^n$  has a Gaussian distribution then

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} |R_{xx}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}_x)^T \boldsymbol{R}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{m}_x)\right), \quad (A-8)$$

where  $\mathbf{m}_x = E\{X\}$  and  $\mathbf{R}_{xx} = E\{(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T\}$  are the mean vector and covariance matrices respectively. When the elements in  $\mathbf{x}$  are identically and independently distributed (i.i.d.) with mean  $\mathbf{m}_x$  and variance  $\sigma_x^2$ , this reduces to

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi\sigma_x^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_x^2}(\boldsymbol{x}-\boldsymbol{m}_x)^T(\boldsymbol{x}-\boldsymbol{m}_x)\right).$$
(A-9)

#### A.2 Conditional Gaussian Distribution

To compute the expectations in (3.9) and (3.10) we use the conditional Gaussian distribution [102] given as

$$f_{Y|X}(y|x) = \frac{1}{(2\pi\sigma_{y|x}^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_{y|x}^2}(y-\hat{m}_{y|x})^2\right),\tag{A-10}$$

which is a Gaussian distribution with mean  $\hat{m}_{y|x}$  and variance  $\sigma_{y|x}^2$  given by

$$\hat{m}_{y|x} = m_y + \rho_{y|x} \frac{\sigma_y}{\sigma_x} (x - m_x)$$
(A-11)

$$\sigma_{y|x}^{2} = \sigma_{y}^{2}(1 - \rho_{y|x}^{2})$$
(A-12)

$$\rho_{y|x} = \frac{E\left\{(x - m_x)(y - m_y)\right\}}{\sigma_y \sigma_x}$$
(A-13)

where  $m_y = m_x$  and  $\sigma_x = \sigma_y = \sigma$  are the means and the variances of the Gaussian random variables X and Y.

### A.3 The Q-Function

The Q function has different forms. In this thesis we are concerned with the Gaussian Q-function and we will review the alternate forms, bounds and techniques to compute some integrals involving the Q-function.

Let  $X \sim \mathcal{N}(0,1)$  be a normalized Gaussian random variable. Then,

$$P\{X < x\} = Q(x) \tag{A-14}$$

where Q(x) is the Gaussian Q-function defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) \mathrm{dt}.$$
 (A-15)

This is the one sided tail probability of a normalized Gaussian random variable. The upper and lower limits of this integral present computational and analytical difficulties and hence an alternate form is highly desired. This is given as

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2 t}\right) dt$$
 (A-16)

and is also known as the Craig representation [111]. This representation is only valid for values of  $x \ge 0$ . For negative values of x, we could use the property that

$$Q(x) + Q(-x) = 1$$
 (A-17)

Other forms of the Q-function, the Marcum Q-function, and the Nuttal Q-function, are not considered here. They can be found in [112].

An upper bound on the Q-function based on Criag's expression is given by

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2 t}\right) dt$$
$$\leq \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2}\right) dt \qquad (A-18)$$
$$\Rightarrow Q(x) \leq \frac{1}{2} \exp\left(-\frac{x^2}{2}\right).$$

This is not a very tight bound and tighter bounds can be obtained from [112] and references therein. Note, however that for large values of x, the approximation is very close.

Many times we need to evaluate expectations of the Q-function, i.e., compute the expected value of a Q-function whose argument is a random variable, say Y. This involves integrating over a Q-function and its product with some distribution  $p_Y(y)$  i.e. compute

$$\int_{-\infty}^{\infty} p_Y(y) Q(f(y)) \mathrm{dy},\tag{A-19}$$

where f(y), is some function of Y. The most frequently encountered form in the study of communication systems is has  $f(y) = a\sqrt{y}$ . If y has a one-sided distribution (or a distribution symmetric about 0), computing such expectations can be done by using the Laplace transform of the density,  $\phi_Y(s) = \mathcal{L} \{ p_Y(y) \}.$  We then have

$$\int_{0}^{\infty} p_{Y}(y)Q(a\sqrt{y})dy = \frac{1}{\pi} \int_{0}^{\pi/2} \int_{0}^{\infty} p_{Y}(y) \exp\left(-\frac{a^{2}y}{2\sin^{2}t}\right) dydt$$
$$= \frac{1}{\pi} \int_{0}^{\pi/2} \left[\mathcal{L}\left\{p_{Y}(y)\right\}\right]_{s=\frac{a^{2}}{2\sin^{2}t}} dt \qquad (A-20)$$
$$= \frac{1}{\pi} \int_{0}^{\pi/2} \phi_{Y}\left(\frac{a^{2}}{2\sin^{2}t}\right) dt,$$

where we have used the Craig representation of the Q-function. One of the integrals we see often when dealing with wireless systems is

$$\int_0^\infty x^{2m-1} \exp\left(\frac{-mb^2 x^2}{2}\right) Q(ax) dx = \frac{2^{m-2}}{(mb^2)^m} \left[1 - \sqrt{\frac{a^2}{a^2 + mb^2}}\right] \sum_{k=0}^{m-1} \binom{2k}{k} \left(\frac{mb^2}{4(a^2 + mb^2)}\right)^k (A-21)$$

where the closed form is available from [112].

### A.4 Rayleigh Random Variable

A generalized Rayleigh distribution results from the norm of a Gaussian random vector with zero mean i.i.d. entries. Let  $\boldsymbol{x} \in X \sim \mathcal{N}(\boldsymbol{0}, \sigma_h^2 \boldsymbol{I}_n)$ , then the norm of X given by

$$R = \boldsymbol{x}^T \boldsymbol{x} = \sqrt{\sum_{i=1}^n x^2[i]}$$
(A-22)

has a generalized Rayleigh distribution of order n. For n = 2m the density function is given by

$$p_R(r) = \frac{2r^{2m-1}}{(2\sigma_h^2)^m(m-1)!} \exp\left(-\frac{r^2}{2\sigma_h^2}\right), \qquad r \ge 0.$$
 (A-23)

In the case of a wireless Rayleigh fading channel a single path corresponds to a Rayleigh distribution of order n = 2, and the density function is given by

$$p_R(r) = \frac{r}{\sigma_h^2} \exp\left(-\frac{r^2}{2\sigma_h^2}\right), \qquad r \ge 0.$$
(A-24)

The square of a generalized Rayleigh random variable  $Y = R^2$ , or the squared norm of a Gaussian vector with n zero-mean i.i.d. entries is called a *central Chi-square* distribution with n degrees of freedom. For n = 2m, the density function of a central chi-square random variable is given as

$$p_Y(y) = \frac{1}{2\sigma_h^2 \Gamma(m)} \left(\frac{y}{2\sigma_h^2}\right)^{m-1} \exp\left(-\frac{y}{2\sigma_h^2}\right), \qquad y \ge 0$$
(A-25)

and for n = 2 the density function is given by

$$p_Y(y) = \frac{1}{2\sigma_h^2} \exp\left(-\frac{y}{2\sigma_h^2}\right), \qquad y \ge 0.$$
 (A-26)

Like a Rayleigh random variable, a Rician random variable also results from the norm of a Gaussian random vector with i.i.d. entries, but in this case the mean of the elements is non-zero. The non-zero mean corresponds to the presence of a *specular*, or line of sight, component in the received signal in case of a wireless channel. The corresponding squared random variable is called a *non-central Chi-square* distribution [22].

