

UNIVERSITE DE NICE-SOPHIA ANTIPOLIS

ECOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THESE

pour obtenir le titre de

DOCTEUR EN SCIENCES

de l'Université de Nice-Sophia Antipolis

Discipline: Automatique, Traitement du Signal et des Images

présentée et soutenue par

Umer SALIM

**Communication with Channel Uncertainties,
Feedback Acquisition and Optimization**

Thèse dirigée par *Prof. Dirk SLOCK*
soutenue le 27 Janvier 2010 à EURECOM

Jury:

M. Merouane DEBBAH	Prof. à Supélec, France	Rapporteur
Mme Ana PEREZ	Prof. à UPC, Espagne	Rapporteur
M. David GESBERT	Prof. à Eurécom, France	Examineur
M. Wolfgang GERSTACKER	Prof. à Univ. de Erlangen-Nürnberg, Allemagne	Examineur
M. Luc DENEIRE	Prof. à Université de Nice, France	Examineur
M. Dirk SLOCK	Prof. à Eurécom, France	Directeur de thèse

Acknowledgements

I would like to extend my sincere thanks to my advisor Prof. Dirk Slock for his invaluable support and brilliant ideas. He is a passionate researcher but above all a true gentleman.

I am grateful to the committee members of my jury, Professor Mérouane Debbah, Professor Ana Perez, Professor David Gesbert, Professor Wolfgang Gerstacker and Professor Luc Deneire for their valuable inputs and time spent reading this thesis.

I am very much thankful to David Gesbert for letting me attend M3 meetings which I really liked and profited. He has a unique gift of identifying the problem of interest from the "mess" (PhD students have the equally unique ability to make a mess out of anything) you discuss with him and describe in simplest of words the true problem statement.

I am particularly indebted to some fellow PhD students at Eurecom. I was extremely lucky to have the best persons as my office mates of 110 B, Najam and Randa, whom I consider my trustworthy friends. Well, it's obvious that they were not that lucky to have me in the office. I am very grateful to Ruben, Randa and Rizwan (inundated by R's, yes) for many useful technical discussions during the course of my PhD. Randa and Rizwan, among other things, made extremely valuable suggestions to improve the quality of my thesis manuscript which implies that someone made them read some parts of my thesis (Quelle horreur!).

I would like to thank all of my friends at Eurecom, all my former colleagues and current PhD students. I will always cherish my good memories earned at Eurecom.

I am all thanks to my Pakistani friends in Cote d'Azur. We really have had a wonderful time in many funny and banal activities, the most notably going out for dinners.

Finally, I would like to thank my family who stood by me unconditionally through all thick and thin.

Abstract

In wireless communications, there has always been a quest for being able to transmit large amounts of information at the expense of minimal resource utilization. The study of capacity/rate both for single-user and multi-user channels without initial assumption of channel state information (CSI) is important as this capacity not only indicates the upper bound of the rate limit achievable by any transmission scheme but also shows power efficient and inefficient regimes of operation.

For the channels without any assumption of CSI, the capacity analysis could become highly cumbersome even for channel models of moderate complexity. Due to intractability of this analysis, we focus on the high signal-to-noise ratio (SNR) regime in the first part of this thesis which lets us characterize the dominant capacity term, the term with $\log(P)$ where P is the transmit power. For single-user single-antenna symbol-by-symbol stationary channels, the exact pre-log (the coefficient of $\log(P)$) is specified for underspread channels. A novel transmission scheme is shown to achieve non-zero pre-log for overspread channels. Considering multi-user downlink (DL) channels with a multi-antenna base station (BS), tight lower and upper bounds of the multiplexing gain are derived for relatively simple block-stationary channels with no assumption of CSI. This analysis also reveals how the quality of CSI at the BS should be scaled with SNR to preserve the multiplexing gain.

A multi-user DL channel shows promising gains even with single-antenna user terminals. Achievability of these gains requires the presence of good quality CSI at the BS which might involve a significant overhead on the uplink (UL) resource. In the second part of this thesis, the focus is on the design and optimization of CSI feedback at the transmitter. We study a multi-user DL system without any assumption of CSI and derive sum rate bounds when CSI acquisition is completely accounted for. These bounds allow maximizing the sum rate by achieving the cost-benefit trade-off of CSI feedback. Further, we propose a novel CSI feedback acquisition strategy

for reciprocal channels which combines the use of training sequence and quantized feedback contrary to the classical scheme for such channels where only pilot sequences are used. The results show the superiority of the new hybrid acquisition over traditional CSI acquisition schemes.

For multi-user channels with CSI, the sum rate maximization under fixed power constraints has been widely studied. The dual of this problem, namely, the minimization of transmit power required to achieve specific rate/quality targets at the users' side is also an equally important design problem and of high interest to service providers. In the third part of this thesis, we study the problem of transmit power minimization in conjunction with user scheduling for various user selection schemes. Some analytical results are derived in the limiting case and it is shown that semi-orthogonal greedy user selection performs much better than some other user selection schemes.

Contents

Acknowledgements	i
Abstract	iii
Contents	v
List of Figures	xi
Acronyms	xiii
Notations	xv
1 Introduction	1
1.1 Evolution in Wireless Communications	1
1.1.1 From Single-Antenna to Multi-Antenna Links	2
1.1.2 Single-User to Multi-User Paradigm Shift	2
1.1.3 The Role of Channel Information	3
1.2 Major Themes in the Thesis	4
1.2.1 Channel Non-Coherence and DOF Perspective	4
1.2.2 DL Sum Rate Maximization and CSIT Feedback	6
1.2.3 Transmit Power Minimization with User Selection	7
1.3 Thesis Outline and Contributions	8
I DOF for Single-User and Multi-User Channels	13
2 DOF for SISO Doubly Selective Channels	15
2.1 Introduction	15
2.1.1 Motivation	15
2.1.2 The State of the Art	16
2.1.3 Contribution	17
2.1.4 Organization	17
2.2 System Model	17
2.3 Basis Expansion Model Representation	18
2.4 Underspread Channels	20

2.5	Overspread Channels	22
2.5.1	Transmission Scheme	22
2.5.2	Conditions for the Existence of the PreLog	24
2.5.3	Analogy with MIMO Systems	25
2.6	Conditional Optimality of the Scheme	26
2.6.1	Optimality for Critically Spread Channels	27
2.6.2	General Condition	27
2.7	Boundaries of Different Regimes of Capacity	28
2.7.1	Boundaries of Logarithmic Scaling	29
2.7.2	Boundaries of Double Logarithmic Scaling	29
2.8	Conclusions	29
2.A	Achievability for Underspread Channels	30
2.B	Upper Bound of MI for Underspread Channels	31
2.C	Achievability for Overspread Channels	33
3	MU MIMO: DOF with no CSI	35
3.1	Introduction	35
3.1.1	Motivation	35
3.1.2	The State of the Art	36
3.1.3	Contribution	36
3.1.4	Organization	37
3.2	System Model	37
3.3	Broadcast channel with No CSIT	39
3.4	TDD Broadcast Channel - Transmission Strategy	40
3.4.1	Uplink Training Phase	41
3.4.2	BS Transmission Strategy: ZF Precoding	42
3.4.3	Downlink Training Phase	43
3.4.4	Coherent Data Phase	44
3.5	TDD Broadcast Channel - DOF Bounds	45
3.5.1	Achievable Rate of the Scheme	45
3.5.2	High SNR DOF of the Achievable Sum Rate	46
3.5.3	Upper Bound of the Multiplexing Gain	48
3.5.4	CSIT Quality Refinement	48
3.6	FDD Broadcast Channel	49
3.6.1	Transmission Strategy	49
3.6.2	High SNR DOF of the Sum Rate	50
3.6.3	UL Data Transmission	51
3.7	Conclusions	51
3.A	MMSE Estimate for DL Training	53

II	CSIT Feedback : Optimization and Acquisition	57
4	Feedback Optimization in MU TDD Systems	59
4.1	Introduction	59
4.1.1	Motivation	59
4.1.2	Contribution	60
4.1.3	The State of the Art	61
4.1.4	Organization	62
4.2	System Model	62
4.3	Transmission Scheme with Oblivious Users	63
4.3.1	Uplink Training Phase	64
4.3.2	Semi-Orthogonal User Selection and ZF Precoding	64
4.3.3	Downlink Training Phase	65
4.3.4	Coherent Data Phase	66
4.3.5	Sum Rate Lower Bound	66
4.4	Transmission Scheme with Informed Users	69
4.4.1	Transmission Scheme	69
4.4.2	Sum Rate Lower Bound	70
4.5	Accuracy of the Approximate Sum Rate Expressions	71
4.6	Asymptotic Analysis with Oblivious Users	73
4.6.1	Noise Limited Regime	73
4.6.2	Interference Limited Regime	74
4.6.3	Asymptotically Large Number of Users	74
4.7	Asymptotic Analysis with Informed Users	74
4.7.1	Noise Limited Regime	75
4.7.2	Interference Limited Regime	75
4.7.3	Asymptotically Large Number of Users	75
4.8	Feedback Load Optimization	76
4.8.1	Optimal Users vs DL SNR	76
4.8.2	Optimal Users vs Channel Coherence Time	78
4.9	Conclusions	80
4.A	Average Power Constrained Users	81
5	Novel CSIT Acquisition for Reciprocal Channels	83
5.1	Introduction	83
5.1.1	Motivation	83
5.1.2	Contribution	84
5.1.3	The State of the Art	85
5.1.4	Organization	86
5.2	System Model and CSIR Acquisition	86

5.3	Classical CSIT Acquisition	87
5.3.1	FDD Systems	87
5.3.2	TDD Systems	88
5.4	Optimal Training and Feedback Combining in TDD Systems	88
5.5	Joint Channel Estimation and Feedback Detection Algorithms	90
5.5.1	Iterative Estimation and Detection	90
5.5.2	Simplified Iterative Estimation and Detection	92
5.5.3	Single-Shot Estimation and Detection	93
5.6	Outage Based Training and Feedback Partitioning	93
5.6.1	Definitions and Initial Setup	93
5.6.2	Resource Split between Training and Feedback	95
5.7	Optimization Setup with Practical Constellations	98
5.7.1	Resource Split for Fixed Constellations	98
5.7.2	Resource Split with Continuous Rates & Parity Bits	99
5.8	Simulation Results	100
5.8.1	Continuous Constellations	100
5.8.2	Fixed Discrete Constellations	102
5.8.3	Discrete Constellations and Coding	103
5.8.4	Imperfect CSIR Analysis	106
5.9	Conclusions	106

III Transmit Power Minimization with User Selection 109

6	Transmit Power Minimization with User Selection	111
6.1	Introduction	111
6.1.1	Motivation	111
6.1.2	The State of the Art	112
6.1.3	Contribution	112
6.1.4	Organization	113
6.2	System Model	113
6.3	Overview of Transmit Power Minimization Problem	114
6.4	Review of User Selection Algorithms	115
6.4.1	Norm-Based User Selection (NUS)	115
6.4.2	Angle-Based User Selection (AUS)	115
6.4.3	Semi-Orthogonal User Selection (SUS)	116
6.5	Transmit Power with User Selection	116
6.5.1	Main Results	116
6.5.2	Performance Comparison	117
6.6	Conclusions	122

6.A	Some Useful Distributions	123
6.B	Proof of Theorem 4	124
6.B.1	Norm-Based User Selection	124
6.B.2	Angle-Based User Selection	125
6.B.3	Semi-Orthogonal User Selection	125
6.B.4	Performance Upper Bound	126
7	Conclusions and Future Perspectives	127
7.1	Conclusions	127
7.2	Future Perspectives	129
	Bibliography	131

List of Figures

2.1	Transmission Scheme Example for Overspread Channels . . .	23
2.2	Channel Matrix with Modified Transmission Scheme	27
3.1	Multi-User MIMO System Model	38
3.2	Transmission Phases for TDD System	40
3.3	Transmission Phases for FDD System	49
3.4	Channel Projections and corresponding ZF vectors for users 1 and 2 in 2-D null space of all other users' channel vectors. .	54
4.1	Transmission Phases for Oblivious Users	63
4.2	Transmission Phases for Informed Users	69
4.3	Sum Rate versus SNR. Approximation expression captures closely the true sum rate vs. SNR behavior for a broad range of number of users feeding back.	72
4.4	Sum Rate versus Number of Users. Approximate expression shows good match with the behavior of the true sum rate vs. number of feeding back users for many different values of SNR.	73
4.5	Optimal Users versus SNR	77
4.6	Sum Rate with Optimal Users versus SNR	78
4.7	Optimal Users versus Coherence Length	79
4.8	Sum Rate with Optimal Users versus Coherence Length . . .	80
5.1	Uplink frame structure: Total feedback length is divided be- tween UL training and quantized feedback phases.	87
5.2	Mean-Square CSIT Errors: $T_{fb} = 20$ and $M = 4$. The novel hybrid scheme performs much better than the classical train- ing based CSIT acquisition. Gains are significant even with naive use of practical constellations without any coding. . . .	101

5.3	Optimal Lengths and Outage Rate: $T_{fb} = 20$ and $M = 4$. With increase in SNR, both the length of the quantized feedback interval T_q and the outage rate b increase gradually. . .	102
5.4	Mean-Square CSIT Errors: $T_{fb} = 20$ and $M = 4$ (a) QPSK and (b) 16-QAM. The novel hybrid scheme with QPSK performs better than the classical one from 9 to 25 dB of SNR, but 16-QAM outperforms both after 21 dB.	104
5.5	Mean-Square CSIT Errors with Convolutional Coding: $T_{fb} = 20$ and $M = 4$. At certain SNR intervals, coding strategy performs better than no coding optimal resource split outcome.	105
5.6	Mean-Square CSIT Errors with Imperfect CSIR: $T_{fb} = 20$ and $M = 4$ (a) QPSK and (b) 16-QAM. For an imperfect CSIR of reasonable quality, the novel scheme performs much better than the classical scheme and the performance approaches to the perfect CSIR case for a good enough CSIR.	107
6.1	Min. Avg. Transmit Power vs. M for $K = 10$, $K_s = 2$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS is the best strategy and follows closely the power lower bound. NUS also performs close to SUS with increasing number of transmit antennas. . .	118
6.2	Min. Avg. Transmit Power vs. Nb. of Users for $M = 4$, $K_s = 2$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS performs the best and NUS becomes sub-optimal when number of users increases.	119
6.3	Min. Avg. Transmit Power vs. M for $K = 8$, $K_s = 4$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS is the best strategy and follows closely the power lower bound. NUS also becomes optimal for a reasonably large number of transmit antennas. .	120
6.4	Min. Avg. Transmit Power vs. Nb. of Users for $M = 4$, $K_s = 4$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS performs the best and NUS becomes strictly sub-optimal when number of users increases.	121

Acronyms

Here are the main acronyms used in this document. The meaning of an acronym is usually indicated once, when it first appears in the text.

AUS	Angle-based User Selection
AWGN	Additive White Gaussian Noise
BEM	Basis Expansion Model
BF	Beamforming
BS	Base Station
CDF	Cumulative Distribution Function
CDI	Channel Direction Information
CQI	Channel Quality Information
CSI	Channel State Information
CSIR	Channel State Information at the Receiver
CSIT	Channel State Information at the Transmitter
DL	Downlink
DMT	Diversity-Multiplexing Tradeoff
DOF	Degrees of Freedom
DPC	Dirty Paper Coding
FB	Feedback
FDD	Frequency-Division Duplex
HDA	Hybrid Digital Analog
IDFT	Inverse Discrete Fourier Transform
ISI	Inter-Symbol Interference
LHS	Left Hand Side
LS	Least Squares
MAC	Multiple Access Channel
MI	Mutual Information
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error

MRC	Maximum Ratio Combining
MSE	Mean Square Error
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MU	Multi-user
NUS	Norm-based User Selection
PDF	Probability Density Function
QAM	Quadrature Amplitude Modulation
OFDM	Orthogonal Frequency Division Multiplexing
ORBF	Orthogonal Random Beam Forming
QoS	Quality of Service
QPSK	Quadrature Phase-Shift Keying
RBF	Random Beamforming
RHS	Right Hand Side
RVQ	Random Vector Quantization
SDMA	Spatial Division Multiple Access
SIC	Successive Interference Cancellation
SIMO	Single-Input Multiple-Output
SINR	Signal-to-Interference-and-Noise Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SU	Single-user
SUS	Semi-orthogonal User Selection
TDD	Time-Division Duplex
UL	Uplink
w.r.t.	with respect to
ZF	Zero Forcing
ZFBF	Zero Forcing Beamforming

Notations

Here is a list of main operators and symbols used in this document. For vectors and matrices, the dimensions are frequently indicated when they appear in the text. We have tried to keep the notation consistent throughout the thesis, but rarely symbols have different definitions in different chapters and in that case they are defined very explicitly to avoid any confusion.

\mathbb{E}	Expectation operator
\otimes	Kronecker product
$(.)^T$	Transpose operation
$(.)^\dagger$	Hermitian operation
\mathcal{CN}	Complex Normal Distribution
\mathbb{R}	Set of real numbers
\mathbb{C}^M	M -dimensional complex space
$\mathcal{H}(\cdot)$	Entropy of the argument
$I(\cdot; \cdot)$	Mutual information between the two arguments
x	Scalar Variable
\mathbf{x}	Vector Variable
\mathbf{X}	Matrix Variable
$ x $	Absolute value of scalar variable
$\lfloor x \rfloor$	Floor operation
$\lceil x \rceil$	Ceil operation
$\ \mathbf{x}\ $	Euclidean norm of vector variable
$\bar{\mathbf{x}}$	Unit norm vector of vector \mathbf{x}
$ \mathbf{X} $	Determinant of the matrix \mathbf{X}
\mathbf{I}_M	Identity matrix of M -dimensions
\mathbf{F}	Partial IDFT matrix
M	Number of transmit antennas
M_r	Number of receive antennas
K	Number of users in the system

N	Length of the transmission block
T	Channel coherence interval (channel uses)
L	Channel delay spread (channel uses)
B_d	Normalized Doppler bandwidth
P	Transmit power constraint
P_u	User's power constraint in DL channel
$h, \mathbf{h}, \mathbf{H}$	Channel (scalar, vector, matrix)
$z, \mathbf{z}, \mathbf{Z}$	Channel noise (scalar, vector, matrix)
$y, \mathbf{y}, \mathbf{Y}$	Observed signal (scalar, vector, matrix)
$\bar{\mathbf{V}}$	Beamforming matrix of normalized columns
σ_h^2	MSE per channel coefficient
ϵ	Outage probability
b	Outage rate
γ	SINR target

Chapter 1

Introduction

The past few decades have seen the rapidest growth of technology and the modernization of infrastructure in the realm of wireless communications. Many wireless systems were standardized, put into effect, later replaced by new systems and newer generations are in the process of standardization. This flurry of research in modern communication systems is the consequence of the quest for being able to transmit large amounts of data reliably. This has been motivated in part by the emergence of novel applications ranging from entertainment providing ones like audio/video transfers, online gaming applications to the medical ones serving to save human lives.

1.1 Evolution in Wireless Communications

For a long time, power and bandwidth were treated as the classical communication resources hence time and frequency were the only available dimensions for user or data multiplexing until the discovery of a new spatial communication dimension which emerges when multiple antennas are employed both at the transmitting side and at the receiving side of the communication link. The proper exploitation of this spatial dimension promises huge gains in spectral efficiency without any extra investment of the classical communication resources. This spatial dimension can be used to get one or more of the following benefits depending upon system/application requirements: a higher spectral efficiency, an increase in communication reliability and

spatial separation of the users. The basic concept of communication system, a single antenna source transmitting data to a single antenna receiver, has evolved with the passage of time with multiple entities transmitting, receiving or both, and each equipped with multiple antennas.

1.1.1 From Single-Antenna to Multi-Antenna Links

In a single-user (SU) single-input single-output (SISO) system, the dominant term of capacity is $\log(P)$ with channel state information at the receiver (CSIR) where P denotes the signal-to-noise ratio (SNR) of the received signal [1], [2], [3]. If the receiver is equipped with multiple antennas, the resulting single-input multiple-output (SIMO) link provides diversity gain to increase the transmission reliability. Multiple receive antennas may also provide power gain. Similarly if the transmitter is equipped with multiple antennas, the resulting multiple-input single-output (MISO) link may provide the diversity or power gain although that might require the availability of the channel state information at the transmitter (CSIT) [3] in certain scenarios.

The presence of multiple antennas only on one end of the communication link may provide diversity/power/array gain or a combination thereof. When multiple antennas are employed at both ends of the communication link, the new spatial communication dimension may yield degree-of-freedom (DOF) gain [3]. This DOF gain can be exploited by spatially multiplexing several data streams on this multi-antenna link. In a SU multiple-input multiple-output (MIMO) system, the spectral efficiency gains of the order of the minimum of transmit and receive dimensions are achievable with the availability of CSIR [4], [5]. Thus for a link having M antennas at both ends, the first order term of the capacity is $M \log(P)$ and the coefficient M of $\log(P)$ is referred to as the pre-log, multiplexing gain or DOF.

1.1.2 Single-User to Multi-User Paradigm Shift

The spatial communication dimension emerges when multiple antennas are used at both ends of a transmission link. This dimension also kicks in when there are multiple users (links) in the system. In a multi-user (MU) communication system where a base station (BS) equipped with M antennas is communicating with K single-antenna users, the spectral efficiency gains of the order of $\min(M, K)$ are achievable as compared to a SISO system operating over the same amount of classical communication resources. This gain is achievable irrespective of the direction of the data transmission whether

it's directed from the BS to multiple users [6], [7], [8], usually termed as the downlink (DL) transmission, or from multiple users to the BS [9], [10], [11], usually termed as the uplink (UL) transmission, conditioned upon the availability of the channel state information (CSI) at the BS.

In the past, wireless systems were mainly carrying voice traffic which is symmetric in UL and DL directions. In the modern era, the data services have dominated the wireless traffic. The highly asymmetric nature of this traffic, requiring extremely high rates in the DL direction, has shifted the multi-user research focus in the same direction. Apart from the spatial multiplexing gain, these DL channels enjoy another gain due to selection possibility over surplus number of users, coined as multi-user diversity [12] benefit. It has been shown in [13], [14] that the sum capacity of the Gaussian broadcast channel has a scaling factor with the number of users as $M \log \log(K)$, where K is the total number of users in the system whose channel information is available at the BS. There are two other advantages of this broadcast channel. It requires mobile users to have single antenna each so user terminals are quite inexpensive and simple. The second advantage is that the channel matrix, in case of a broadcast channel, is much well conditioned as compared to that of a point-to-point MIMO link which is plagued by line-of-sight channel conditions and spatial correlation [15].

1.1.3 The Role of Channel Information

The channel information is always required at the receiver side for coherent data detection. The existence or non-existence of CSIT, on the other hand, may have a very different impact on the performance of a system depending upon its nature. For SU systems, the presence of CSIT may bring at best the power gain over the same system working without CSIT but the spatial multiplexing gain of such a system remains unchanged.

These are MU DL systems where the presence of CSIT plays the cardinal role. A MU DL system, with M -antenna BS and K single-antenna users, has the multiplexing gain of $\min(M, K)$ and the multi-user diversity benefit of $M \log \log(K)$ in the sum rate in the presence of CSIT. For the same system with no CSIT, the multiplexing gain reduces to one as the first order term of the sum capacity is only $\log(P)$ because of the optimality of transmitting to a single user [2], [16], [3]. Furthermore the multi-user diversity benefit is completely lost in the no CSIT system. For $K \geq M$, the CSIT of M users is indispensable to achieve the full multiplexing gain [17], [6] and capturing the multi-user diversity benefit of $M \log \log(K)$ in the sum rate requires the CSIT availability of all of these K users where normally K could be much

larger than M .

For the DL transmission, the requirement of the presence of good quality CSIR is easily met with the help of pilot transmission from the BS to the users. But the acquisition of CSIT at the BS is not that straightforward. If the communication system is operating under time-division duplexing (TDD) mode, the reciprocity can be exploited to get the CSIT by simple pilot transmission on the UL [18], [19]. For the systems operating under frequency-division duplexing (FDD) mode, each receiver needs to estimate its channel first and then feed a reasonable function of this estimate back to the BS (see [20], [17], [21], [22] and the references therein for further details). No matter what the duplexing mode is, the CSIT acquisition consumes the UL communication resources, the amount of which could be highly significant under certain scenarios, hence the CSIT feedback function and the acquisition strategy require a careful design.

1.2 Major Themes in the Thesis

In this section, we give a high-level view of the themes, the major research directions, explored in this thesis. We describe the importance of these problems, show their relevance and raise pertinent questions. Then in the next section, we describe relatively specifically how we have tried to solve these problems highlighting our contributions.

1.2.1 Channel Non-Coherence and DOF Perspective

Since Shannon's landmark paper [1] gave birth to the theory of information, widespread efforts have been being made aimed at unveiling the fundamental rate limits of communication channels. This fundamental rate limit of a communication channel is coined as the channel capacity in the jargon of information theory. Initially CSI was mostly assumed to be perfectly available while computing the capacity (or capacity region) for different channels/systems.

Although sparingly there were research activities dealing with channels without channel knowledge assumption, termed as non-coherent channels, this subject got real attention since the end of last decade. Since then many communication channels have been analyzed and their capacities investigated under this realistic non-coherent scenario. The fundamental importance of this subject stems from the fact that all communication channels are non-coherent in nature and the CSI needs to be estimated, acquired or

fed back at the expense of communication resources. The information theoretic analysis of non-coherent channels indicates the fundamental rate limits without explicitly specifying how these can be approached, further motivating researchers/engineers to develop transmission schemes/strategies to achieve these limits.

Unfortunately, this fundamental information theoretic capacity analysis is intractable save for the most simplistic unrealistic communication channels. This difficulty orients the stream of research to various asymptotic regimes, like low or high signal-to-noise-ratio (SNR) asymptotics or the analysis with asymptotically large number of antennas etc. The asymptotically large SNR analysis of the capacity gives an expansion where $\log(P)$ and $\log \log(P)$ are the dominant terms. The coefficient of $\log(P)$ is often termed as the multiplexing gain, the pre-log or the degrees of freedom (DOF) of the communication channel. The DOF of a channel/system is a very important concept and one of its interpretations is the number of independent streams that can be multiplexed over that channel.

SU multi-antenna block fading non-coherent channels were treated in [23], [24] and later in [25] showing the existence of the $\log(P)$ regime in the high SNR capacity and specifying the pre-log. For symbol-by-symbol stationary channels, the high SNR capacity analysis was conducted in [26], [27], [28], [29]. It turns out that the non-coherent capacity is extremely sensitive to the channel variation mechanism and normally underspread¹ channels show a capacity growth with $\log(P)$ (positive pre-log) whereas overspread channels show only double logarithmic growth with SNR (zero pre-log) rendering the communication highly power inefficient.

For multi-user DL channels, even if each single link is inherently simple and well-studied, the system-level high SNR capacity/throughput analysis might still become intractable due to the necessity of CSI acquisition and the fact that the optimal feedback strategies are yet unknown. On the other hand, MU channels require transmission and reception of independent data streams to/from multiple users, making this setup just in line with the DOF perspective. And hence the question of how many users can be transmitted data simultaneously when CSI is not assumed rather acquired becomes very interesting and relevant.

¹Typical wireless channels are underspread in nature, i.e., the product of the channel coherence time and the coherence bandwidth is much larger than one [29]. See [3] for details.

1.2.2 DL Sum Rate Maximization and CSIT Feedback

In a DL channel having a BS equipped with M transmit antennas and K ($K \geq M$) single antenna users, the first order term of the sum capacity is $M \log(P)$ [6], [30], [17]. The DL channels enjoy another gain due to selection possibility over surplus number of users, coined as multi-user diversity benefit [12], showing the gains increasing with $M \log \log(K)$ [13], [14].

These promising advantages of the broadcast MIMO don't come for free as with no CSIT and perfect CSIR, the first order term of the sum capacity is only $\log(P)$ because of the optimality of transmitting to a single user [2], [16], [3]. Thus the CSIT of M users is required to achieve the full multiplexing gain [17], [6]. Capturing the full multi-user diversity benefit of $M \log \log(K)$ in the sum rate, the CSIT from all of the system users is required.

It is known that for a DL channel with perfect CSI, dirty paper coding (DPC) is the capacity achieving strategy [31] but if CSI is not assumed to be known, there are two basic fundamental questions yet unanswered. First what is the maximum throughput of the system if it has to account for the resource consumption for channel estimation and feedback transmission? Second what are the optimal feedback and data transmission strategies for such a non-coherent system achieving the optimal throughput? These questions become intractable because of the presence of enormous number of underlying sub-problems for which the optimal solutions are not known. The situation gets aggravated with the fact that the optimal solutions to some of the sub-problems might be known if these problems are treated standalone, but when analyzed the whole system, even the optimal solutions for the sub-problems might not hold anymore.

Due to intractability of the most fundamental questions, it is still interesting to study the maximum system throughput setup by introducing the sub-optimal solutions for some of the sub-problems and try to determine suitable/optimal solutions for relatively less studied sub-problems. One such problem formulation could be to determine the amount of feedback which maximizes the system throughput by using known transmission and user scheduling techniques. There is an enormous volume of research publications analyzing the CSIT acquisition techniques and the feedback gains in different scenarios but the primordial issue, which is usually ignored, is the cost of obtaining the feedback at the BS. Similarly the time varying nature of the channel and the finite channel coherence times are not given proper attention while designing feedback schemes. Both the feedback gain and the acquisition overhead increase with the amount of feedback and hence the problem of how much feedback achieves the feedback gain-cost trade-off is

of inherent fundamental nature and of wide practical importance.

It is well-known that to preserve the multiplexing gain of the DL channel with SNR the quality of CSIT is of prime importance and the poor quality CSIT causes the sum rate saturation [17], [32]. That means the future high speed DL data transfers will require very high quality CSIT at the BS and hence a significant portion of the UL communication resources will be used to carry this CSI traffic. Thus a very important sub-problem to the global problem of the system throughput maximization can be formulated as follows: How the CSIT quality can be refined further over the classical CSIT acquisition schemes for fixed resource consumption? This problem has been addressed for FDD systems [20], [17], [22] to some extent and a lot of schemes have been proposed. On the other hand, the only existing CSIT acquisition strategy for TDD systems is through exploitation of the channel reciprocity, i.e., BS gets the CSIT by estimating the UL training sequences transmitted by the users. Due to DL and UL traffic load management flexibility and easy CSIT acquisition, TDD is attracting a lot of attention in standardization groups as a duplexing strategy for upcoming wireless networks. Hence developing novel resource-efficient CSIT acquisition strategies for TDD systems is a problem of utmost practical importance and may have widespread impact on the future implementations of wireless standards.