### A.5 Expectation of a *Q*-function with a First-Order Rayleigh Argument

We will frequently come across the evaluation of the following expectation

$$E_r\{Q(ar)\},\tag{A-27}$$

where R is a generalized Rayleigh distribution of order n = 2m and a is some constant. Let us first derive an upper bound on this expectation using a bound on the Q-function given by (A-18).

$$E_{r}\{Q(ar)\} \leq E_{r}\left\{\exp\left(-\frac{a^{2}r^{2}}{2}\right)\right\}$$

$$\leq \int_{0}^{\infty} \frac{2r^{2m-1}}{(2\sigma_{h}^{2})^{m}(m-1)!} \exp\left(-\frac{r^{2}}{2\sigma_{h}^{2}}\right) \exp\left(-\frac{a^{2}r^{2}}{2}\right) dr$$

$$\leq \int_{0}^{\infty} \frac{2r^{2m-1}}{(2\sigma_{h}^{2})^{m}(m-1)!} \exp\left(-\frac{r^{2}}{2}\left(\frac{1}{\sigma_{h}^{2}}+a^{2}\right)\right) dr$$

$$\leq \frac{1}{(1+\sigma_{h}^{2}a^{2})^{m}} \int_{0}^{\infty} \frac{2r^{2m-1}}{\left(2\frac{\sigma_{h}^{2}}{1+\sigma_{h}^{2}a^{2}}\right)^{m}(m-1)!} \exp\left(-\frac{r^{2}}{2}\left(\frac{1}{\frac{\sigma_{h}^{2}}{1+\sigma_{h}^{2}a^{2}}}\right)\right) dr$$

$$E_{r}\left\{Q(ar)\right\} \leq \frac{1}{\left(1+\sigma_{h}^{2}a^{2}\right)^{n/2}}.$$
(A-28)

In getting to the last step we observe that the integral in the previous step corresponds to a chi-square distribution. This bound is use very often in the analysis of communication systems involving diversity. We use this bound in Chapter 1 and Chapter 2 very frequently. Specifically we use it in getting (1.61) and (2.58).

We can evaluate the expectation,

$$E_r \{Q(ar)\} = \int_0^\infty \frac{2r^{2m-1}}{(2\sigma_h^2)^m (m-1)!} \exp\left(-\frac{r^2}{2\sigma_h^2}\right) Q(ar) dr \quad .$$
(A-29)

Using  $b^2 = \frac{1}{\sigma_h^2 m}$  in the integral (A-21), we get

$$E_r \left\{ Q(ar) \right\} = \frac{1}{2(n/2 - 1)!} \left[ 1 - \sqrt{\frac{a^2 \sigma_h^2}{a^2 \sigma_h^2 + 1}} \right] \sum_{k=0}^{n/2 - 1} {\binom{2k}{k}} \left( \frac{1}{4(a^2 \sigma_h^2 + 1)} \right)^k.$$
(A-30)

We can in fact obtain a closed form solution using the Laplace Transform technique mentioned above. Observe that we can write

$$E_r \left\{ Q(ar) \right\} = E_y \left\{ Q(a\sqrt{y}) \right\},\tag{A-31}$$

where y is a chi-square random variable of degree n. To find the Laplace transform of the density of a chi-square random variable with n = 2m degrees of freedom we use the fact that

$$L\{t^m \exp(-bt)\} = (-1)^m \frac{d^m}{\mathrm{ds}^m} \frac{1}{s+b} = \frac{m!}{(s+b)^{m+1}}, \qquad m \in \mathbb{Z}^+.$$
 (A-32)

The Laplace transform of the chi-square density will then be

$$\phi_Y(s) = L \left\{ \frac{1}{2\sigma_h^2 \Gamma(m)} \left( \frac{y}{2\sigma_h^2} \right)^{m-1} \exp\left(-\frac{y}{2\sigma_h^2}\right) \right\}$$

$$= \frac{1}{(2\sigma_h^2)^m \Gamma(m)} L \left\{ (y)^{m-1} \exp\left(-\frac{y}{2\sigma_h^2}\right) \right\}$$

$$= \frac{1}{(2\sigma_h^2)^m} \left( \frac{1}{s + \frac{1}{2\sigma_h^2}} \right)^m, \quad \Gamma(m) = (m-1)!$$

$$= \left( \frac{1}{2\sigma_h^2 s + 1} \right)^m.$$
(A-33)

Then, we can compute the above expectation as

$$E_Y \{Q(a\sqrt{y})\} = \int_0^\infty p_Y(y)Q(a\sqrt{y})dy$$
  
=  $\frac{1}{\pi} \int_0^{\pi/2} \phi_Y\left(\frac{a^2}{2\sin^2 t}\right) dt$   
=  $\frac{1}{\pi} \int_0^{\pi/2} \left(\frac{\sin^2 t}{\sigma_h^2 a^2 + \sin^2 t}\right)^m dt.$  (A-34)

Since *m* is an integer, it will result into the form given by (A-30). We use (A-30) in Chapter 2, (2.59) and to give the average bit-error-rate in (1.74) where n = 2L and  $a = \sqrt{\frac{2\mathcal{E}_b}{N_0}}$ .

From (A-28) and (A-30) observe that

$$\frac{1}{2(n/2-1)!} \left[ 1 - \sqrt{\frac{a^2 \sigma_h^2}{a^2 \sigma_h^2 + 1}} \right] \sum_{k=0}^{n/2-1} {\binom{2k}{k}} \left( \frac{1}{4(a^2 \sigma_h^2 + 1)} \right)^k \le \frac{1}{\left(1 + \sigma_h^2 a^2\right)^{n/2}}, \tag{A-35}$$

since the former is an upper bound.

### **B** Some properties of a Hermitian Matrix

A Hermitian matrix  $\mathbf{R} \in \mathbb{C}^{K \times K}$  is similar to a diagonal matrix and has an eigenvalue decomposition which is given as [35, 18]

$$\boldsymbol{R} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^H \tag{B-2}$$

where

$$\Lambda = \operatorname{diag} \left\{ \lambda_1, \cdots, \lambda_K \right\} \tag{B-3}$$

are the eigenvalues of the matrix  $\boldsymbol{R}$ , and

$$\boldsymbol{U} = \left[ \begin{array}{ccc} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_K \end{array} \right] \tag{B-4}$$

is the orthogonal matrix formed by are the corresponding eigenvectors  $u_i$ . The eigenvalues of the inverse matrix are given as  $\{\lambda_1^{-1}, \dots, \lambda_K^{-1}\}$ . To see this observe that

$$\boldsymbol{R}^{-1} = \left[\boldsymbol{U}^{H}\right]^{-1} \Lambda^{-1} \boldsymbol{U}^{-1}$$
  
=  $\boldsymbol{U} \Lambda^{-1} \boldsymbol{U}^{H}, \qquad \because \boldsymbol{U}^{-1} = \boldsymbol{U}^{H}.$  (B-5)

The trace of the matrix  $\boldsymbol{R}$  is given as

$$\operatorname{tr} \{ \boldsymbol{R} \} = \sum_{i=1}^{K} \lambda_i. \tag{B-6}$$

To prove this observe that

$$\boldsymbol{R} = \sum_{i=1}^{K} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^H, \tag{B-7}$$

and hence

$$\operatorname{tr} \{\boldsymbol{R}\} = \operatorname{tr} \left\{ \sum_{i=1}^{K} \lambda_{i} \boldsymbol{u}_{i} \boldsymbol{u}_{i}^{H} \right\}$$
$$= \sum_{i=1}^{K} \lambda_{i} \operatorname{tr} \{\boldsymbol{u}_{i} \boldsymbol{u}_{i}^{H} \}$$
$$= \sum_{i=1}^{K} \lambda_{i} \boldsymbol{u}_{i}^{H} \boldsymbol{u}_{i}$$
$$= \sum_{i=1}^{K} \lambda_{i}, \qquad \boldsymbol{u}_{i} \boldsymbol{u}_{i}^{H} = 1.$$

### Appendix

#### **B.1** Proof for Condition on Power Scaling Factor

To prove (2.56), that the constant  $\alpha = \sqrt{\frac{K}{\sum_{k=1}^{K} \frac{1}{\lambda_k}}} \leq 1$ , observe the fact that when  $\mathbf{R}$  corresponds to a spreading cross-correlation matrix,

$$\operatorname{tr} \{ \boldsymbol{R} \} = \sum_{i=1}^{K} \lambda_i = K, \tag{B-8}$$

since the spreading codes are normalized. We will prove that  $\sum_{i=1}^{K} 1/\lambda_i \ge K$  by contradiction. But first we need to note that since  $\mathbf{R}$  is Grammian, it will be positive definite [35], and hence will have all eigenvalues

$$\lambda_i > 0, \qquad \forall i = 1, \cdots, K. \tag{B-9}$$

Let  $\lambda_K$  be the largest eigenvalue. Now assume to the contrary that

$$\sum_{i=1}^{K} 1/\lambda_i < K,\tag{B-10}$$

which implies that

$$\sum_{i=1}^{K} 1/\lambda_i - K < 0$$
  
$$\Rightarrow \sum_{i=1}^{K} \frac{1-\lambda_i}{\lambda_i} < 0$$
  
$$\Rightarrow \frac{\sum_{i=1}^{K} 1-\lambda_i}{\lambda_K} < 0$$
  
$$\Rightarrow \sum_{i=1}^{K} 1-\lambda_i < 0,$$

which contradicts with (B-8). Note that  $\alpha = 1$  corresponds to the case tat  $\lambda_1 = \cdots = \lambda_K = 1$ , which corresponds to  $\mathbf{R} = \mathbf{I}$ .

# C Optimal Precoder Derivation for Flat Fading Channels with Perfect Transmitter CSI

A proof of the identity (2.48) is given in [100]. That proof is based on the gradient approach. Here we will look at an alternate proof for the identity, where we use the completion of squares

### Appendix

approach [30]. Consider the cost function given by (2.47). We will compute the expectations and solve to minimize the cost given. We will neglect the antenna index a.

$$J = E_{\boldsymbol{b},\tilde{\boldsymbol{n}}} \left\{ ||\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}}||^2 \right\}$$
  

$$= E \left\{ (\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}})^H (\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}}) \right\}$$
  

$$= E \left\{ \operatorname{tr} \left\{ (\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}}) (\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}})^H \right\} \right\}$$
  

$$= \operatorname{tr} \left\{ \boldsymbol{D}\boldsymbol{D}^H - \boldsymbol{D}\boldsymbol{M}^H\boldsymbol{R}\boldsymbol{H}^H - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{D}^H + \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{M}^H\boldsymbol{R}\boldsymbol{H}^H + \frac{N_0}{2M_T}\boldsymbol{R} \right\}$$
  

$$= \operatorname{tr} \left\{ (\boldsymbol{D} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}) (\boldsymbol{D} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M})^H + \frac{N_0}{2M_T}\boldsymbol{R} \right\}$$
  

$$= (\boldsymbol{D} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M})^H (\boldsymbol{D} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}) + \operatorname{tr} \left\{ \frac{N_0}{2M_T}\boldsymbol{R} \right\},$$
  
(C-2)

where we have used the following

$$E \left\{ \boldsymbol{b} \boldsymbol{b}^{H} \right\} = \boldsymbol{I}_{K}$$
$$E \left\{ \tilde{\boldsymbol{n}} \tilde{\boldsymbol{n}}^{H} \right\} = \frac{N_{0}}{2M_{T}} \boldsymbol{I}_{K} \boldsymbol{R},$$

and the fact that the noise is zero-mean. It is clear that to minimize the cost  $J^{(a)}$ , we need to force,

$$\left(\boldsymbol{D}^{(a)} - \boldsymbol{H}^{(a)}\boldsymbol{R}\boldsymbol{M}^{(a)}\right) = 0, \qquad (C-3)$$

the solution to which is given by (2.47), (4.39) with  $\mathbf{R} = \mathbf{I}$ , and (4.52).

An alternate proof of the same is based on vector calculus is given below. We use the following identities [35]:

$$\frac{\partial}{\partial \boldsymbol{X}} \operatorname{tr}(\boldsymbol{X}^{H} \boldsymbol{A} \boldsymbol{X} \boldsymbol{B}) = \boldsymbol{A}^{T} \boldsymbol{X}^{*} \boldsymbol{B}^{T}$$
(C-4)

$$\frac{\partial}{\partial \boldsymbol{X}} \operatorname{tr}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}^T.$$
 (C-5)

We will also use the fact that

$$\frac{\partial}{\partial \boldsymbol{X}} \operatorname{tr}(\boldsymbol{X}^{H} \boldsymbol{A}) = \boldsymbol{0}.$$
 (C-6)

Equation C-6 can be obtained from the following definitions [35]:

$$\frac{\partial x^*}{\partial x} = 0, \tag{C-7}$$
and

$$\frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{11}} & \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{12}} & \cdots & \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{1m}} \\\\ \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{21}} & \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{22}} & \cdots & \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{2m}} \\\\ \vdots & \vdots & \ddots & \vdots \\\\ \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{m1}} & \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{m2}} & \cdots & \frac{\partial \operatorname{tr}(\boldsymbol{Y})}{\partial x_{mm}} \end{bmatrix}.$$
(C-8)

Considering

$$\frac{\partial \operatorname{tr}(\boldsymbol{X}^{H}\boldsymbol{A})}{\partial x_{ij}} = \frac{\partial \sum_{l} [\boldsymbol{X}]_{:,l}^{H} [\boldsymbol{A}]_{:,l}}{\partial x_{ij}} = \frac{\partial x_{ij}^{*} a_{ij}}{\partial x_{ij}} = 0,$$
(C-9)

we arrive at (C-6).