1.2.3 Transmit Power Minimization with User Selection

In multi-antenna DL systems, the maximization of the sum rate has been widely studied. Conditioned upon the availability of perfect CSI, the capacity region is known and hence the optimal and a wide variety of sub-optimal (but less complicated) transmission strategies have been treated and analyzed. In many practical wireless systems, maximizing the throughput may not be the primary objective. A very important design objective for multi-antenna MU systems is to achieve a particular link quality over all links with minimum transmission power which is equivalent to achieving certain signal-to-interference-and-noise ratios (SINR) or data rates over corresponding links. This problem, in some sense, is the dual problem of the sum rate maximization under a fixed power constraint. Certainly from an operator's perspective, the minimization of average transmit power to achieve these SINR targets is of prime importance.

The problem of the minimization of the DL transmit power required to meet users' SINR constraints by joint optimization of transmit beamforming (BF) vectors and power allocations was solved in [33] and [34] by exploiting the duality of the UL and the DL channels. For Gaussian MU channels

(either UL or DL), they showed that the problem of the minimization of transmit power corresponding to certain SINR targets bears a relatively simple solution due to the added structure which may be exploited by successive interference cancellation (SIC) in UL or by dirty paper coding (DPC) based encoding for known interference in DL channels and the results were presented in [35], [36] and [34]. The authors of [37] studied the problem of transmit power minimization under different user selection algorithms. They studied the Gaussian multi-user system but without exploiting the extra structure of this system through SIC in UL or through DPC in DL. For the case of 2 users transmitted simultaneously, they obtained analytical expressions for the average transmit power required for guaranteed rates with norm-based user selection (NUS) and angle-based user selection (AUS) schemes.

The current state-of-the-art for the joint problem of transmit power minimization and user scheduling raises more questions than the answers it provides. Still open questions in this area include how does the minimum average transmit power decay with the number of users or the number of BS transmit antennas when SIC or DPC is employed. Similarly the optimal user selection scheme for transmit power minimization has never been investigated. In the context of the sum rate maximization, the semi-orthogonal user selection (SUS) has been shown to behave very close to the optimal [38] and is widely believed to be the best greedy user selection strategy but no analytical results for average transmit power are known when SUS is employed. Hence the characterization of the minimum transmit power when SUS is employed and its performance comparison with other selection schemes are very interesting research problems.

1.3 Thesis Outline and Contributions

The research work conducted in this thesis can be divided in three different parts. Part I and II focus primarily on non-coherent SU and MU channels. Part I, comprising chapters 2 and 3, further focuses on high SNR analysis following the degrees of freedom perspective. Part II, comprising chapters 4 and 5, is the analysis of MU DL channels for the sum rate maximization where explicit feedback is taken into account and optimized. Part III, consisting of chapter 6 alone, is the performance comparison of various user selection algorithms for transmit power minimization to achieve certain SINR constraints. In the following paragraphs, we give a brief overview of the dissertation and describe the contributions on a per chapter basis.

Chapter 2 - DOF for SISO Doubly Selective Channels

In this chapter, we consider stationary time- and frequency-selective channels with no assumption of CSI. We investigate the capacity behavior of these doubly selective channels as a function of the channel parameters delay spread, Doppler bandwidth and channel spread factor (the product of the delay spread and the Doppler bandwidth). We study different large SNR capacity regimes dominated either by $\log(P)$ or $\log \log(P)$ depending upon the channel conditions (delay spread, Doppler Bandwidth and channel spread factor). For critically spread channels (channel spread factor of 1), it is widely believed that the dominant term of the high-SNR expansion of the capacity is of order $\log \log(P)$ or in other words, the pre-log is zero. We provide a very simple novel transmission scheme showing that for critically spread channels (even for mildly overspread channels) a non-zero pre-log exists under certain conditions.

The DOF specification for doubly selective channels and the discovery of the scheme showing non-zero pre-log for overspread channels are the main contributions of this chapter. These results have been published in:

- Umer Salim and Dirk Slock, “Asymptotic capacity of underspread and overspread stationary time- and frequency-selective channels”, *Proceedings of Information Theory and Applications Workshop (ITA 2008)*, San Diego, USA, January 28-30, 2008.

Chapter 3 - MU MIMO: DOF with no CSI

A MU DL system is considered with no initial CSI assumption at any of the single-antenna receivers or the BS transmitter. It is shown that with no feedback allowed the DL capacity region is bounded by the capacity of a point-to-point MISO link and hence the multiplexing gain of the DL sum rate is $(1 - 1/T)$ for a block fading channel of coherence length T . When the BS is allowed to acquire CSI, operating under TDD mode, we give a simple transmission scheme through which BS and all users get necessary CSI and the high SNR sum rate shows multiplexing gain of $M[1 - (M+1)/T]$. A tight upper bound to this multiplexing gain is also provided. This analysis reveals the important fact that the CSI quality at the transmitter must be refined with the increase in DL SNR, otherwise the multiplexing capability of the BS due to multiple antennas is lost. When this MU system is working under FDD mode, another simple practically realizable transmission strategy is proposed which provides necessary CSI to both sides with minimal resource utilization and the high SNR DOF for the DL channel are specified. The

given strategy makes the system fully scalable for data transmission in both directions.

The work in this chapter has been published in:

- Umer Salim and Dirk Slock, “Broadcast channel : Degrees of Freedom with no CSIR”, *Proceedings of 46th Allerton Conference on Communication, Control and Computing (Allerton 2008)*, Monticello, Illinois, USA, September 23-26, 2008.
- Umer Salim and Dirk Slock, “Multiuser MIMO downlink : multiplexing gain without free channel information”, *Proceedings of 4th IEEE Broadband wireless access workshop, colocated with IEEE GLOBECOM 2008*, New Orleans, LA, USA, November 30-December 4, 2008.

Chapter 4 - Feedback Optimization in MU TDD Systems

We study a TDD broadcast channel with initial assumption of channel information neither at the BS nor at the users’ side. We give two transmission strategies through which the BS and the users get necessary CSI. A novel lower bound of the sum rate is derived which reflects the rate loss compared to a system with perfect CSIT and the corresponding approximate sum rate expressions are developed for both schemes. These expressions capture fully the benefits of the CSIT feedback, enjoying multi-user diversity gain and better inter-user interference cancellation, and the cost of CSIT feedback. These sum rate expressions, owing to their simplicity, can be optimized for any set of system parameters to unveil the trade-off between the cost and the gains associated to feedback.

The novel problem formulation for the sum rate maximization of a DL channel, with no initial CSI assumption and where CSI acquisition is completely accounted for, and the derivation of a sum rate lower bound capturing the cost-benefit trade-off of feedback are the contributions of this chapter. This work has been published in:

- Umer Salim and Dirk Slock, “How many users should inform the BS about their channel Information?”, *Proceedings of 6th International Symposium on Wireless Communication Systems (ISWCS 2009)*, Siena, Italy, September 7-10, 2009.
- Umer Salim and Dirk Slock, “Transmission Strategies and Sum Rate Maximization in Multi-User TDD Systems”, *Proceedings of IEEE Global Communications Conference (Globecom 2009)*, Hawaii, USA, November 30-December 4, 2009.

and submitted as:

- Umer Salim and Dirk Slock, “How much FEEDBACK is Required for TDD Multi-Antenna Broadcast Channels with User Selection?”, *EURASIP Journal on Advances in Signal Processing*, (under revision).

Chapter 5 - Novel CSIT Acquisition Strategy for Reciprocal Channels

The CSIT acquisition for MU reciprocal systems is the problem treated in this chapter for which the channel estimation at the BS making use of the UL pilot sequences forms the classical CSIT acquisition strategy. We show that this traditional TDD setup fails to fully exploit the channel reciprocity in its true sense and the system can benefit from a combined CSIT acquisition strategy mixing the use of limited feedback and that of a training sequence. This combining gives rise to a very interesting joint estimation and detection problem for which two iterative algorithms are proposed. An outage rate based framework is also developed which gives the optimal resource split between training and feedback. The potential of this hybrid combining is demonstrated in terms of the improved CSIT quality under a global training and feedback resource constraint.

The novel hybrid CSIT acquisition strategy for MU reciprocal channels is the major contribution of this chapter. The other contributions include the proposition of iterative algorithms for the joint estimation and detection problem with the convergence proof, and the outage rate based framework which gives the optimal resource split between training and quantized feedback. The work in this chapter has been published in:

- Umer Salim, David Gesbert, Dirk Slock and Zafer Beyaztas, “Hybrid pilot/quantization-based feedback in multi-antenna TDD systems”, *Proceedings of IEEE Global Communications Conference (Globecom 2009)*, Hawaii, USA, November 30-December 4, 2009.

and submitted as:

- Umer Salim, David Gesbert and Dirk Slock, “Combining Training and Quantized Feedback in Multi-Antenna Reciprocal Channels”, *IEEE Transactions on Wireless Communications*, (under revision).

Chapter 6 - Transmit Power Minimization with User Selection

In this chapter, we study the joint problem of the minimization of the average transmit power at the BS with user selection. The analytical results

for the minimum average transmit power to attain specific SINR targets are almost non-existent due to the intricate intertwined structure of the beam-forming vectors of all users and the corresponding power allocation scalars, the situation becoming even more complicated with the introduction of a user selection process. Nevertheless, we characterize analytically the average transmit power required to meet guaranteed performance with various user selection algorithms, namely SUS, NUS and AUS, in case when only 2 users are selected for simultaneous transmission. The SUS performs better than the other presented selection algorithms due to its better user selection mechanism. When more users are selected for simultaneous transmission, the performance of SUS improves further relative to NUS and AUS.

The performance comparison of various user selection algorithms for the objective of average transmit power minimization and the derivation of analytical results, although under certain restrictions, are the contributions of this chapter. These results have been published in:

- Umer Salim and Dirk Slock, “Performance of Different User Selection Algorithms for Transmit Power Minimization”, *Proceedings of 43rd Asilomar Conference on Signals Systems and Computers (Asilomar 2009)*, California, USA, November 1-4, 2009.

In thesis writing, there are two somewhat contradictory objectives that the authors like to achieve, i) avoiding repetition ii) having self-contained chapters. Although both of these conflicting themes have their own merits, we have certainly given preference to the second objective. Hence we have tried to make self-contained chapters with as little repetition as possible.

Part I

DOF for Single-User and Multi-User Channels

Chapter 2

DOF for SISO Doubly Selective Channels

2.1 Introduction

2.1.1 Motivation

Information theoretic capacity analysis for different types of channel models started with somewhat unusual assumption that the channel is perfectly known at the receiver or sometimes even assuming that the channel is known at the transmitter. Inherently all channels are non-coherent in nature and hence channel estimation is required to get CSIR and then some kind of feedback (for FDD systems) or estimation (for reciprocal TDD systems) may provide CSIT. The area of capacity analysis for non-coherent fading channels has received considerable attention in recent years as it gives fundamental limits of data communication without any CSI assumption. These fundamental capacity bounds for non-coherent channels serve as a benchmark for judging the efficiency of all transmission schemes which make explicit use of pilots or training sequences to estimate the channel and later use these estimates for data detection. In this chapter, we deal with single-user SISO channels where there is no feedback link and hence CSIT is non-existent throughout the transmission. Furthermore the focus is kept on high SNR regime where pre-log, the coefficient of $\log(P)$, and pre-loglog, the coefficient

of $\log \log(P)$, become the important capacity determining parameters.

2.1.2 The State of the Art

Usually block fading models have been assumed for obtaining the capacity bounds in the no CSIR case. In the standard version of this model [23], the fading remains constant over a block of T symbol periods, and changes independently from one block to another. Capacity bounds are obtained by introducing training segments in an ad hoc fashion. For the standard block fading model, the capacity is shown [23], [24] to grow logarithmically with SNR at large values of SNR, thus $\log(P)$ was shown to be the dominant term of capacity. Later Liang and Veeravalli [25] allowed the fading to vary inside the block with a certain correlation matrix characterized by its rank Q and showed for SISO channels that the capacity pre-log is $(1 - Q/T)$. For block constant frequency selective channels with L taps, the pre-log was shown to be $(1 - L/T)$ in [39].

Non-coherent capacity has also been analyzed with the channel fading process taken to be symbol-by-symbol stationary. In this model, fading is not independent but time selective without any block structure. Surprisingly, this model leads to very different capacity results: contrary to $\log(P)$ capacity growth in block fading channels, here the capacity grows only double logarithmically with SNR at asymptotically large SNR [26], [40], [27] when the fading process is non-bandlimited (the Doppler spectrum spans the full transmission bandwidth); in this case the channel prediction error is non-zero even if the infinite channel past is known.

For symbol-by-symbol stationary Gaussian fading channels, if the Doppler spectrum is bandlimited (of limited support), then the fading process is called non-regular and the prediction error given the infinite past goes to zero. Lapidoth [28] studied the SISO case for this kind of fading processes showing that the capacity grows logarithmically with SNR and the capacity pre-log is the Lebesgue measure of the frequencies where the spectral density of the fading process (Doppler spectrum) has nulls.

Etkin and Tse [29] study the same channel model of bandlimited fading for MIMO systems; they show that the pre-log exists even for MIMO systems with no CSIR for which they provide a lower bound of the MIMO capacity pre-log.

2.1.3 Contribution

The channels of interest in this chapter are non-coherent SISO doubly selective channels where the channel has multiple taps varying in time as stationary processes characterized by a Doppler spectrum. Their coherent counterparts have a pre-log of one. For such non-coherent channels under strict underspread assumption, we show that the loss in pre-log is equal to the spread factor of the channel (the product of the delay spread and the Doppler bandwidth). This result is not counter-intuitive as channel spread in time or frequency introduces more channel parameters that need to be estimated for coherent detection of the data. This result implies that the pre-log should be zero when channel spread factor becomes one but we present a simple scheme which shows the existence of $\log(P)$ regime for overspread channels. The channel conditions are specified which govern the range of existence of this regime. At higher channel spread factors, the $\log(P)$ term vanishes and a $\log \log(P)$ term becomes the dominant capacity term, the range of which is also specified.

2.1.4 Organization

After describing the system model in section 2.2, its representation using a basis expansion model (BEM) is given in section 2.3. Section 2.4 presents the capacity analysis for underspread channels. In section 2.5, a simple transmission scheme is introduced showing the existence of the pre-log for overspread channels with the associated conditions for the existence of this pre-log and an analogy of the scheme from antenna deactivation strategy for frequency-flat MIMO channels is also provided. In section 2.6, we discuss the optimality of our transmission scheme. Then in section 2.7, we specify the boundaries of the high SNR capacity regimes of $\log(P)$ and $\log \log(P)$. The chapter ends with some concluding remarks in section 2.8.

2.2 System Model

We consider a discrete-time SISO fading channel at symbol rate, having L taps whose time- n output $y[n] \in \mathbb{C}$ is given by

$$y[n] = \sqrt{P} \sum_{l=0}^{L-1} h[n, l] x[n-l] + z[n] \quad (2.1)$$

where $x[n] \in \mathbb{C}$ denotes the time- n channel input, the complex scalar $h[n, l] \in \mathbb{C}$ represents the l -th coefficient of the FIR (finite impulse response) channel filter at time n consisting of circularly symmetric complex Gaussian components of zero mean and unit variance, and $z[n] \in \mathbb{C}$ denotes the additive white Gaussian noise. Here \mathbb{C} denotes the complex field. The system is normalized so that the channel input has an average power constraint of $\mathbb{E}[|x[n]|^2] \leq 1$.

The channel fading process $\{h[n, l]\}$ for each tap l is assumed to be stationary, ergodic and bandlimited. They are independent and identically distributed (i.i.d.) across different taps l . The hypothesis of the bandlimitedness of the fading process is motivated by the physical limitations on mobile speeds. For a mobile speed v , the maximum Doppler frequency magnitude f_{max} for each path is $f_{max} = v/\lambda_c$ where λ_c is the carrier wavelength. The bandwidth of each fading process will be upper bounded by the two-sided Doppler bandwidth $2f_{max}$. We define the normalized Doppler bandwidth as $B_d = 2f_{max}T_s$ where T_s represents the symbol period, assuming the Doppler spectrum has support between the two extreme Doppler shifts. In general, B_d will denote the support of the Doppler spectrum. The hypothesis of bandlimited Doppler spectrum is an approximation because the Doppler shifts do not remain constant. Similarly, the hypothesis of limited delay spread is an approximation. Limited values for Doppler and delay spreads can be justified at a given working SNR.