In the following discussion we will neglect the antenna index, a, for simplicity.

$$J = E_{\boldsymbol{b},\tilde{\boldsymbol{n}}} \left\{ (\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} - \boldsymbol{n})^{H} (\boldsymbol{D}\boldsymbol{b} - \boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b}\tilde{\boldsymbol{n}}) \right\}$$
  
$$J = E \left\{ \boldsymbol{b}^{H}\boldsymbol{D}^{H}\boldsymbol{D}\boldsymbol{b} - \boldsymbol{b}^{H}\boldsymbol{D}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} - \boldsymbol{b}^{H}\boldsymbol{D}^{H}\tilde{\boldsymbol{n}} - \boldsymbol{b}^{H}\boldsymbol{M}^{H}\boldsymbol{R}^{H}\boldsymbol{H}^{H}\boldsymbol{D}\boldsymbol{b} + \tilde{\boldsymbol{n}}^{H}\boldsymbol{n}\boldsymbol{D}\boldsymbol{b} + \tilde{\boldsymbol{n}}^{H}\boldsymbol{h}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}}^{H}\tilde{\boldsymbol{n}}\right\}.$$
(C-10)  
$$+ \boldsymbol{b}^{H}\boldsymbol{M}^{H}\boldsymbol{R}^{H}\boldsymbol{H}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \boldsymbol{b}^{H}\boldsymbol{M}^{H}\boldsymbol{R}^{H}\boldsymbol{H}^{H}\tilde{\boldsymbol{n}} - \tilde{\boldsymbol{n}}^{H}\boldsymbol{D}\boldsymbol{b} + \tilde{\boldsymbol{n}}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\boldsymbol{b} + \tilde{\boldsymbol{n}}^{H}\tilde{\boldsymbol{n}}\right\}.$$

Taking the expectation with respect to the bits b and the noise  $\tilde{n}$  and neglecting the terms independent of M we get

$$J' = \underbrace{\operatorname{tr}\left\{\boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{H}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\right\}}_{T_{1}} - \underbrace{\operatorname{tr}\left\{\boldsymbol{D}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\right\}}_{T_{2}} - \underbrace{\operatorname{tr}\left\{\boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{H}^{H}\boldsymbol{D}\right\}}_{T_{3}}.$$
 (C-11)

Using (C-4) for  $T_1$  we get

$$\frac{\partial}{\partial \boldsymbol{M}} T_1 = \boldsymbol{R} \boldsymbol{H}^T \boldsymbol{H}^* \boldsymbol{R} \boldsymbol{M}^*. \tag{C-12}$$

Using (C-5) for  $T_2$  we get

$$\frac{\partial}{\partial \boldsymbol{M}}T_2 = \boldsymbol{R}\boldsymbol{D}\boldsymbol{H}.$$
 (C-13)

Using (C-6) for  $T_3$ 

$$\frac{\partial}{\partial \boldsymbol{M}}T_3 = \boldsymbol{0}.$$
 (C-14)

Thus, we get

$$J' = \mathbf{R}\mathbf{H}^T\mathbf{H}^*\mathbf{R}\mathbf{M}^* - \mathbf{R}\mathbf{D}\mathbf{H}.$$
 (C-15)

Equating this cost to zero and using the fact that

$$egin{array}{rcl} D^H &=& D^T = D \ H^T &=& H \ H^* &=& H^H, \end{array}$$

we get

$$M^{*} = (HH^{H}R)^{-1}DH$$
  
=  $R^{-1}(H^{(a)H})^{-1}(H)^{-1}HD$   
=  $R^{-1}(H^{H})^{-1}D$  (C-16)  
 $M = R^{-1*}(H^{H})^{-1*}D^{*}$   
=  $R^{-1}H^{-1}D$ .

This result is also used to obtain the precoders in Chapter 2, (2.47), and Chapter 4, (4.39) and (4.52) along with (4.53) to obtain the optimal precoders.

### C.1 Equi-correlated Spreading Sequences

The matrix inversion lemma [35], is used to compute the inverse of a small rank adjustment, defined for a full rank matrix A, as the addition of a rank one matrix as shown below

$$\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{x} \boldsymbol{y}^H. \tag{C-17}$$

The inverse of B can be computed from that of A by using the following equation,

$$B^{-1} = A^{-1} - A^{-1} \frac{xy^{H}}{1 + y^{H} A^{-1} x} A^{-1}.$$
 (C-18)

For  $\boldsymbol{R}$ , given by (2.62), we can write,

$$\boldsymbol{R} = \boldsymbol{A} + \boldsymbol{x} \boldsymbol{y}^{H}, \tag{C-19}$$

where

$$A = (1 - \rho)I_K$$
$$x = [1, 1, \cdots, 1]^T$$
$$y = \rho x.$$

Then,

$$\begin{aligned} \boldsymbol{R}^{-1} &= \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \frac{\boldsymbol{x} \boldsymbol{y}^{H}}{1 + \boldsymbol{y}^{H} \boldsymbol{A}^{-1} \boldsymbol{x}} \boldsymbol{A}^{-1} \\ &= \frac{1}{1 - \rho} \boldsymbol{I}_{K} - (\frac{1}{1 - \rho})^{2} \boldsymbol{I}_{K} \frac{\rho \boldsymbol{1}_{K}}{\frac{1 + (K - 1)\rho}{1 - \rho}} \\ &= \frac{1}{1 - \rho} \boldsymbol{I}_{K} - \frac{\rho}{1 + (K - 2)\rho - (K - 1)\rho^{2}} \boldsymbol{1}_{K} \end{aligned}$$

and summing the diagonal elements we get (2.63).

## D Proof for Partial Channel Information

The key step is computing the conditional expectations (3.1)

$$J^{(a)} = E_{\boldsymbol{b}, \tilde{\boldsymbol{n}}, \{\boldsymbol{H}^{(i)}\}} \left\{ \left\| \boldsymbol{D}^{(a)} \boldsymbol{b} - \boldsymbol{H}^{(a)} \boldsymbol{R} \boldsymbol{M}^{(a)} \boldsymbol{b} + \tilde{\boldsymbol{n}} \right\|^{2} \left| \{ \hat{\boldsymbol{H}}^{(i)} \} \right\}.$$
(D-2)

Neglecting the antenna index a, and proceeding in a manner similar to that for the derivation for perfect channel information, we arrive at,

$$J' = \underbrace{E_{\left\{\boldsymbol{H}^{(i)}\right\}}\left\{\operatorname{tr}\left\{\boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{H}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\right\}\left|\left\{\hat{\boldsymbol{H}}^{(i)}\right\}\right\}}_{T_{1}} - \underbrace{E_{\left\{\boldsymbol{H}^{(i)}\right\}}\left\{\operatorname{tr}\left\{\boldsymbol{D}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M}\right\}\left|\left\{\hat{\boldsymbol{H}}^{(i)}\right\}\right\}}_{T_{2}}}_{T_{2}} (D-3)$$

Observe the expectation to be taken over the actual channel, conditioned upon the partial channel information as compared to the expression in (C-11). Consider

$$T_{2} = E\left\{ \operatorname{tr}\left\{ \boldsymbol{D}^{H}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M} \middle| \left\{ \hat{\boldsymbol{H}}^{(i)} \right\} \right\} \right\}$$
$$= E\left\{ \operatorname{tr}\left\{ \boldsymbol{D}\boldsymbol{H}\boldsymbol{R}\boldsymbol{M} \middle| \left\{ \hat{\boldsymbol{H}}^{(i)} \right\} \right\} \right\}.$$
(D-4)