The capacity pre-log is normally defined as

$$\text{PreLog} = \lim_{P \rightarrow \infty} \frac{C(P)}{\log(P)} \quad (2.2)$$

whenever the capacity $C(P)$ is of order $\log(P)$, and the capacity pre-loglog is given by

$$\text{PreLogLog} = \lim_{P \rightarrow \infty} \frac{C(P)}{\log \log(P)} \quad (2.3)$$

whenever $C(P)$ is of order $\log \log(P)$.

2.3 Basis Expansion Model Representation

We shall assume here, without loss of generality, that the Doppler spectrum is contiguous and that the demodulation is synchronized to the lower edge of the Doppler Spectrum. To get a proper model for the doubly selective channel, we start by considering block transmission with block length N . Continuous transmission results will then be obtained by letting the block

size N grow to infinity. Observing a signal over a block can always be thought of as if the block considered is one period of a periodic process, in which case the signal has a Fourier series expansion. This leads to a Basis Expansion Model (BEM) for the time-varying channel coefficients in which the basis functions are complex exponentials with frequencies at the multiples of $1/N$ [41]. As the Doppler spectrum is bandlimited, we shall take the BEM to be correspondingly bandlimited. We should note here that we do not necessarily demand of the BEM to provide an exact description of the channel statistics over the block of length N , as long as the description becomes exact as the block length tends to infinity. The BEM leads to the following representation for the channel coefficients over a block that starts at time zero w.l.o.g.,

$$h[n, l] = \sum_{m=0}^{N_d-1} \Psi[m, l] e^{j2\pi nm/N}, \quad n = 0, 1, \dots, N-1 \quad (2.4)$$

where $N_d = \lceil N B_d \rceil$. In the above equation, $\Psi[m, l]$ are independent, uncorrelated, zero mean proper complex Gaussian random variables whose variances are the values of the spectrum of the corresponding fading process at the respective frequencies m/N . If the block transmission is alternatively thought of as an isolated block (instead of a period of a periodic process), then the windowing in time domain with a rectangular block of size N leads to an interpolation in frequency domain between the frequencies m/N with $\frac{\sin \pi N f}{N \sin \pi f}$ which leads to something non-bandlimited, as indeed a signal cannot be both time- and bandlimited. However, the process becomes bandlimited as the block size N tends to infinity (see also [25]). To avoid inter-block interference and facilitate the description in the frequency-domain, we add a cyclic prefix of length $L-1$ making the total block length $N+L-1$. At the receiver the first $L-1$ received samples corresponding to the prefix are neglected and the remaining N outputs, the inputs and the noise get collected in vector form as $\mathbf{y} = [y[0] y[1] \dots y[N-1]]^T$, $\mathbf{x} = [x[0] x[1] \dots x[N-1]]^T$, $\mathbf{z} = [z[0] z[1] \dots z[N-1]]^T$, leading to the system equation

$$\mathbf{y} = \sqrt{P} \mathbf{H} \mathbf{x} + \mathbf{z} \quad (2.5)$$

where $\mathbf{H} \in \mathbb{C}^{N \times N}$ is the channel matrix for this block and has the circulant structure due to the addition of cyclic prefix.

We also need a system representation in which the roles of channel and input are reversed. For this, we define a diagonal matrix $\mathbf{X}_i = \text{diag}(x[i], x[i+1], \dots, x[i+N-1])$ and $\mathbf{X} = [\mathbf{X}_0 \mathbf{X}_{-1} \dots \mathbf{X}_{-(L-1)}]$. Hence $\mathbf{X} \in \mathbb{C}^{N \times NL}$ is

$$\mathbf{H} = \begin{bmatrix} h[0,0] & & h[0,L-1] & \dots & h[0,1] \\ \vdots & h[1,0] & & \ddots & \vdots \\ \vdots & & & & h[L-2,L-1] \\ h[L-1,L-1] & \vdots & & & \\ & h[L,L-1] & \ddots & & \\ & & \ddots & & \\ & & & & h[N-1,0] \end{bmatrix}$$

the system input for one block of length N . If $\mathbf{h}_l = [h[0,l] \ h[1,l] \ \dots \ h[N-1,l]]^T$ and $\mathbf{h} = [\mathbf{h}_0^T \ \mathbf{h}_1^T \ \dots \ \mathbf{h}_{L-1}^T]^T$ then eq. (2.5) can be written as

$$\mathbf{y} = \sqrt{P} \mathbf{X} \mathbf{h} + \mathbf{z}. \quad (2.6)$$

Similar to channel components, by putting the uncorrelated coefficients of BEM in vectors $\Psi_l = [\Psi[0,l] \ \dots \ \Psi[N_d,l]]^T$, eq. (2.4) takes the form of $\mathbf{h}_l = \mathbf{F} \Psi_l$ where $\mathbf{F} \in \mathbb{C}^{N \times N_d}$ is the (partial) IDFT matrix whose element at n -th row and m -th column is given by $e^{j2\pi nm/N}$. By regrouping BEM coefficients of all channel taps in a vector $\Psi = [\Psi_0^T \ \Psi_1^T \ \dots \ \Psi_{L-1}^T]^T$, we can write

$$\mathbf{h} = \mathbf{F}_c \Psi, \quad (2.7)$$

where $\mathbf{F}_c = \mathbf{I}_L \otimes \mathbf{F}$ with \mathbf{I}_L denoting an L -dimensional identity matrix and \otimes represents the Kronecker product. With this eq. (2.6) can be written as

$$\mathbf{y} = \sqrt{P} \mathbf{X} \mathbf{F}_c \Psi + \mathbf{z}. \quad (2.8)$$

2.4 Underspread Channels

Typically wireless channels are underspread in nature [3], so first of all we study the capacity pre-log for doubly selective channels when they are underspread (the product of the delay spread and the normalized Doppler bandwidth is strictly less than one). We derive a lower and an upper bound for the mutual information (MI) between the input \mathbf{x} and the output \mathbf{y} , denoted as $I(\mathbf{x}; \mathbf{y})$, of non-coherent doubly selective channels for which a BEM was developed in the previous section. These bounds allow us to specify the pre-log for such non-coherent underspread channels.

Lemma 1 (Lower Bound of Mutual Information). *For an underspread doubly selective channel with delay spread of L symbol intervals and where each channel tap can be represented using N_d independent BEM coefficients over a block length of N symbol intervals, the following lower bound of the mutual information at large SNR is achievable using Gaussian i.i.d. inputs.*

$$\lim_{P \rightarrow \infty} \frac{1}{N} I(\mathbf{x}; \mathbf{y}) \geq \left(1 - \frac{LN_d}{N}\right) \log(P) + O(1) \quad (2.9)$$

where $O(1)$ represents a term which does not grow with SNR.

Proof. The proof of this lemma appears in Appendix 2.A. \square

Lemma 2 (Upper Bound of Mutual Information). *For the same underspread doubly selective channel, an upper bound of the mutual information at large SNR can be characterized as:*

$$\lim_{P \rightarrow \infty} \frac{1}{N} I(\mathbf{x}; \mathbf{y}) \leq \left(1 - \frac{LN_d}{N}\right) \log(P) + o(\log(P)) \quad (2.10)$$

where $o(\log(P))$ indicates that there might be lower order terms which depend upon SNR (such as a $\log(\log(P))$ term) but they are negligible as compared to $\log(P)$ at very large values of SNR.

Proof. The proof of this lemma appears in Appendix 2.B. \square

Theorem 1 (Pre-Log for Underspread Doubly Selective Channels). *For a doubly selective underspread channel having a delay spread of L symbol intervals and normalized Doppler bandwidth of B_d , the pre-log is given by:*

$$\text{PreLog} = 1 - LB_d \quad (2.11)$$

Proof. Using the results from the previous two lemmas about the lower and upper bounds on the mutual information of strictly underspread channels, one can conclude that the pre-log is given by

$$\text{PreLog} = \left(1 - \frac{LN_d}{N}\right). \quad (2.12)$$

Now we can let the block length N to go to infinity. The factor N_d which is the total number of Fourier coefficients required to describe a single channel tap over block length N has its dependence upon N and the limiting value of N_d/N with large block length turns out to be $2f_{max}T_s$, a quantity

we described as the normalized Doppler bandwidth in section 2.2. So the capacity pre-log for underspread channels becomes $1 - LB_d$. It shows that the loss in pre-log for a non-coherent SISO channel is equal to the channel spread factor which is the average number of channel parameters per symbol interval that can parameterize the channel. \square

2.5 Overspread Channels

In this section we treat the case of a channel which is overspread. Hence the channel spread factor (the product of the delay spread of the channel and the normalized Doppler bandwidth) is greater than one which would imply that the pre-log obtained in the previous section for such doubly selective channels ($1 - LB_d$) has already become zero. In fact according to the pre-log expression of $(1 - LB_d)$, the pre-log will become zero as soon as the channel is critically spread ($LB_d = 1$). Below we give a very simple scheme which shows that the $\log(P)$ term exists for overspread channels under certain conditions.

2.5.1 Transmission Scheme

Our transmission scheme to realize $\log(P)$ growth for overspread channels is based upon zero padding. The zero padding is done in such a manner that at the receiver side, each transmitted symbol appears without inter-symbol interference (ISI) for at least one symbol interval. To achieve one output sample free of ISI corresponding to each input symbol, each information symbol is followed by $\lfloor L/2 \rfloor$ deterministic zeros. Hence each information symbol gets received with no ISI at $(\lfloor L/2 \rfloor + 1)$ -th symbol instant after its transmission. For this scheme $\lfloor L/2 \rfloor$ input symbols are wasted (zero-padded) corresponding to each single information symbol but focusing on ISI free output samples gives the advantage that the effective channel becomes frequency flat and each information symbol comes multiplied with the same channel tap, the $(\lfloor L/2 \rfloor + 1)$ -th tap. This scheme is explained in Figure 2.1.

Now we need to see what fraction of symbols we are able to transmit in this zero-padded scheme where $\lfloor L/2 \rfloor$ symbols get wasted for each single information symbol. So the fraction of the information symbols is

$$n_{tx} = \frac{1}{\lfloor L/2 \rfloor + 1} \quad (2.13)$$

Now keeping in mind that here we are interested in only a single channel tap (which appears with ISI free output symbol) requiring N_d BEM coefficients

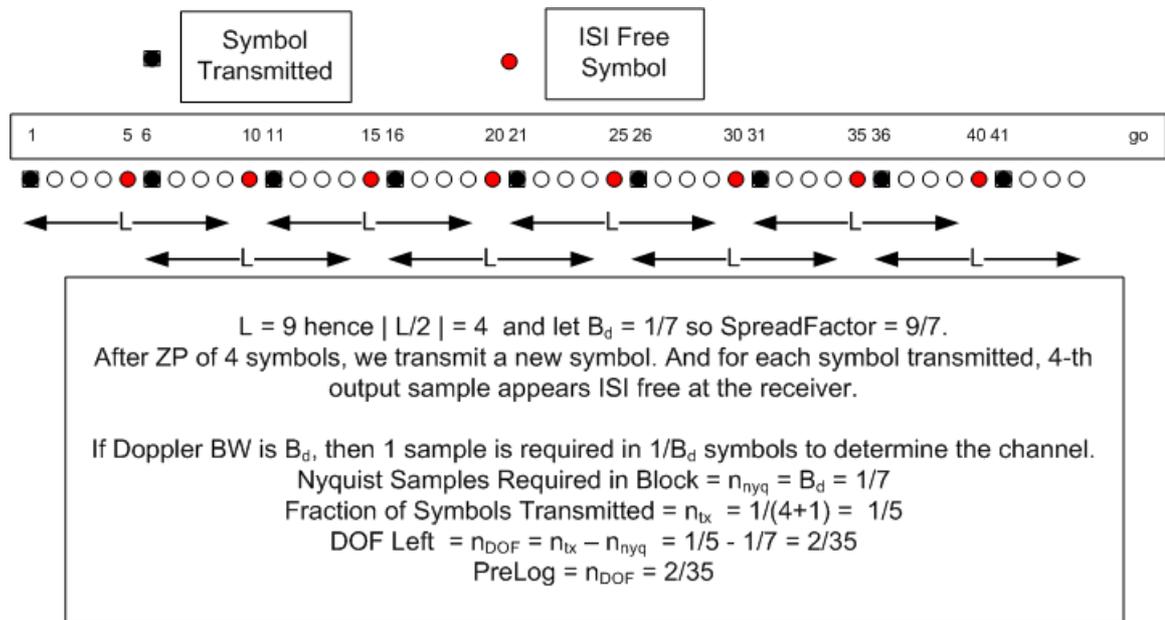


Figure 2.1: Transmission Scheme Example for Overspread Channels

to be estimated to be fully known over a block length N as we argued in section 2.3. And to estimate a single channel tap, per symbol coefficients required N_d/N was shown to be equal to the normalized Doppler bandwidth B_d in section 2.4. We denote this fraction by n_{nyq} , the minimum number of samples required to estimate the channel

$$n_{nyq} = \lim_{N \rightarrow \infty} \frac{N_d}{N} = B_d \quad (2.14)$$

If we want to estimate the channel by sending pilot symbols, we need to transmit B_d fraction of pilots among the non-zero transmit symbols and then this particular channel tap can be estimated by estimating its BEM coefficients. But in this scheme, the total number of information symbols transmitted is the fraction $1/(\lfloor L/2 \rfloor + 1)$. Now there is the possibility that some degrees of freedom (DOF) are left even after estimating this particular channel tap but that will depend upon the relative values of channel delay spread L and normalized Doppler bandwidth B_d .

$$n_{DOF} = n_{tx} - n_{nyq} = \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} - B_d \quad (2.15)$$

So we can have coherent transmission albeit with imperfect channel estimate over this fraction n_{DOF} (if this fraction is non-negative!) and so it corresponds to a coherent channel with positive pre-log. Hence pre-log per symbol interval is given by

$$\text{PreLog} = n_{DOF} = \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \left(1 - B_d (\lfloor \frac{L}{2} \rfloor + 1) \right). \quad (2.16)$$

Formal information theoretic proof for the achievability of the above pre-log for overspread channels has been given in Appendix 2.C. Although we don't have a proof for the upper bound of the pre-log for this transient regime but we conjecture that the pre-log achievable by this zero-padding scheme is the pre-log for this regime.

2.5.2 Conditions for the Existence of the PreLog

First of all, the Doppler spectrum should not be of full support i.e. the normalized Doppler bandwidth should be less than one $B_d \leq 1$. If the normalized Doppler bandwidth is one, even for frequency flat channels, channel prediction becomes impossible hence coherent regime can never come into play and the $\log(P)$ term does not exist [28].

The zero-padding scheme gives more strict restriction on the normalized Doppler bandwidth than $B_d \leq 1$. In our zero-padded transmission scheme, we transmit a fraction $1/(\lfloor L/2 \rfloor + 1)$ number of symbols and the fractional number of Nyquist samples required for minimal channel representation is B_d . Hence the number of transmitted symbols over any block length should be greater than Nyquist symbols required to have some positive DOF where coherent operation can be carried out to obtain capacity growth with $\log(P)$. So this gives us the condition

$$B_d \leq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1}. \quad (2.17)$$

We can find out the channel parameter values where the pre-log given by the zero-padded transmission scheme surpasses the pre-log $(1 - LB_d)$ derived in section 2.4. This gives us a lower bound on the normalized Doppler bandwidth. Combining this lower bound with the upper bound given above, we get

$$\frac{\lfloor \frac{L}{2} \rfloor}{(\lfloor \frac{L}{2} \rfloor + 1)(L - 1)} \leq B_d \leq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1}. \quad (2.18)$$

The left inequality shows the condition for an underspread channel where the pre-log of this zero-padding scheme takes over the classical pre-log of $(1 - LB_d)$ and the right inequality shows the condition under which an overspread channel shows positive pre-log with this scheme. The multiplication of the above inequality with L gives us the corresponding bounds on the channel spread factor.

2.5.3 Analogy with MIMO Systems

Suppose we are working with a MIMO system having M transmit and M_r receive antennas. And each channel coefficient of $M_r \times M$ MIMO matrix is frequency flat and has normalized Doppler bandwidth of B_d . When the channel is largely underspread ($B_d < 1$), from [32] the pre-log of this MIMO system can be represented as

$$\text{PreLog} = M'(1 - M'B_d), \quad (2.19)$$

where M' is given by the following expression

$$M' = \min \left(M, M_r, \frac{1}{2B_d} \right). \quad (2.20)$$

M' is in fact the optimal number of transmit antennas which need to be activated to get this pre-log corresponding to the capacity of this channel. The active number of transmit antennas depends upon the channel spread factor (which is equal to the normalized Doppler bandwidth for frequency flat channels) in a fashion that as channel spread factor increases, one needs to activate lesser and lesser number of transmit antennas.

The reason for this deactivation of transmit antennas is that the existence of $\log(P)$ dominant regime requires coherent data detection and hence channel coefficients need to be estimated. Although the deactivation of transmit antennas reduces the spatial signaling dimensions, but so do the number of channel parameters which need to be estimated making this optimal at relatively higher spread factors. The same reasoning makes our scheme work where we sacrifice temporal signaling dimensions to reduce the number of active channel parameters, getting the benefit of coherent detection and resultantly positive/higher pre-log.

2.6 Conditional Optimality of the Scheme

In our transmission scheme with zero padding, we transmit one information symbol in each block of $(\lfloor L/2 \rfloor + 1)$ symbols. One can argue if more than one symbol is transmitted and zero padding of the same size is done, there might be the possibility of having more DOF and resultantly a higher pre-log factor. In Figure 2.2, we explain this modified transmission scheme and develop generalized pre-log expressions when more symbols are transmitted and from this analysis we show the optimality of the scheme proposed in section 2.5.1.

Suppose we choose to transmit s symbols and do zero padding of $\lfloor L/2 \rfloor$ symbols so the fraction of the symbols transmitted is

$$n_{tx} = \frac{s}{\lfloor L/2 \rfloor + s} \quad (2.21)$$

From the figure, we see that if s symbols are transmitted, to detect these s symbols at the receiver involves the estimation of at least $2s - 1$ channel taps. For one channel tap, the BEM coefficients required per symbol interval is B_d . Hence for our case where we have $2s - 1$ channel taps involved, the number of observations required is

$$n_{nyq} = (2s - 1)B_d \quad (2.22)$$

So DOF or the pre-log (the number of symbols available for coherent detection after estimating the minimum required Nyquist samples) is given

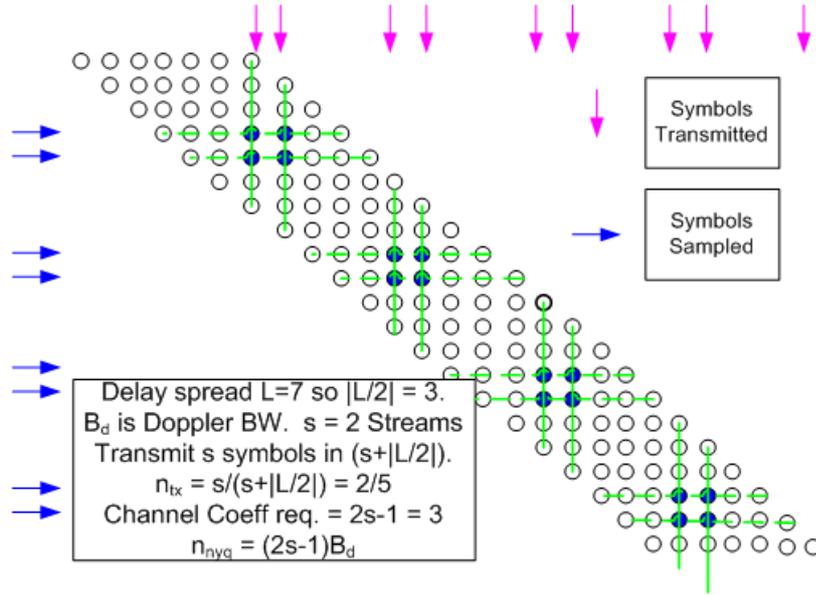


Figure 2.2: Channel Matrix with Modified Transmission Scheme

by

$$\text{PreLog} = n_{tx} - n_{nyq} = \frac{s}{\lfloor L/2 \rfloor + s} - (2s-1)B_d. \quad (2.23)$$

2.6.1 Optimality for Critically Spread Channels

The above expression of pre-log can be specialized to critically spread channels (the spread factor of 1) which gives $B_d = 1/L$. Hence in that case, the pre-log is given by

$$\text{PreLog} = \begin{cases} \frac{L-(2s-1)^2}{2L(\lfloor L/2 \rfloor + s)} & \text{for } L \text{ odd integer} \\ \frac{L-2s(2s-1)}{2L(\lfloor L/2 \rfloor + s)} & \text{for } L \text{ even integer} \end{cases}$$

This expression of pre-log gets maximized for $s = 1$ which gives the transmission scheme given in section 2.5.1 hence proving the optimality of our scheme at spread factor of 1 among other similar zero padded schemes.

2.6.2 General Condition

If the scheme with s ($s \geq 2$) streams is better than zero-padding scheme described in the previous section, then its pre-log should be higher than the

pre-log of that scheme eq. (2.16) which gives us the following condition after some manipulation

$$B_d \leq \frac{1}{2(\lfloor \frac{L}{2} \rfloor + 1)} \frac{\lfloor \frac{L}{2} \rfloor}{\lfloor \frac{L}{2} \rfloor + s}. \quad (2.24)$$

On the other hand, the pre-log of the scheme with s streams should also be higher than the conventional pre-log of $1 - LB_d$ which gives another condition on B_d

$$B_d \geq \frac{1}{L + 1 - 2s} \frac{\lfloor \frac{L}{2} \rfloor}{\lfloor \frac{L}{2} \rfloor + s}. \quad (2.25)$$

Combining the above two inequalities with some algebra, we get the following condition

$$\frac{1}{L + 1 - 2s} \leq B_d \left(\frac{\lfloor \frac{L}{2} \rfloor + s}{\lfloor \frac{L}{2} \rfloor} \right) \leq \frac{1}{2(\lfloor \frac{L}{2} \rfloor + 1)}. \quad (2.26)$$

If we pick the terms on the extreme left and the extreme right of the above inequalities, we get

$$s \leq \frac{L}{2} - \lfloor \frac{L}{2} \rfloor - \frac{1}{2}, \quad (2.27)$$

which is false for any value of $s \geq 2$. This shows that the pre-log for any scheme with $s \geq 2$ information symbols and zero padding of $\lfloor L/2 \rfloor$ symbols never beats the conventional pre-log of $1 - LB_d$ and the pre-log of zero-padding scheme given in eq. (2.16) at the same time, hence proving the conditional optimality of the scheme among other such schemes.

2.7 Boundaries of Different Regimes of Capacity

In this section, we characterize the boundaries of logarithmic capacity regime (where capacity scales with $\log(P)$) and double logarithmic regime (where capacity scales with $\log \log(P)$). We showed the optimality of our zero padding transmission scheme among other schemes which may employ zero padding in the previous section. This scheme is an extreme case of zero padding and corresponds to the worst case scenario with higher spread factors so the boundaries of different capacity regimes like $\log(P)$ and $\log \log(P)$ will be the same as given by this scheme.

2.7.1 Boundaries of Logarithmic Scaling

For underspread channels, the dominant term of capacity is $\log(P)$. From our scheme which we explained in the previous sections, we conclude that the $\log(P)$ regime will be there as long as the following condition is satisfied.

$$B_d \leq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \quad (2.28)$$

It is important to mention that even overspread channels might satisfy this condition and in that case show capacity growth with $\log(P)$.

2.7.2 Boundaries of Double Logarithmic Scaling

The regime where the dominant term of capacity is $\log \log(P)$ starts when the logarithmic regime $\log(P)$ ends. Now we want to know when this double logarithmic regime also ceases to exist. According to our transmission scheme, no matter how large is the delay spread, we just do zero padding in a manner such that at the receiver side, we get at least one ISI free sample and the information symbols can be thought of passing through a frequency flat channel. As energy detection (non-coherent) of frequency flat channels gives $\log \log(P)$ growth of capacity [26], this $\log \log(P)$ regime will only stop when the channel faces an infinitely long delay spread. And when this is the case, high SNR capacity will be bounded giving no increase with increasing SNR. This result has been rigorously proved later in [42].

2.8 Conclusions

In this chapter, we have derived the pre-log expression for underspread time- and frequency selective channels. We proved the existence of $\log(P)$ regime of capacity growth for overspread channels under certain conditions of the delay spread and the Doppler bandwidth of the channel with the help of a very simple transmission scheme utilizing zero padding. The optimality of this novel scheme is shown over other schemes employing zero padding for overspread channels. We have specified the boundary where $\log(P)$ regime converts to $\log \log(P)$ regime. It is further indicated that for infinite delay spread channels, even this double logarithmic regime ceases to exist.

2.A Achievability for Underspread Channels

To show achievability, we select Gaussian i.i.d. inputs, denoted as $\mathbf{x}^{\mathbf{G}}$. The mutual information between the input and the output of the doubly selective channel eq. (2.5) over the block length N is given by

$$\begin{aligned} I(\mathbf{x}^{\mathbf{G}}; \mathbf{y}) &= I(\mathbf{x}^{\mathbf{G}}, \mathbf{H}; \mathbf{y}) - I(\mathbf{H}; \mathbf{y} | \mathbf{x}^{\mathbf{G}}) \\ &= I(\mathbf{x}^{\mathbf{G}}; \mathbf{y} | \mathbf{H}) + I(\mathbf{H}; \mathbf{y}) - I(\mathbf{H}; \mathbf{y} | \mathbf{x}^{\mathbf{G}}) \\ &\geq I(\mathbf{x}^{\mathbf{G}}; \mathbf{y} | \mathbf{H}) - I(\mathbf{H}; \mathbf{y} | \mathbf{x}^{\mathbf{G}}). \end{aligned} \quad (2.29)$$

Equalities here follow by the introduction of the channel matrix \mathbf{H} and by using the chain rule of mutual information [2] multiple times and the inequality follows from the non-negativity of the mutual information [2].

First term in the above inequality is the mutual information when the channel is known and can be evaluated readily

$$I(\mathbf{x}^{\mathbf{G}}; \mathbf{y} | \mathbf{H}) = \mathcal{H}(\mathbf{y} | \mathbf{H}) - \mathcal{H}(\mathbf{y} | \mathbf{H}, \mathbf{x}^{\mathbf{G}}) = \mathcal{H}(\mathbf{y} | \mathbf{H}) - \mathcal{H}(\mathbf{z}).$$

Here we use the definition of mutual information as the difference of differential entropies [2], where $\mathcal{H}(\cdot)$ denotes the differential entropy of its argument. As the input $\mathbf{x}^{\mathbf{G}}$ is i.i.d. Gaussian so \mathbf{y} given \mathbf{H} is also Gaussian distributed with zero mean and a covariance of $\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger | \mathbf{H}] = P\mathbf{H}\mathbf{H}^\dagger + \mathbf{I}_N$, so

$$I(\mathbf{x}^{\mathbf{G}}; \mathbf{y} | \mathbf{H}) = \mathbb{E} \log |P\mathbf{H}\mathbf{H}^\dagger + \mathbf{I}_N|. \quad (2.30)$$

$\mathbf{H}\mathbf{H}^\dagger$ will be a full rank matrix due to its block diagonal structure and Gaussian entries, hence at high SNR, the above mutual information can be expressed as

$$\lim_{P \rightarrow \infty} I(\mathbf{x}^{\mathbf{G}}; \mathbf{y} | \mathbf{H}) = N \log(P) + O(1) \quad (2.31)$$

Now we bound the second mutual information term in eq. (2.29) using the model from eq. (2.6)

$$I(\mathbf{H}; \mathbf{y} | \mathbf{x}^{\mathbf{G}}) = I(\mathbf{h}; \mathbf{y} | \mathbf{X}^{\mathbf{G}}) = \mathcal{H}(\mathbf{y} | \mathbf{X}^{\mathbf{G}}) - \mathcal{H}(\mathbf{y} | \mathbf{h}, \mathbf{X}^{\mathbf{G}}),$$

where the entropy of \mathbf{y} given \mathbf{h} and $\mathbf{X}^{\mathbf{G}}$ is equal to the entropy of the i.i.d. Gaussian noise vector \mathbf{z} because of the invariability of the entropy due to deterministic translations [2] and \mathbf{y} given $\mathbf{X}^{\mathbf{G}}$ is zero mean Gaussian distributed with covariance $\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger | \mathbf{X}^{\mathbf{G}}] = P\mathbf{X}^{\mathbf{G}}\mathbf{K}_h\mathbf{X}^{\mathbf{G}\dagger} + \mathbf{I}_N$ where \mathbf{K}_h de-

notes the covariance matrix of the NL length channel vector \mathbf{h} .

$$\begin{aligned}
I(\mathbf{H}; \mathbf{y} | \mathbf{x}^G) &\stackrel{a}{=} \mathbb{E} \log |P\mathbf{X}^G \mathbf{K}_h \mathbf{X}^{G\dagger} + \mathbf{I}_N| \\
&\stackrel{b}{=} \mathbb{E} \log |P\mathbf{K}_h \mathbf{X}^{G\dagger} \mathbf{X}^G + \mathbf{I}_{NL}| \\
&\stackrel{c}{\leq} \log |P\mathbf{K}_h \mathbb{E}(\mathbf{X}^{G\dagger} \mathbf{X}^G) + \mathbf{I}_{NL}| \\
&\stackrel{d}{=} \log |P\mathbf{K}_h + \mathbf{I}_{NL}| \tag{2.32}
\end{aligned}$$

Equality (b) follows from the determinant identity $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$, (c) uses the Jensen's inequality and (d) follows as $\mathbb{E}(\mathbf{X}^{G\dagger} \mathbf{X}^G) = \mathbf{I}_{NL}$. As $\mathbf{h} = \mathbf{F}_c \boldsymbol{\Psi}$ so $\mathbf{K}_h = \mathbf{F}_c \mathbf{K}_\Psi \mathbf{F}_c^\dagger$ where \mathbf{K}_Ψ is the diagonal covariance matrix of LN_d length BEM coefficient vector $\boldsymbol{\Psi}$ due to its uncorrelated entries and the above equation becomes

$$\begin{aligned}
I(\mathbf{H}; \mathbf{y} | \mathbf{x}^G) &\stackrel{a}{\leq} \log |P\mathbf{F}_c \mathbf{K}_\Psi \mathbf{F}_c^\dagger + \mathbf{I}_{NL}| \\
&\stackrel{b}{=} \log |P\mathbf{K}_\Psi \mathbf{F}_c^\dagger \mathbf{F}_c + \mathbf{I}_{LN_d}| \\
&\stackrel{c}{=} \log |P\mathbf{K}_\Psi + \mathbf{I}_{LN_d}| \\
&\stackrel{d}{=} \sum_{i=1}^{LN_d} \log [P\psi_i + 1]. \tag{2.33}
\end{aligned}$$

In (b), we again use the determinant identity $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$ and (c) follows as $\mathbf{F}_c^\dagger \mathbf{F}_c = \mathbf{I}_{LN_d}$. ψ_i denotes the i -th diagonal element of \mathbf{K}_Ψ . At high SNR, this mutual information is given by

$$\lim_{P \rightarrow \infty} I(\mathbf{H}; \mathbf{y} | \mathbf{x}^G) \leq LN_d \log(P) + O(1). \tag{2.34}$$

Combining equations (2.29), (2.31) and (2.34), we get the following lower bound of the mutual information

$$\lim_{P \rightarrow \infty} I(\mathbf{x}^G; \mathbf{y}) \geq (N - LN_d) \log(P) + O(1). \tag{2.35}$$

2.B Upper Bound of MI for Underspread Channels

To derive the upper bound on the mutual information between the input and the output of the channel eq. (2.5) over the block length N , we split the output vector $\mathbf{y} \in \mathbb{C}^N$ in two vectors, one consisting of first LN_d entries and the other having the rest of $N - LN_d$ entries, respectively denoted as

\mathbf{y}_1 and \mathbf{y}_2 . The noise vector $\mathbf{z} \in \mathbb{C}^N$ is divided in \mathbf{z}_1 and \mathbf{z}_2 in the same manner.