Observe that the elements  $\mathbf{R}$  and  $\mathbf{M}$  are arbitrary and we do not take any expectation over them. As a result we can move the expectation operator inside the trace operator (since both are linear operators,  $E\{ax_1 + bx_2\} = aE\{x_1\} + bE\{x_2\}$ , while  $tr\{\mathbf{AX}\} = \sum_i [\mathbf{A}]_{:,i} [\mathbf{X}]_{i,:}$ . Thus,  $E\{tr\{\mathbf{AX}\}\} = E\{\sum_i [\mathbf{A}]_{:,i} [\mathbf{X}]_{i,:}\} = \sum_i [\mathbf{A}]_{:,i} E\{[\mathbf{X}]_{i,:}\}$ . Consequently, we can write

$$T_2 = E\left\{ \operatorname{tr}\left\{ \boldsymbol{D}\boldsymbol{H} \middle| \left\{ \hat{\boldsymbol{H}}^{(i)} \right\} \right\} \right\} \boldsymbol{R}\boldsymbol{M}.$$
 (D-5)

Using (3.3), and  $\boldsymbol{D}^{H} = \boldsymbol{D}$ , and  $\boldsymbol{H}^{H} = \boldsymbol{H}^{*}$  we get

$$T_2 = \operatorname{tr}\left\{ \boldsymbol{C}^*_{\boldsymbol{H}\boldsymbol{D}} \boldsymbol{R} \boldsymbol{M} \right\}.$$
 (D-6)

Using, (C-5), we get

$$\frac{\partial}{\partial \boldsymbol{M}} T_2 = \boldsymbol{R} \boldsymbol{C}_{\boldsymbol{H} \boldsymbol{D}}^H. \tag{D-7}$$

Considering

$$T_{3} = E\left\{ \operatorname{tr}\left\{ \boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{H}^{H}\boldsymbol{D}\right\} \middle| \left\{ \hat{\boldsymbol{H}}^{(i)} \right\} \right\}$$
  
=  $\operatorname{tr}\left\{ \boldsymbol{M}^{H}\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{H}\boldsymbol{D}} \right\}.$  (D-8)

Using (C-6), we get

$$\frac{\partial}{\partial \boldsymbol{M}}T_3 = \boldsymbol{0}.\tag{D-9}$$

 $\operatorname{Consider}$ 

$$T_{1} = E \left\{ \operatorname{tr} \left\{ \boldsymbol{M}^{H} \boldsymbol{R} \boldsymbol{H}^{H} \boldsymbol{H} \boldsymbol{R} \boldsymbol{M} \right\} \left| \left\{ \hat{\boldsymbol{H}}^{(i)} \right\} \right\}$$
  
=  $\operatorname{tr} \left\{ \boldsymbol{M}^{H} \boldsymbol{R} \boldsymbol{C}_{\boldsymbol{H} \boldsymbol{H}} \boldsymbol{R} \boldsymbol{M} \right\}.$  (D-10)

Using (C-4),

$$\frac{\partial}{\partial \boldsymbol{M}}T_1 = \boldsymbol{R}\boldsymbol{C}_{\boldsymbol{H}\boldsymbol{H}}\boldsymbol{R}\boldsymbol{M}^*.$$
 (D-11)

Thus equating J' = 0, and using the fact that  $C_{HD}^{H} = C_{HD}^{*}$  and  $C_{HH}^{*} = C_{HD}$ , we arrive at

$$\boldsymbol{M}^* = \boldsymbol{R}^{-1} \boldsymbol{C}_{\boldsymbol{H}\boldsymbol{H}}^{-1} \boldsymbol{C}_{\boldsymbol{H}\boldsymbol{D}}^{\boldsymbol{H}}$$
(D-12)

$$\boldsymbol{M} = \boldsymbol{R}^{-1} \boldsymbol{C}_{\boldsymbol{H}\boldsymbol{H}}^{-1} \boldsymbol{C}_{\boldsymbol{H}\boldsymbol{D}}.$$
 (D-13)

Reintroducing the antenna index we get (3.4).

# E Discrete-Time Model for CDMA Communication over Frequency Selective Channel

Suppose that we are using CDMA for a single user system with a frequency selective channel. Corresponding to the discrete-time transmitted signal given in (4.1), the continuous-time transmitted signal can be written using (1.101) as

$$x(t) = b \sum_{i=0}^{N-1} s[i]p(t - iT_c) 0 \le t \le NT_c,$$
 (E-2)

where we have incorporated the energy term  $\sqrt{\mathcal{E}}_b$  in b. Let the channel between the transmitter and the user be given as a multi-tap filter given as

$$\sum_{l=0}^{L-1} h_i^{(a)}[l]\delta(\tau - lT_c),$$
(E-3)

which as we see from (1.53), corresponds to the case where  $B = 1/T_c$ . The channel is frequency selective, and hence the received signal can be written using  $x(t) = s_m(t)$  in (1.54) to get

$$r(t) = \sum_{l=0}^{L-1} h[l] x\left(t - \frac{l}{B}\right) + n(t), \qquad 0 \le t \le T_d.$$
(E-4)

The output of the chip match filter during interval i will be

$$r[i] = \int_{iT_c}^{(i+1)T_c} r(t)p(t-iT_c)dt$$
  
=  $\int_{iT_c}^{(i+1)T_c} \left(\sum_{l=0}^{L-1} h[l]x(t-lT_c) + n(t)\right) p(t-iT_c)dt$  (E-5)

The noise-free output can be written as

$$\begin{aligned} r'[i] &= \int_{iT_c}^{(i+1)T_c} \left( \sum_{l=0}^{L-1} h[l]x(t-lT_c) \right) p(t-iT_c) \mathrm{dt} \\ &= \int_{iT_c}^{(i+1)T_c} \left( \sum_{l=0}^{L-1} h[l]b \sum_{j=0}^{N-1} s[j]p(t-jT_c-lT_c) \right) p(t-iT_c) \mathrm{dt} \\ &= \sum_{l=0}^{L-1} h[l]b \sum_{j=0}^{N-1} s[j] \int_{iT_c}^{(i+1)T_c} \left( p(t-(j+l)T_c) \right) p(t-iT_c) \mathrm{dt} \\ &= \sum_{l=0}^{L-1} h[l]b \sum_{j=0}^{N-1} s[j]\delta[i-j-l] \\ &= \sum_{l=0}^{L-1} h[l]s[i-l]b \end{aligned}$$
(E-6)

Note that the receiver receives a signal for a total interval of  $N + LT_c$  seconds and we have assumed the receiver to be chip-synchronous with the transmitter. Therefore, the output r'[i] will be available for  $i = 0, \dots, N + L - 1$ . Adding the noise components at the output of the chip match filter we get

$$r[i] = \sum_{l=0}^{L-1} h[l]s[i-l]b + n[i], \qquad i = 0, \cdots, N+L-1$$
(E-7)

and correspondingly the vector formed by collecting all the  ${\cal N}+L$  match-filter outputs will be ,

$$\mathbf{r} = \begin{bmatrix} \sum_{l=0}^{L-1} h[l]s[1-l]b + n[1] \\ \sum_{l=0}^{L-1} h[l]s[2-l]b + n[2] \\ \vdots \\ \sum_{l=0}^{L-1} h[l]s[N+L-l]b + n[N+L] \end{bmatrix}$$
(E-8)  
$$= \sum_{l=0}^{L-1} h[l] \begin{bmatrix} s[1-l] \\ s[2-l] \\ \vdots \\ s[N+L-l] \end{bmatrix} b + \mathbf{n},$$
(E-9)

where  $\boldsymbol{n} = [n[1], \cdots, n[N+L]]^T$  is a white Gaussian noise vector. Using the fact that

$$s[i] = 0$$
  $i < 0, i \ge N$  (E-10)

we get

$$s[i-l] = 0$$
  $i < l, i \ge N + l.$  (E-11)

Using the notation

$$\boldsymbol{s}[l] = \begin{bmatrix} \boldsymbol{0}_{l \times 1} \\ \boldsymbol{s} \\ \boldsymbol{0}_{N+L-l \times 1} \end{bmatrix}$$
(E-12)

we can write (E-9) as (4.3). To observe the resulting signal from multi-user transmission observe that corresponding to the discrete-time transmitted signal

$$\boldsymbol{x} = \boldsymbol{S}\boldsymbol{b},\tag{E-13}$$

we can write the transmitted signal in (E-2) as

$$x(t) = \sum_{k=1}^{K} \sum_{i=0}^{N-1} s_k[i] p(t - iT_c) 0 \le t \le NT_c b[k],$$
(E-14)

and proceeding with the same steps as in (E-4) to (4.3) we can write the signal received by user k as

$$\boldsymbol{r}_{k} = \sum_{l=0}^{L-1} h[l]\boldsymbol{S}[l]\boldsymbol{b} + \boldsymbol{n}, \qquad (E-15)$$

where the matrix

$$\boldsymbol{S}[l] = \left[ \begin{array}{ccc} \boldsymbol{s}_1[l] & \cdots & \boldsymbol{s}_K[l], \end{array} \right]$$
(E-16)

where  $s_k[l]$  is defined in a manner similar to s[l] as given in (E-12). We will use extend this notation to see how the received signal is modelled when precoding is used.

### E.1 Received signal with Precoding

In this section we see how the received signal can be written as (4.28). The signal transmitted from precoder  $M^{(a)}[l], x^{(a)}[l] \in \mathbb{C}^{N+L-1\times 1}$  can be given as

$$\boldsymbol{x}^{(a)}[l] = \alpha \tilde{\boldsymbol{S}}[l] \boldsymbol{M}^{(a)}[l] \boldsymbol{b}.$$
(E-17)

Note that when the transmitted signal was of length N as given in (4.1), the received signal  $\boldsymbol{r}$  given in (E-7) was of length N + L. Therefore the signal resulting from transmission of  $\boldsymbol{x}^{(a)}[l]$  will be of length N + 2L - 1. With steps similar to those in the above section, we can write the noise-free signal received from all paths due to transmission from  $\boldsymbol{M}^{(a)}[l]$ , which we represent as  $\boldsymbol{r}_{k}^{a'}[l] \in \mathbb{C}^{N+2L-1\times 1}$  as

$$\boldsymbol{r}_{k}^{a'}[l] = \sum_{j=0}^{L-1} h_{k}^{a}[j] \tilde{\boldsymbol{S}}[l,j] \boldsymbol{M}^{(a)}[l] \boldsymbol{b},$$
(E-18)

where we have neglected the noise term for simplicity and similar to (E-12) we have defined

$$\tilde{\boldsymbol{S}}[l,j] = \begin{bmatrix} \boldsymbol{0}_{j \times K} \\ \tilde{\boldsymbol{S}}[l] \\ \boldsymbol{0}_{N+L-j \times K} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0}_{j \times K} \\ \boldsymbol{0}_{L-1-l \times K} \\ \boldsymbol{S} \\ \boldsymbol{0}_{l \times K} \\ \boldsymbol{0}_{N+L-j \times K} \end{bmatrix}.$$
(E-19)

However, our receiver receives the signal only from intervals L to N + L - 1. The resulting received signal  $\boldsymbol{r}_k^a[l] \in \mathbb{C}^{N \times 1}$  can be written as

$$\boldsymbol{r}_{k}^{a}[l] = \left[\boldsymbol{r}_{k}^{a'}[l]\right]_{L:N+L-1,:}$$
  
=  $\sum_{j=0}^{L-1} h_{k}^{a}[j] \left[\tilde{\boldsymbol{S}}[l,j]\right]_{L:N+L-1,:} \boldsymbol{M}^{(a)}[l]\boldsymbol{b}.$  (E-20)

Observe that we can write

$$\begin{bmatrix} \tilde{\boldsymbol{S}}[l,j] \end{bmatrix}_{L:N+L-1,:} = \boldsymbol{S}[j-l] \triangleq \begin{cases} \begin{bmatrix} \boldsymbol{0}_{j-l \times K} \\ [\boldsymbol{S}]_{1:N-j+l,:} \end{bmatrix} & j > l \\ \begin{bmatrix} [\boldsymbol{S}]_{N-l+j+1:N,:} \\ \boldsymbol{0}_{l-j \times K} \end{bmatrix} & j < l \\ \boldsymbol{S} & j = l \end{cases}$$
(E-21)

to get the form in (4.28). Another way to observe the received signal during the reception interval will be to look at the corresponding part of  $x^{(a)}$  during the reception interval, which is

$$[\boldsymbol{x}^{(a)}]_{L:N+L-1,:} = [\alpha \breve{\boldsymbol{S}} \boldsymbol{\mathcal{M}}^{(a)} \boldsymbol{b}]_{L:N+L-1,:}.$$
(E-22)

The reception over the  $j^{th}$  path will be a version of  $x^{(a)}$ , delayed by j chips. Observe that the entire received signal from path j during the reception interval will be

$$h_{k}^{(a)}[j][\boldsymbol{x}^{(a)}]_{L-j:N+L-1-j,:} = h_{k}^{(a)}[j][\boldsymbol{\breve{S}M}^{(a)}\boldsymbol{b}]_{L-j:N+L-1-j,:}$$
(E-23)

The part corresponding to our precoder  $M^{(a)}[l]$ , will be

$$h_{k}^{(a)}[j]\alpha[\breve{\boldsymbol{S}}]_{L-j:N+L-1-j,(l-1)K+1:lK}\boldsymbol{M}^{(a)}[l]\boldsymbol{b} = \alpha h_{k}^{(a)}[j] \left[\tilde{\boldsymbol{S}}[l]\right]_{L-j:N+L-1-j,:} \boldsymbol{M}^{(a)}[l]\boldsymbol{b}$$

$$= \alpha h_{k}^{(a)}[j]\boldsymbol{S}[j-l]\boldsymbol{M}^{(a)}[l]\boldsymbol{b}.$$
(E-24)

This provides an alternate, and maybe simpler explanation for (4.28).