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}) &= I(\mathbf{x}; \mathbf{y}_1, \mathbf{y}_2) \\ &= I(\mathbf{x}; \mathbf{y}_1) + I(\mathbf{x}; \mathbf{y}_2 | \mathbf{y}_1) \end{aligned} \quad (2.36)$$

Now we try to upper bound both of the terms in the above equation separately. We treat the second term as

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}_2 | \mathbf{y}_1) &\stackrel{a}{=} \mathcal{H}(\mathbf{y}_2 | \mathbf{y}_1) - \mathcal{H}(\mathbf{y}_2 | \mathbf{y}_1, \mathbf{x}) \\ &\stackrel{b}{\leq} \mathcal{H}(\mathbf{y}_2) - \mathcal{H}(\mathbf{y}_2 | \mathbf{y}_1, \mathbf{x}, \mathbf{H}) \\ &\stackrel{c}{\leq} (N - LN_d) (\mathcal{H}(y[N-1]) - \mathcal{H}(z[N-1])). \end{aligned} \quad (2.37)$$

(a) is the definition of MI in terms of differential entropy, (b) follows because conditioning reduces the entropy and (c) uses the independence bound [2] and because with \mathbf{x} and \mathbf{H} known, the randomness in \mathbf{y} is only due to the noise. $y[N-1]$ is zero mean and its variance is $\mathbb{E}[y[N-1]y[N-1]^\dagger] = PL + 1$ due to i.i.d. Gaussian channel taps of unit variance. Hence at high SNR, this gives

$$\lim_{P \rightarrow \infty} I(\mathbf{x}; \mathbf{y}_2 | \mathbf{y}_1) \leq (N - LN_d) \log(P) + O(1). \quad (2.38)$$

For the first term $I(\mathbf{x}; \mathbf{y}_1)$ in eq. (2.36), the input vector $\mathbf{x} \in \mathbb{C}^N$ is split in two vectors, $\mathbf{x}_1 = [x[0] \cdots x[LN_d - 1] \ x[N-1+L] \cdots x[N-1]]^T$ and $\mathbf{x}_2 = [x[LN_d] \cdots x[N-2+L]]^T$.

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}_1) &\stackrel{a}{=} I(\mathbf{x}_1; \mathbf{y}_1) + I(\mathbf{x}_2; \mathbf{y}_1 | \mathbf{x}_1) \\ &\stackrel{b}{=} I(\mathbf{x}_1; \mathbf{y}_1) \end{aligned} \quad (2.39)$$

(a) follows from the chain rule and (b) follows because given \mathbf{x}_1 , the only randomness in \mathbf{y}_1 is due to the corresponding channel coefficients and the noise, both of which are independent of \mathbf{x}_2 and hence $I(\mathbf{x}_2; \mathbf{y}_1 | \mathbf{x}_1) = 0$. The term $I(\mathbf{x}_1; \mathbf{y}_1)$ represents the mutual information for an overspread channel as the number of observations available are LN_d and same is the number of minimal independent BEM coefficients which need to be estimated. So this term gives no growth with $\log(P)$ and has capacity of the order of $o(\log(P))$. Thus combining equations (2.36), (2.38) and (2.39), the upper bound of the mutual information at high SNR is given by

$$\lim_{P \rightarrow \infty} I(\mathbf{x}; \mathbf{y}) \leq (N - LN_d) \log(P) + o(\log(P)). \quad (2.40)$$

2.C Achievability for Overspread Channels

Here we derive a lower bound on the achievable data rate of overspread channels with the zero-padded transmission scheme described in section 2.5.1. In this scheme, we transmit one symbol and do zero padding of $\lfloor L/2 \rfloor$ symbols and so on. Thus the input vector \mathbf{x} can be split in two vectors, one vector \mathbf{x}_a containing all the non-zero input samples and the other \mathbf{x}_b containing all the zero-padded input samples. So \mathbf{x}_a has samples of \mathbf{x} from indices $\{i(\lfloor L/2 \rfloor + 1), i = 0, 1, \dots, N/(\lfloor L/2 \rfloor + 1)\}$. Similarly we split the output samples in two groups, ones which appear with no ISI and the other samples where we get sum of multiple input samples weighted with channel coefficients. We denote \mathbf{y}_a as the vector of output samples which appear without ISI and hence they contain sample values of \mathbf{y} at time indices $\{i(\lfloor L/2 \rfloor + 1) + \lfloor L/2 \rfloor, i = 0, 1, \dots, N/(\lfloor L/2 \rfloor + 1)\}$ and \mathbf{y}_b denotes the vector of the rest of the output samples. So the achievable data rate, denoted by R_N , is

$$\begin{aligned}
 R_N &\stackrel{a}{=} I(\mathbf{x}; \mathbf{y}) = I(\mathbf{x}_b; \mathbf{y}) + I(\mathbf{x}_a; \mathbf{y} | \mathbf{x}_b) \\
 &\stackrel{b}{=} I(\mathbf{x}_a; \mathbf{y} | \mathbf{x}_b) \\
 &\stackrel{c}{=} I(\mathbf{x}_a; \mathbf{y}_a | \mathbf{x}_b) + I(\mathbf{x}_a; \mathbf{y}_b | \mathbf{x}_b, \mathbf{y}_a) \\
 &\stackrel{d}{\geq} I(\mathbf{x}_a; \mathbf{y}_a | \mathbf{x}_b)
 \end{aligned} \tag{2.41}$$

(b) follows as \mathbf{x}_b is deterministically zero, giving $I(\mathbf{x}_b; \mathbf{y}) = 0$ and (d) follows from the non-negativity of the mutual information.

Each element in \mathbf{x}_a has a one-to-one relationship with the corresponding element of \mathbf{y}_a given by:

$$y_a[i + \lfloor L/2 \rfloor] = \sqrt{P}x_a[i]h[i + \lfloor L/2 \rfloor, \lfloor L/2 \rfloor] + z[i + \lfloor L/2 \rfloor] \tag{2.42}$$

where $i = 0, 1, \dots, N/(\lfloor L/2 \rfloor + 1)$. This equation represents the input-output relationship for a frequency flat time varying channel for which high SNR capacity results are known in the non-coherent case [25]. For the mutual information term $I(\mathbf{x}_a; \mathbf{y}_a | \mathbf{x}_b)$, both the input and the output have length $N/(\lfloor L/2 \rfloor + 1)$ which plays the role of the block length in this case. Now there is only a single channel tap which needs to be estimated for the coherent data detection and requires the estimation of N_d BEM coefficients for this block, the rank of the channel covariance matrix for this particular tap. Hence in a straightforward manner, using the result from [25], we can write

$$\lim_{P \rightarrow \infty} R_N \geq \left(\frac{N}{\lfloor \frac{L}{2} \rfloor + 1} - N_d \right) \log(P) + O(1). \tag{2.43}$$

This is the rate over the block length of N symbol intervals, so the pre-log per channel use is given by

$$\text{PreLog} \geq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \left(1 - (\lfloor \frac{L}{2} \rfloor + 1) \frac{N_d}{N} \right). \quad (2.44)$$

While deriving large block length asymptotics for underspread channels, we showed that for very large values of N , the factor N_d/N is equal to B_d , the normalized Doppler bandwidth. Hence the pre-log for the novel zero-padded transmission scheme is lower bounded as

$$\text{PreLog} \geq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \left(1 - (\lfloor \frac{L}{2} \rfloor + 1) B_d \right). \quad (2.45)$$

Chapter 3

MU MIMO: DOF with no CSI

3.1 Introduction

3.1.1 Motivation

This chapter focuses on high SNR capacity analysis like Chapter 2 where pre-log or DOF becomes the capacity determining parameter but both differ in certain aspects. Contrary to the previous chapter which treats single-user point-to-point channels, this chapter treats multi-user downlink channels which require the availability of CSI at the transmitter to be fully operational. The other major difference is in the time variation pattern of the channel model. Chapter 2 treats continuously time-varying per-symbol stationary channel whereas in this chapter block fading cyclo-stationary channels are considered. A fully cognizant DL channel, comprising of an M -antenna BS communicating with K single-antenna users ($K \geq M$) in the DL direction, can enjoy data rates M times larger than a single antenna BS [6], [30], [17]. The fact that practically all channels are non-coherent in nature poses a serious question to this cognizant scaling law of the DL channel capacity. Hence a natural question to ask is what is the multiplexing gain of a DL channel if impractical assumptions of the presence of CSI are removed and the users and the BS have to acquire channel knowledge at the

expense of available communication resources.

We make no assumption of the presence of CSI but the BS and the users are not prevented from learning, feeding back the channel and subsequently use this information for precoding/decoding of data. If CSI is not assumed to be there, the users (receivers) need to estimate the channels implicitly (data aided) or explicitly by some kind of training (pilots transmission) to get CSIR. The CSIT acquisition varies depending upon the duplexing mode of the system. For FDD systems, the CSIT is acquired when the users feedback their estimated forward channel information on the reverse link. On the other hand for TDD systems, the reciprocity implies that the forward channel matrix is the transpose of the reverse channel matrix [18], hence facilitating the CSIT acquisition by simple pilot transmission from the user terminals to the BS.

3.1.2 The State of the Art

Most of the references related to this work have already been described in Chapter 1 so we keep the discussion to minimum. With perfect assumption of CSI, the broadcast channel with M -antenna BS and K ($K \geq M$) single-antenna users shows multiplexing gain of M [6], [30], [17] and a multi-user diversity [12] benefit of $M \log \log(K)$. The same broadcast channel with perfect CSIR and no CSIT has no multi-user diversity gain and a multiplexing gain of only one as single user transmission becomes optimal in such a scenario [2], [16], [3]. For $K \geq M$, the CSIT of M users is indispensable to achieve the full multiplexing gain [17], [6] and capturing the multi-user diversity benefit of $M \log \log(K)$ in the sum rate requires the CSIT availability of all of these K users where normally K could be much larger than M .

3.1.3 Contribution

In this chapter, contrary to the previous chapter where we analyzed underspread and overspread channels, the channels of concern are strictly underspread fading channels, although high SNR regime rests the focus of attention. When no feedback is allowed to the BS throughout the transmission, it is shown that the DL capacity region is bounded by the capacity of a point-to-point MISO link and hence the multiplexing gain of the DL sum rate is $(1 - 1/T)$ for a block fading channel of coherence length T . Our next objective is to analyze the DOF of the broadcast channel when the system is allowed to acquire CSIT. For a broadcast channel operating under TDD

regime, we give a transmission scheme where the BS is made to learn users' channels and with which multiplexing gain (DOF) of $M[1 - (M + 1)/T]$ is achievable. The breakdown of this strategy reveals that the CSIT quality should be imperatively refined with DL SNR otherwise the multiplexing gain of the multi-antenna DL channel is lost. An upper bound to the DOF of this DL channel is also obtained by letting all the users collaborate among themselves. The brief analysis of the broadcast channel operating under FDD mode is also carried out by providing a practical transmission scheme, serving to furnish CSI to the BS and all the users followed by data transmission, and the achievable DOF corresponding to this scheme are specified. This scheme enables the efficient use of UL FDD system resource without any further exchange of information, making the system practically scalable for bi-directional data transmission.

3.1.4 Organization

The system model is described in section 3.2. In section 3.3, we analyze the capacity of a broadcast channel when no feedback is allowed to the BS. Next two sections deal with a TDD broadcast channel when feedback is allowed. A complete transmission strategy to provide CSI is given in section 3.4. Section 3.5 gives lower and upper bounds for the sum rate and the multiplexing gain, with a short note on the significance of the CSIT quality refinement. After this, the chapter focuses on FDD broadcast channels giving a suitable DL transmission strategy and specifying high SNR DOF in section 3.6. The conclusions of this chapter are summarized in section 3.7.

3.2 System Model

The system we consider consists of a BS having M transmit antennas and K single-antenna user terminals. In the downlink, the signal received by k -th user y_k can be expressed as

$$y_k = \mathbf{h}_k^\dagger \mathbf{x} + z_k, \quad k = 1, 2, \dots, K \quad (3.1)$$

where $\mathbf{h}_1^\dagger, \mathbf{h}_2^\dagger, \dots, \mathbf{h}_K^\dagger$ are the DL channel vectors of users 1 through user K with $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$, $\mathbf{x} \in \mathbb{C}^{M \times 1}$ denotes the channel input and z_1, z_2, \dots, z_K are independent complex Gaussian additive noise terms with zero mean and unit variance. We denote the concatenation of the channels by $\mathbf{H}^\dagger = [\mathbf{h}_1 \mathbf{h}_2 \cdots \mathbf{h}_K]$, so \mathbf{H} is the $K \times M$ forward channel matrix with k -th row equal to the channel of k -th user (\mathbf{h}_k^\dagger). The channel input from the BS must

satisfy an average transmit power constraint of P i.e., $\mathbb{E}[\|\mathbf{x}\|^2] \leq P$ which is equal to the average SNR at each user as the noise at each user has been normalized to have unit variance.

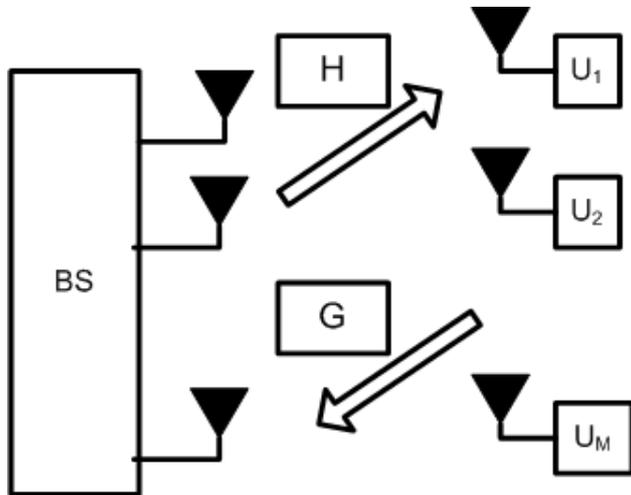


Figure 3.1: Multi-User MIMO System Model

The channel is assumed to be block fading having coherence length of T symbol intervals during which fading remains the same, with independent fading from one block to the next [23]. The entries of the forward channel matrix \mathbf{H} are independent and identically distributed (i.i.d.) complex Gaussian with zero mean and unit variance.