# F Signal Model for Jointly Designed Precoders for Channels without ISI

We will show how simple matrix manipulation results in the form of (4.46).

$$\begin{aligned} \mathbf{y}^{(a')} &= [\mathbf{y}_{1}^{(a')}, \dots, \mathbf{y}_{K}^{(a')}]^{T} \\ &= \begin{bmatrix} \sum_{j=0}^{L-1} \sum_{l=0}^{L-1} h_{1}^{(a)}[j]\mathbf{s}_{1}^{H}\mathbf{S}[j-l]\mathbf{M}^{(a)}[l]\mathbf{b} \\ \sum_{j=0}^{L-1} \sum_{l=0}^{L-1} h_{2}^{(a)}[j]\mathbf{s}_{2}^{H}\mathbf{S}[j-l]\mathbf{M}^{(a)}[l]\mathbf{b} \\ \vdots \\ \sum_{j=0}^{L-1} \sum_{j=0}^{L-1} \prod_{l=0}^{L-1} h_{K}^{(a)}[j]\mathbf{s}_{K}^{H}\mathbf{S}[j-l]\mathbf{M}^{(a)}[l]\mathbf{b} \end{bmatrix} \\ &= \sum_{l=0}^{L-1} \sum_{j=0}^{L-1} \prod_{l=0}^{L-1} \left[ \begin{array}{c} h_{1}^{(a)}[j] & 0 & \cdots & 0 \\ 0 & h_{2}^{(a)}[j] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & h_{K}^{(a)}[j] \end{array} \right] \left[ \begin{array}{c} \mathbf{s}_{1}^{H}\mathbf{S}[j-l] \\ \vdots \\ \mathbf{s}_{K}^{H}\mathbf{S}[j-l] \end{array} \right] \mathbf{M}^{(a)}[l]\mathbf{b} \\ &= \sum_{l=0}^{L-1} \sum_{j=0}^{L-1} \mathbf{H}^{(a)}[j] \left[ \begin{array}{c} \mathbf{s}_{1}^{H} \\ \mathbf{s}_{1}^{H} \\ \vdots \\ \mathbf{s}_{K}^{H} \end{array} \right] \mathbf{S}[j-l]\mathbf{M}^{(a)}[l]\mathbf{b} \\ &= \sum_{l=0}^{L-1} \sum_{j=0}^{L-1} \mathbf{H}^{(a)}[j] \left[ \begin{array}{c} \mathbf{s}_{1}^{H} \\ \mathbf{s}_{1}^{H} \\ \vdots \\ \mathbf{s}_{K}^{H} \end{array} \right] \mathbf{S}[j-l]\mathbf{M}^{(a)}[l]\mathbf{b} \\ &= \sum_{l=0}^{L-1} \sum_{j=0}^{L-1} \mathbf{H}^{(a)}[j]\mathbf{R}_{j-l}\mathbf{M}^{(a)}[l]\mathbf{b}, & \because \mathbf{R}_{j-l} = \mathbf{S}^{H}\mathbf{S}[j-l] \\ &\mathbf{y}^{(a')} = \underbrace{\left[ \mathbf{H}^{(a)}[0] & \dots & \mathbf{H}^{(a)}[L-1] \\ \mathbf{H}^{(a)} \\ &\mathbf{H}^{(a)} \\ &= \underbrace{\left[ \begin{array}{c} \mathbf{R}_{0} & \cdots & \mathbf{R}_{-(L-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{L-1} & \cdots & \mathbf{R}_{0} \\ &\mathbf{R} \\ \end{array} \right] \mathbf{M}^{(a)}\mathbf{b}. \end{aligned}$$

# F.1 Constant for Peak to Average Power Ratio

Using (4.41) the average transmitted power will be given by

$$P_{avg} = p_0 \beta K + (1 - p_0) E \left\{ \sqrt{\frac{zK}{E \{\varrho\}}} \right\}$$
$$= p_0 \beta K + (1 - p_0) zK$$
$$= K,$$
(F-3)

where  $p_0$  is the probability that the transmission takes place at peak power  $\beta K$ . Thus,

$$z = \frac{P_{avg} - p_0\beta K}{(1 - p_0)K},$$

and since we have chosen that the average transmit power is K, we get (4.42).

### F.2 Proof for Multipath Precoding with ISI

Taking the expectation in (4.73) and neglecting the terms independent of  $\tilde{\mathcal{M}}^{(a)}$  we can write the above cost function as

$$J^{(a)} = \underbrace{\operatorname{tr}\left\{\tilde{\mathcal{M}}^{(a)H}\tilde{\mathcal{H}}^{(a)H}\tilde{\mathcal{H}}^{(a)}\tilde{\mathcal{M}}^{(a)}\right\}}_{T_{1}} - \underbrace{\operatorname{tr}\left\{\tilde{\mathcal{D}}^{(a)H}\tilde{\mathcal{H}}^{(a)}\tilde{\mathcal{M}}^{(a)}\right\}}_{T_{2}} - \underbrace{\operatorname{tr}\left\{\tilde{\mathcal{M}}^{(a)H}\tilde{\mathcal{H}}^{(a)H}\tilde{\mathcal{D}}^{(a)}\right\}}_{T_{3}} \quad (F-4)$$

To compute  $T_1$  we observe that from (4.70) and (4.71)

$$\tilde{\boldsymbol{\mathcal{H}}}^{(a)} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)}\boldsymbol{\mathcal{R}}_{1} & \boldsymbol{0} & \dots & \dots & \boldsymbol{0} \\ \boldsymbol{\mathcal{H}}_{-1}^{(a)}\boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)}\boldsymbol{\mathcal{R}}_{1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \dots & \boldsymbol{\mathcal{H}}_{-1}^{(a)}\boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} & \boldsymbol{\mathcal{H}}_{1}^{(a)}\boldsymbol{\mathcal{R}}_{1} \\ \boldsymbol{0} & \dots & \dots & \boldsymbol{0} & \boldsymbol{\mathcal{H}}_{-1}^{(a)}\boldsymbol{\mathcal{R}}_{-1} & \boldsymbol{\mathcal{H}}_{0}^{(a)}\boldsymbol{\mathcal{R}}_{0} \end{bmatrix}$$

$$ilde{\mathcal{M}}^{(a)} \triangleq egin{bmatrix} \mathcal{M}^{(a)} & \mathbf{0} & \mathbf{0} \\ dots & \ddots & dots \\ \mathbf{0} & \mathbf{0} & \mathcal{M}^{(a)} \end{bmatrix}$$

we get

 $ilde{oldsymbol{\mathcal{H}}}^{(a)} ilde{oldsymbol{\mathcal{M}}}^{(a)} =$ 

$egin{array}{l} \mathcal{H}_0^{(a)}\mathcal{R}_0\mathcal{M}^{(a)}\ \mathcal{H}_{-1}^{(a)}\mathcal{R}_{-1}\mathcal{M}^{(a)} \end{array}$	$oldsymbol{\mathcal{H}}_1^{(a)} oldsymbol{\mathcal{R}}_1 oldsymbol{\mathcal{M}}^{(a)} \ oldsymbol{\mathcal{H}}_0^{(a)} oldsymbol{\mathcal{R}}_0 oldsymbol{\mathcal{M}}^{(a)}$	$egin{array}{c} 0 \ \mathbf{\mathcal{H}}_1^{(a)} \mathbf{\mathcal{R}}_1 \mathbf{\mathcal{M}}^{(a)} \end{array}$	0		0 - 0	
-	·	·	·	·	:	$(\mathbf{F}_{-5})$
: :	÷.,	÷.,	÷.,	÷.,		(1-5)
0			$\mathcal{H}_{-1}^{(a)}\mathcal{R}_{-1}\mathcal{M}^{(a)}$	$\mathcal{H}_{0}^{(a)}\mathcal{R}_{0}\mathcal{M}^{(a)}$	$\mathcal{H}_{1}^{(a)}\mathcal{R}_{1}\mathcal{M}^{(a)}$	
L O			0	$\mathcal{H}_{-1}^{(a)}\mathcal{R}_{-1}\mathcal{M}^{(a)}$	$\mathcal{H}_0^{(a)} \mathcal{R}_0 \mathcal{M}^{(a)}$ _	

and

	$ ilde{\mathcal{M}}^{(a)H} ilde{\mathcal{H}}^{(a)H}$	=					
[	$\mathbf{\mathcal{M}}^{(a)H}\mathbf{\mathcal{R}}_{0}^{H}\mathbf{\mathcal{H}}_{0}^{(a)H}$	$\mathcal{M}^{(a)H}\mathcal{R}_{-1}^{H}\mathcal{H}_{-1}^{(a)H}$	0			0	1
	$\mathcal{M}^{(a)H} \mathcal{R}_1^H \mathcal{H}_1^{(a)H}$	$\mathcal{M}^{(a)H}\mathcal{R}_{0}^{H}\mathcal{H}_{0}^{(a)H}$	·	0		0	
	:	·	·	·	·	:	
		·	·	·	·	:	
	0			$\mathcal{M}^{(a)H} \mathcal{R}_1^H \mathcal{H}_1^{(a)H}$	$\mathcal{M}^{(a)H} \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{R}_0 \mathcal{M}^{(a)}$	$\mathcal{M}^{(a)H} \mathcal{R}^{H}_{-1} \mathcal{H}^{(a)H}_{-1}$	
L				0	$\mathcal{M}^{(1)} \mathcal{K}_1^{-1} \mathcal{K}_1^{-1}$	$\mathcal{M}^{(1)} \mathcal{K}_{0}^{-} \mathcal{K}_{0}^{-}$	1

We have

$$T_1 = \operatorname{tr}\{\tilde{\boldsymbol{\mathcal{M}}}^{(a)H} \tilde{\boldsymbol{\mathcal{H}}}^{(a)H} \tilde{\boldsymbol{\mathcal{H}}}^{(a)} \tilde{\boldsymbol{\mathcal{M}}}^{(a)}\}.$$
 (F-6)

Using the fact that the trace corresponds only to the diagonal entries we get

$$T_{1} = p \operatorname{tr} \{ \mathcal{M}^{(a)H} \mathcal{R}_{0}^{H} \mathcal{H}_{0}^{(a)H} \mathcal{H}_{0}^{(a)} \mathcal{R}_{0} \mathcal{M}^{(a)} \} - (p-1) \operatorname{tr} \{ \mathcal{M}^{(a)H} \mathcal{R}_{-1}^{H} \mathcal{H}_{-1}^{(a)H} \mathcal{H}_{-1}^{(a)} \mathcal{R}_{-1} \mathcal{M}^{(a)} \}$$

$$- (p-1) \operatorname{tr} \{ \mathcal{M}^{(a)H} \mathcal{R}_{1}^{H} \mathcal{H}_{1}^{(a)H} \mathcal{H}_{1}^{(a)} \mathcal{R}_{1} \mathcal{M}^{(a)} \}.$$
(F-7)

Similarly computing the matrix products of the terms in the arguments of the tr $\{\cdot\}$  functions and using the block diagonal nature of  $\tilde{\mathcal{D}}^{(a)}$ , we get

$$T_2 = p \operatorname{tr} \left\{ \mathcal{D}^{(a)H} \mathcal{H}_0^{(a)} \mathcal{R}_0 \mathcal{M}^{(a)} \right\}$$
(F-8)

$$T_3 = p \operatorname{tr} \left\{ \mathcal{M}^{(a)H} \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{D}^{(a)} \right\},$$
 (F-9)

Using rules from vector calculus and following an approach similar to that use for the previous derivations given in Appendix C, we obtain

$$\frac{\partial}{\partial \mathcal{M}^{(a)}} T_1 = p \mathcal{R}_0^H \mathcal{H}_0^{(a)H} \mathcal{H}_0^{(a)} \mathcal{R}_0 \mathcal{M}^{(a)*} - (p-1) \mathcal{R}_1^H \mathcal{H}_1^{(a)H} \mathcal{H}_1^{(a)} \mathcal{R}_1 \mathcal{M}^{(a)*}$$
(F-10)  
$$-(p-1) \mathcal{R}_{-1}^H \mathcal{H}_{-1}^{(a)H} \mathcal{H}_{-1}^{(a)} \mathcal{R}_{-1} \mathcal{M}^{(a)*}$$

$$\frac{\partial}{\partial \mathcal{M}^{(a)}} T_2 = p \mathcal{R}_0^T \mathcal{H}_0^{(a)T} \mathcal{D}^{(a)}$$
(F-11)

$$\frac{\partial}{\partial \mathcal{M}^{(a)}} T_3 = 0 \tag{F-12}$$

and equating  $J^{(a)'} = 0$ , we get (4.74)

### Proof for with Partial CSI at Transmitter

With partial CSI we will need to take the expectation in (4.73) over the channel as well. This will simple be a modification of F-4 to include the conditional expectation over the channel as,

$$J^{(a)} = \underbrace{E_{\{\mathcal{H}^{(i)}\}} \left\{ \operatorname{tr} \left\{ \tilde{\mathcal{M}}^{(a)H} \tilde{\mathcal{H}}^{(a)H} \tilde{\mathcal{H}}^{(a)} \tilde{\mathcal{M}}^{(a)} \right\} \left| \{\hat{\mathcal{H}}^{(i)} \} \right\}}_{T_1} - \underbrace{E_{\{\mathcal{H}^{(i)}\}} \left\{ \operatorname{tr} \left\{ \tilde{\mathcal{D}}^{(a)H} \tilde{\mathcal{H}}^{(a)} \tilde{\mathcal{M}}^{(a)} \right\} \left| \{\hat{\mathcal{H}}^{(i)} \}}_{T_2} \right.}_{T_2} \left( \text{E-13} \right)$$

$$- \underbrace{E_{\{\mathcal{H}^{(i)}\}} \left\{ \operatorname{tr} \left\{ \tilde{\mathcal{M}}^{(a)H} \tilde{\mathcal{H}}^{(a)H} \tilde{\mathcal{D}}^{(a)} \right\} \left| \{\hat{\mathcal{H}}^{(i)} \} \right.}_{T_3} \right\}$$

Moving on the same lines as in Section D and using (4.76), (4.78), (4.79) and (4.79), to get

$$\frac{\partial}{\partial \mathcal{M}^{(a)}} T_1 = p \mathcal{R}_0^H \mathcal{C}_{HH_0}^{(a)} \mathcal{R}_0 \mathcal{M}^{(a)*} - (p-1) \mathcal{R}_1^H \mathcal{C}_{HH_1}^{(a)} \mathcal{R}_1 \mathcal{M}^{(a)*}$$

$$-(p-1) \mathcal{R}_{-1}^H \mathcal{C}_{HH_{-1}}^{(a)} \mathcal{R}_{-1} \mathcal{M}^{(a)*}$$
(F-14)

$$\frac{\partial}{\partial \mathcal{M}^{(a)}} T_2 = p \mathcal{R}_0^T \mathcal{C}_{H_0 D}^{(a)*}$$
(F-15)

$$\frac{\partial}{\partial \mathcal{M}^{(a)}} T_3 = 0 \tag{F-16}$$

and equating (F-14) and (F-15) we get (4.80).