The uplink channel matrix for all users is $\mathbf{G} = [\mathbf{g}_1 \mathbf{g}_2 \cdots \mathbf{g}_K]$, where $\mathbf{g}_k \in \mathbb{C}^{M \times 1}$ is the channel for k -th user, comprising of i.i.d. zero mean unit variance Gaussian distributed entries. When the system is operating under TDD mode, $\mathbf{G} = \mathbf{H}^T$ due to perfect reciprocity assumption. When the system operates under FDD mode, we assume that the bandwidths allocated for UL and DL undergo independent fading realization and the UL channel is also T length block fading with synchronized switching point.

In broadcast channels, the number of users (K) could be larger than the number of BS transmit antennas (M). With CSI available, $K = M$ single-antenna users are sufficient to achieve full DOF M of the DL channel [43] and surplus users can provide only multi-user diversity gain [12] adding nothing to the pre-log. In this chapter, our main point of concern is the multiplexing gain or the DOF of non-coherent broadcast channel so we restrict the number of users K equal to M . Practically this mimics the sit-

uation when a scheduling algorithm chooses M out of K users independent of their channel realizations.

3.3 Broadcast channel with No CSIT

For the discussion in this section, the BS is kept oblivious of the CSIT throughout the transmission. This portrays a practical scenario where the user terminals are inexpensive devices with only the reception capabilities.

For a no CSIT broadcast channel where all the users are symmetrically distributed and each user (receiver) knows its own channel perfectly, the sum capacity of this channel is equal to the capacity of the point-to-point channel from the transmitter to any one of the receivers. Thus TDMA is the optimal strategy in this case of no CSIT [2], [16], [3] and the multiplexing gain of such a broadcast channel with CSIR and no CSIT becomes only one.

In this section, we focus on the broadcast channel where even the users have no channel information (no CSIR case). Because of the symmetry of the fading distributions among users, these channels fall under the category of “bottleneck channels” of Cover [16]. So any code transmitted by the BS which is decodable at any user i is also decodable at any other user j implying that every user can decode all the information transmitted by the BS for all the users. Hence the capacity region of such a broadcast channel is bounded by the capacity of a single user channel from the BS to any one of the users. And the maximum sum rate with the restriction of no feedback is given by

$$R_{\text{sum}}^{\text{NO-FB}} = C_{\text{SU}}, \quad (3.2)$$

where C_{SU} is the single-user capacity of a non-coherent MISO link from M -antenna BS to any of the single antenna users. Although, for the case of interest (no CSIT, no CSIR), exact expression even for C_{SU} is not known but high SNR asymptotics are available. Using the non-coherent capacity result of block fading channel from [24], we can write

$$R_{\text{sum}}^{\text{NO-FB}} = \left(1 - \frac{1}{T}\right) \log(P) + o(1), \quad (3.3)$$

where $o(1)$ is a term that does not depend upon SNR at large SNR. Hence the exact multiplexing gain of a broadcast channel, with no initial CSI and no CSIT throughout, is given by

$$\text{PreLog} = 1 - \frac{1}{T}. \quad (3.4)$$

The achievability of this DOF is straightforward. The BS activates any one of its M transmit antennas and we also focus on a single user. This reduces the broadcast channel to a point-to-point SISO channel. In each coherence block of length T , first symbol is dedicated to training when the selected user estimates the only channel coefficient present. On the rest of $T - 1$ symbol intervals, user decodes the data based upon this channel knowledge, so extracting $T - 1$ DOF in each coherence block of T symbol intervals, matching the rate of equation (3.3).

The analysis carried out in this section tells us that for a broadcast channel with M transmit antennas and K single antenna users, having no initial CSI and no CSIT throughout the transmission, the capacity region is bounded by the capacity of an $M \times 1$ MISO link. Thus, in this particular scenario, the existence of multiple users gives no gain at any SNR.

3.4 TDD Broadcast Channel - Transmission Strategy

For an M -antenna BS communicating with $K = M$ single antenna users with perfect CSI, the first order term of the sum capacity is $M \log(P)$ [43]. This term reduces to $\log(P)$ if only CSIR is available, giving the strong motivation of having a learned transmitter BS. So in this section, a transmission scheme is proposed which provides necessary CSI for a TDD DL system. Later based upon the proposed strategy, rate bounds are derived and analyzed in the next section.

In this scheme, the block fading channel coherence length of T channel uses is divided in three phases: 1) uplink training, 2) downlink training and 3) coherent data transmission. The first phase is the UL training phase

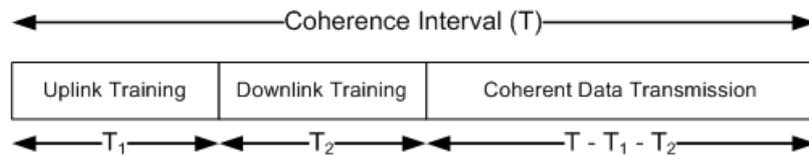


Figure 3.2: Transmission Phases for TDD System

which serves to provide CSIT. Based upon this channel information, BS may choose some transmission strategy which could be a simple linear beamforming strategy like zero forcing (ZF), some non-linear strategy like vector per-

turbation or the optimal DPC. The second phase is the DL training phase where the BS transmits pilots so that the users estimate their corresponding effective channels. When this second phase ends, both sides of the broadcast channel have necessary CSI, albeit imperfect. This makes possible the coherent data transmission and reception in the third data phase. Below we give a detailed analysis of the transmission phases and the BS processing steps for this scheme.

3.4.1 Uplink Training Phase

In this training phase, users transmit pilot sequences which are known to the BS. Due to TDD system with perfect reciprocity, uplink channel estimates at the BS provide CSIT and this training behaves equivalent to the analog feedback from the users. Hence working with TDD systems, we use the terms uplink training and feedback synonymously. As there are $K = M$ users, so the length of this uplink training interval is $T_1 \geq M$. Here we suppose that the power constraint of each user is P_u . For this uplink training, the use of orthogonal training sequences by all users is attractive in the sense that all users can transmit simultaneously to the BS with their full power without interfering with each other. If pilot signal matrix (combined from all users) is $\sqrt{T_1}\mathbf{A}$ where \mathbf{A} is an $M \times T_1$ unitary matrix then $\mathbf{A}\mathbf{A}^\dagger = \mathbf{I}_M$. If \mathbf{Y}_u denotes the $M \times T_1$ matrix of the received signal by M antennas of the BS in this training interval of length T_1 , the system equation for this uplink training phase becomes

$$\mathbf{Y}_u = \sqrt{P_u T_1} \mathbf{G} \mathbf{A} + \mathbf{Z}_u, \quad (3.5)$$

where \mathbf{Z}_u is an $M \times T_1$ matrix having i.i.d. zero mean unit variance complex Gaussian noise entries at the BS antennas during this training phase and \mathbf{G} denotes the $M \times M$ uplink channel matrix. As pilot signal matrix \mathbf{A} is known at the BS, it can formulate an MMSE estimate of the uplink channel matrix \mathbf{G} which is given by

$$\hat{\mathbf{G}} = \frac{\sqrt{P_u T_1}}{P_u T_1 + 1} \mathbf{Y}_u \mathbf{A}^\dagger. \quad (3.6)$$

The TDD reciprocity of this broadcast channel dictates that the DL channel matrix is just the transpose of the UL channel matrix, giving $\hat{\mathbf{H}} = \hat{\mathbf{G}}^\mathbf{T}$. The channel vector for user k can be expressed as $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$ where $\tilde{\mathbf{h}}_k$ is the estimation error vector with i.i.d. Gaussian entries. All entries in the channel matrix are independent hence CSIT estimation error variance for

any channel entry, denoted by σ_h^2 , is given by

$$\sigma_h^2 = \mathbb{E}[|\mathbf{H}_{ij} - \hat{\mathbf{H}}_{ij}|^2] = \frac{1}{P_u T_1 + 1}. \quad (3.7)$$

The CSIT estimation error variance for each channel coefficient goes inversely proportional to the training length T_1 and the power constraint of the user terminals P_u . Furthermore, the length of this UL training phase T_1 depends solely upon the number of users and is completely independent of the number of BS antennas.

3.4.2 BS Transmission Strategy: ZF Precoding

It is known that the DPC is the capacity achieving transmission scheme for MIMO broadcast channels and achieves the full capacity region [31] but its implementation is quite tedious. So a lot of research has been carried out to analyze the performance of simpler linear precoding schemes. ZF precoding, one of the simplest linear precoding strategy, has been shown to behave quite optimally at asymptotically high values of SNR and achieves the full DOF of a coherent broadcast channel [43]. In this chapter, we are mainly interested in analyzing the achievable DOF hence BS uses ZF precoding based upon the knowledge of the forward channel matrix obtained through explicit UL training.

In ZF precoding, beamforming vector for user k (denoted as $\bar{\mathbf{v}}_k$), is selected such that it is orthogonal to the channel vectors of all other users. ZF beamforming vectors are the normalized columns of the inverse of the channel matrix \mathbf{H} . Hence for ZF with perfect CSIT, each user receives only the beam directed to it and no multi-user interference is experienced. For the case in hand, where the BS has imperfect estimate of the channel matrix, there will be some residual interference. If we represent ZF beamforming matrix by $\bar{\mathbf{V}} = [\bar{\mathbf{v}}_1 \bar{\mathbf{v}}_2 \cdots \bar{\mathbf{v}}_M]$, the transmitted signal \mathbf{x} becomes $\mathbf{x} = \bar{\mathbf{V}}\mathbf{u}$ where \mathbf{u} is the data vector with u_k data intended for k -th user. Thus the signal received by k -th user (3.1) can be expressed as

$$\begin{aligned} y_k &= \mathbf{h}_k^\dagger \bar{\mathbf{V}}\mathbf{u} + z_k \\ &= \mathbf{h}_k^\dagger \bar{\mathbf{v}}_k u_k + \sum_{j \neq k} \mathbf{h}_k^\dagger \bar{\mathbf{v}}_j u_j + z_k. \end{aligned} \quad (3.8)$$

The choice of ZF beamforming unit vectors based upon imperfect CSIT makes the k -th user receive some signal intended for j -th user ($j \neq k$) through its effective channel

$$\mathbf{h}_k^\dagger \bar{\mathbf{v}}_j = \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j + \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j = \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j, \quad (3.9)$$

hence the received signal at k -th user becomes

$$\begin{aligned} y_k &= \mathbf{h}_k^\dagger \bar{\mathbf{v}}_k u_k + \sum_{j \neq k} \tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j u_j + z_k \\ &= h_{k,k} u_k + \sum_{j \neq k} h_{k,j} u_j + z_k. \end{aligned} \quad (3.10)$$

$h_{k,k} = \mathbf{h}_k^\dagger \bar{\mathbf{v}}_k$ is the effective scalar channel for user k and $h_{k,j} = \mathbf{h}_k^\dagger \bar{\mathbf{v}}_j$ are the interference coefficients which arise due to ZF beamforming based upon imperfect CSIT.

3.4.3 Downlink Training Phase

ZF based upon perfect CSIT creates non-interfering SISO links from the BS to the users. Here the CSIT and the ZF beamforming vectors are imperfect so each user receives some unwanted signal contribution from the beams directed to other users. This interference is of the same order as of the channel noise so, for this DL training phase, we let the BS activate all the beams simultaneously for T_2 symbol intervals. So in each symbol interval, every user receives the pilot through its effective scalar channel, the Gaussian noise of the channel and the interference due to imperfect channel estimates and ZF beamforming vectors.

$$y_k = h_{k,k} u_k + \sum_{j \neq k} h_{k,j} u_j + z_k \quad (3.11)$$

Based upon this received signal and the known pilots, k -th user forms the MMSE estimate $\hat{h}_{k,k}$ of the effective scalar channel $h_{k,k}$ which is given by

$$\begin{aligned} \hat{h}_{k,k} &= \frac{\mathbb{E}[h_{k,k} y_k^\dagger]}{\mathbb{E}[y_k y_k^\dagger]} y_k \\ &= \frac{\sqrt{\frac{PT_2}{M}}}{\frac{PT_2}{M} + \frac{PT_2}{M}(M-1)\sigma_h^2 + 1} y_k. \end{aligned} \quad (3.12)$$

(See Appendix 3.A for the details of the derivation of this estimator.)

The unit vector $\bar{\mathbf{v}}_k$ is independent of \mathbf{h}_k due to its construction, so effective scalar channel $h_{k,k} = \mathbf{h}_k^\dagger \bar{\mathbf{v}}_k$ is zero mean complex Gaussian with unit variance. As a result, MMSE estimate $\hat{h}_{k,k}$ and the estimation error $\tilde{h}_{k,k}$, giving

$h_{k,k} = \hat{h}_{k,k} + \tilde{h}_{k,k}$, are complex Gaussian distributed as below.

$$\begin{aligned}\hat{h}_{k,k} &\sim \mathcal{CN}\left(0, \frac{\frac{PT_2}{M}}{\frac{PT_2}{M} + \frac{PT_2}{M}(M-1)\sigma_h^2 + 1}\right) \\ \tilde{h}_{k,k} &\sim \mathcal{CN}\left(0, \frac{\frac{PT_2}{M}(M-1)\sigma_h^2 + 1}{\frac{PT_2}{M} + \frac{PT_2}{M}(M-1)\sigma_h^2 + 1}\right)\end{aligned}$$

This simultaneous activation of all the ZF beams from the BS to provide CSIR has the advantage that the length of this DL training phase T_2 becomes independent of the number of BS transmit antennas and the number of users (so can be reduced to one in the extreme case). On the negative side, it has the disadvantage of producing larger estimation error for the estimation of effective channel due to the presence of interference caused by beams meant for other users. So if CSIT is not of good quality (which would imply relatively large imperfections in the ZF beamforming vectors and hence significant power in the interfering beams), it might be a good idea to activate the beams one by one at the BS for this training process removing the interference.

3.4.4 Coherent Data Phase

After the two training phases, first in the uplink and second in the downlink direction, both the BS and all the users have imperfect channel estimates hence coherent data transmission is possible. We adopt the strategy of independent data transmission to all users from the BS with power equally divided among them. So k -th user information signal u_k is Gaussian i.i.d., i.e. $u_k \sim \mathcal{CN}(0, P/M)$. The intuition is that, in case of perfect CSI, Gaussian signals are the optimal ones. With ZF beamforming employed, the signal y_k , received by k -th user (3.10), may be expressed as

$$y_k = \hat{h}_{k,k}u_k + \tilde{h}_{k,k}u_k + \sum_{j \neq k} h_{k,j}u_j + z_k. \quad (3.13)$$

Contrary to eq. (3.10) where user k is unaware of its scalar channel $h_{k,k}$, eq. (3.13) effectively represents a point-to-point coherent channel with channel $\hat{h}_{k,k}$ known at k -th user, although there is Gaussian noise, some interference coming from ZF beamforming vectors of other users and the noise due to imperfect estimation of the effective channel at user's side.

3.5 TDD Broadcast Channel - DOF Bounds

3.5.1 Achievable Rate of the Scheme

If we denote the rate achievable by k -th user as R_k , it is the mutual information I between u_k and y_k with known channel $\hat{h}_{k,k}$

$$R_k = I(u_k; y_k). \quad (3.14)$$

In this case, the expression for the mutual information of known scalar channel cannot be used because of the presence of interference terms whose distributions are unknown. If we combine the noise, the interference and the estimation error contribution present in the observed signal eq. (3.13) in an effective additive noise w_k , then

$$w_k = \tilde{h}_{k,k}u_k + \sum_{j \neq k} h_{k,j}u_j + z_k. \quad (3.15)$$

Now the variance of this effective noise conditioned upon the effective scalar channel estimate $\hat{h}_{k,k}$ is required to evaluate the above mutual information expression which is given as

$$\mathbb{E}[w_k w_k^\dagger | \hat{h}_{k,k}] = \mathbb{E}[\tilde{h}_{k,k}^2] \mathbb{E}[|u_k|^2] + \sum_{j \neq k} \mathbb{E}[|h_{k,j}|^2 | \hat{h}_{k,k}] \mathbb{E}[|u_j|^2] + \mathbb{E}[|z_k|^2].$$

All the expectations in the above equation are known except $\mathbb{E}[|h_{k,j}|^2 | \hat{h}_{k,k}]$ which is difficult to compute.

$$\mathbb{E}[w_k w_k^\dagger | \hat{h}_{k,k}] = \frac{P}{M} \frac{\frac{PT_2}{M}(M-1)\sigma_h^2 + 1}{\frac{PT_2}{M} + \frac{PT_2}{M}(M-1)\sigma_h^2 + 1} + \frac{P}{M} \sum_{j \neq k} \mathbb{E}[|h_{k,j}|^2 | \hat{h}_{k,k}] + 1 \quad (3.16)$$

Due to the use of MMSE estimation in the DL training, the signal becomes uncorrelated with the noise and all interfering terms giving

$$\mathbb{E}[u_k(\tilde{h}_{k,k}u_k + \sum_{j \neq k} h_{k,j}u_j + z_k)^\dagger] = 0. \quad (3.17)$$

The above expectation is zero because of the property of uncorrelated MMSE estimation error, the use of independent signals for different users and that the noise is independent of everything else. With all additive noise terms uncorrelated with the desired signal, we can invoke Theorem 1 from [44] which states that the worst case uncorrelated noise has zero mean Gaussian

distribution. So we can replace the effective scalar additive noise w_k of unknown distribution with a noise of the same second moment but having Gaussian distribution. It will give a lower bound to the rate R_k of k -th user but the mutual information can be written as

$$\begin{aligned} R_k &\geq \mathbb{E}_{\hat{h}_{k,k}} \log \left(1 + \frac{|\hat{h}_{k,k}|^2 \mathbb{E}[u_k]^2}{\mathbb{E}[w_k w_k^\dagger | \hat{h}_{k,k}]} \right) \\ &= \mathbb{E}_{\hat{h}_{k,k}} \log \left(1 + \frac{P}{M} \frac{|\hat{h}_{k,k}|^2}{\mathbb{E}[w_k w_k^\dagger | \hat{h}_{k,k}]} \right) \end{aligned} \quad (3.18)$$

3.5.2 High SNR DOF of the Achievable Sum Rate

The rate for k -th user derived in eq. (3.18) can further be lower bounded as

$$\begin{aligned} R_k &\geq \mathbb{E}_{\hat{h}_{k,k}} \log \left(\frac{P}{M} \frac{|\hat{h}_{k,k}|^2}{\mathbb{E}[w_k w_k^\dagger | \hat{h}_{k,k}]} \right) \\ &= \mathbb{E}_{\hat{h}_{k,k}} \log \left(\frac{P}{M} |\hat{h}_{k,k}|^2 \right) - \mathbb{E}_{\hat{h}_{k,k}} \log \left(\mathbb{E}[w_k w_k^\dagger | \hat{h}_{k,k}] \right) \\ &\geq \mathbb{E}_{\hat{h}_{k,k}} \log \left(\frac{P}{M} |\hat{h}_{k,k}|^2 \right) - \log \left(\mathbb{E}[w_k w_k^\dagger] \right), \end{aligned} \quad (3.19)$$

where the first inequality becomes a good approximation at large SNR and the last inequality follows from the Jensen's inequality. With this, we only need to compute the 2nd moment of w_k which is readily shown to be

$$\sigma_w^2 = \mathbb{E}[w_k w_k^\dagger] = \frac{P}{M} \frac{\frac{PT_2}{M}(M-1)\sigma_h^2 + 1}{\frac{PT_2}{M} + \frac{PT_2}{M}(M-1)\sigma_h^2 + 1} + (M-1) \frac{P}{M} \sigma_h^2 + 1, \quad (3.20)$$

because

$$\mathbb{E}[|h_{k,j}|^2] = \mathbb{E}[|\tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_j|^2] = \sigma_h^2 = \frac{1}{P_u T_1 + 1}. \quad (3.21)$$

As all of the users are symmetrically distributed, so the sum rate of this broadcast channel is given by

$$\begin{aligned} R_{\text{sum}}^{\text{TDD}} &= \frac{T - T_1 - T_2}{T} M R_k \\ &\geq \frac{T - T_1 - T_2}{T} M \left[\mathbb{E}_{\hat{h}_{k,k}} \log \left(\frac{P}{M} |\hat{h}_{k,k}|^2 \right) - \log(\sigma_w^2) \right], \end{aligned} \quad (3.22)$$

where we have also incorporated the DOF loss in the sum rate due to two training phases in the UL and the UL directions.

If we increase the first training phase duration T_1 , it improves the quality of CSIT at the BS and interference at each user due to beamforming vectors of other users decreases but it gives only a gain in SNR offset (see (3.22) and (3.20)) which is logarithmic in nature but the coefficient $(T - T_1 - T_2)$ reduces the DOF of the sum rate linearly with increase in T_1 so the optimal length of the first training phase should be the minimum possible at high SNR, hence $T_1 = M$. This argumentation assumes that the power constraints of user terminals (P_u) are of the same order as that of the BS power constraint P .

Similarly with the increase in the DL training interval T_2 , the users are better able to estimate their effective scalar channels which gives SNR gain, logarithmic in nature but increase in T_2 directly hits DOF due to the coefficient $(T - T_1 - T_2)$ in front of the logarithm. So to exploit the maximum number of DOF at high SNR, the optimal value of T_2 comes out to be 1. Adopting these values, the sum rate becomes

$$R_{\text{sum}}^{\text{TDD}} \geq \frac{T - M - 1}{T} M \left[\mathbb{E}_{\hat{h}_{k,k}} \log \left(\frac{P}{M} |\hat{h}_{k,k}|^2 \right) - \log(\sigma_w^2) \right]. \quad (3.23)$$

It's trivial to show that σ_w^2 is bounded by a finite constant for large values of P (the BS power constraint) and if power constraints of users (P_u) are of the same order as that of P . Furthermore, $\mathbb{E}_{\hat{h}_{k,k}} \log(|\hat{h}_{k,k}|^2)$, which is also a constant, can be exactly specified using Lemma 8 from [25]. So for limiting values of P , the lower bound to the sum rate becomes

$$\lim_{P \rightarrow \infty} R_{\text{sum}}^{\text{TDD}} \geq \frac{T - (M + 1)}{T} M \log(P) + o(1). \quad (3.24)$$

Thus the multiplexing gain of a non-coherent TDD broadcast channel is lower bounded by

$$\text{PreLog} \geq M \left[1 - \frac{M + 1}{T} \right]. \quad (3.25)$$

So for a broadcast channel operating under TDD mode, having M BS antennas, same number of symmetric users, block fading channel of coherence interval T and starting from zero channel state information at both ends, our very simple scheme is able to achieve $M[1 - (M + 1)/T]$ DOF. If we compare this multiplexing gain to the multiplexing gain of the same broadcast channel under the restriction of no feedback to the BS (section 3.3) where DOF is only $(1 - 1/T)$, it is clear that even for very practical values of the block coherence interval T in mobile environments, this lower bound $M[1 - (M + 1)/T]$ is comparatively much larger and to make the BS learn the channel pays off very well.

3.5.3 Upper Bound of the Multiplexing Gain

An upper bound to the sum rate of our scheme can be obtained when one sacrifices minimal lengths for both training intervals but then assumes that the BS knows the DL channels perfectly and each user perfectly knows its effective scalar channel. This will remove all the interference terms from the received signal but DOF achieved will still be $M[1 - (M + 1)/T]$.

A general upper bound to the sum rate of TDD broadcast channel can be obtained by letting all the user terminals cooperate among themselves. If M users are collaborating, we get an equivalent single user point-to-point MIMO square channel of M dimensions. For this point-to-point non-coherent MIMO channel, results are available in the literature [24]. Thus the sum rate of TDD broadcast channel at high SNR is upper bounded by

$$R_{\text{sum}}^{\text{TDD}} \leq \frac{T - M}{T} M \log(P) + o(1). \quad (3.26)$$

And upper bound to the multiplexing gain of the sum rate is given by

$$\text{PreLog} \leq M \left[1 - \frac{M}{T} \right]. \quad (3.27)$$

This shows that our scheme, achieving $M[1 - (M + 1)/T]$ DOF, is very close to this high SNR asymptote of single user MIMO. Although we don't have a matching upper bound yet we conjecture that the DOF achievable by our scheme $M[1 - (M + 1)/T]$ is the true DOF of TDD broadcast channel due to CSI requirement at both ends.

3.5.4 CSIT Quality Refinement

While switching from eq. (3.23) to eq. (3.24) which shows that our scheme is able to achieve $M[1 - (M + 1)/T]$ DOF for this broadcast channel, the effective noise variance σ_w^2 should stay bounded even with DL SNR going to infinity. The expression of this effective noise variance eq. (3.20) reveals that this requires users' power constraint P_u to be of the same order as that of the BS power constraint P . If this is not the case

$$\lim_{P \rightarrow \infty} \frac{P_u}{P} = 0, \quad (3.28)$$

and the CSIT quality at the BS will be very poor (as compared to the DL SNR) and the interference power at each user (the product of the signal power and the CSIT error variance) due to beams meant for other users

(and hence σ_w^2) will go on increasing with the DL power P . This will cause the sum rate to saturate in SNR resulting in complete collapse of DOF. This result parallels the result of [17] for quantized feedback which showed that the feedback rate (the CSIT quality) must increase with the DL SNR (in dBs) to achieve full DOF of the broadcast channel. Our result with analog CSIT acquisition points out that the UL power (which governs the CSIT quality) must scale up with the DL SNR. The rates unbounded in SNR can be achieved by transmitting to a single user (or by time-sharing between users), with fixed uplink power or even with no feedback to the BS, but DOF of the broadcast channel (due to multiple antennas at the BS and multiple users at the receiving side) are lost.

3.6 FDD Broadcast Channel

The CSIT acquisition is relatively difficult for FDD systems as compared to TDD systems. The absence of channel reciprocity requires explicit channel feedback transmission from the users to the BS to furnish CSIT. To cater for such peculiarities, we propose a novel transmission strategy for FDD systems and analyze its achievable rate and multiplexing gain. Due to certain similarities with TDD strategy, the discussion is kept to minimum.

3.6.1 Transmission Strategy

The proposed transmission strategy for FDD systems divides the coherence length of T symbol intervals in four phases: 1) initial UL and DL training, 2) uplink feedback, 3) final DL training and 4) coherent data transmission. The

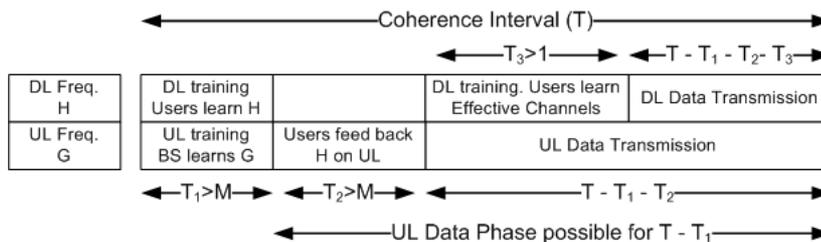


Figure 3.3: Transmission Phases for FDD System

first phase is the initial UL and DL training phase where the users transmit pilots on the UL band so that the BS estimates the UL channels. The BS is

transmitting pilots simultaneously on the DL so that each user estimates its corresponding DL channel. The presence of M BS antennas and M system users requires that the length T_1 of this initial training interval must exceed M , the number of independent parameters which need to be estimated at each user and at each BS antenna.

This initial training interval serves to provide CSIR to both communicating ends but to multiplex data streams of multiple users on the DL, BS requires the CSIT, the information of the DL channel matrix. In the second UL feedback phase, each user transmits its estimated DL channel on the UL. The availability of the UL (imperfect) CSIR at the BS allows the users to use this UL channel as fully operational MIMO MAC with spatial multiplexing enabled. Analog feedback (also termed as unquantized and uncoded) is an attractive option due to its simplicity and optimality in case source and channel bandwidth are the same [45], [46] which becomes the case of interest at high SNR. As each user has only a single antenna so the transmission of M complex coefficients will require this interval length (denoted by T_2) to be at least M hence $T_2 \geq M$. Based upon the CSIT obtained in this feedback interval, the BS chooses some transmission strategy (we adopt ZF as for the TDD scheme) to multiplex users. The third phase is the DL training phase where the BS transmits pilots through the chosen beamforming vectors so that the users estimate their corresponding effective scalar channels, exactly similar to the TDD scheme. As each user needs to estimate its effective scalar channel, the length of this training interval T_3 should be at least 1. When this third phase ends, both sides of the DL channel have necessary CSI. Hence in the fourth data phase, the BS transmits independent data streams to M users simultaneously and each user decodes its stream coherently.

3.6.2 High SNR DOF of the Sum Rate

Treating all the preliminary training and feedback phases in the similar manner as for TDD case provides imperfect CSI to the BS and all the users. In the final data phase, each user receives a mixture of the desired signal, the channel noise and the interference terms which arise due to ZF beamforming based upon imperfect CSIT and imperfect CSIR. If R_k denotes the rate of k -th user, the sum rate of this scheme is given by

$$R_{\text{sum}}^{\text{FDD}} = \frac{T - T_1 - T_2 - T_3}{T} M R_k, \quad (3.29)$$

where we have incorporated the DOF loss in the sum rate due to one feedback and two training phases. If we increase the first training phase duration

T_1 , it improves the CSIR quality at both ends. The increase of the feedback length T_2 improves the CSIT quality at the BS. In the same manner, increase in the final DL training length T_3 lets the users better estimate their effective scalar channels. But all of these gains appear as power offset of the sum rate curve. On the other hand due to the presence of $(T - T_1 - T_2 - T_3)$ in the rate expression, the slope of the high SNR sum rate (the DOF) will reduce if the length of any of the training or feedback interval is used more than the minimal required. So $T_1 = M$, $T_2 = M$ and $T_3 = 1$ are the optimal training/feedback lengths at high SNR.

$$\lim_{P \rightarrow \infty} R_{\text{sum}}^{\text{FDD}} \geq \frac{T - 2M - 1}{T} M \log(P) + o(1) \quad (3.30)$$

This gives the achievable DOF of this FDD broadcast channel as

$$\text{PreLog} \geq M \left[1 - \frac{2M + 1}{T} \right]. \quad (3.31)$$

3.6.3 UL Data Transmission

FDD systems have two different frequency bands, one for UL and the other for DL transmission. Strictly speaking, as soon as the first phase ends and the BS obtains the CSIR of the UL channel, this information is sufficient to use the UL channel as MIMO MAC and full DOF (equal to M) can be achieved during the rest of the coherence interval $(T - T_1)$ [47]. But the users have to feedback the DL channel to the BS to make it fully operational. So after the second feedback phase, UL frequency band can be used as MIMO MAC for $(T - T_1 - T_2)$ symbol intervals achieving M DOF per symbol interval. Using the asymptotic high (uplink) SNR expression from Theorem 1 of [47], the sum rate in the UL direction would be

$$\lim_{P_u \rightarrow \infty} R_{\text{UL}} = \frac{T - 2M}{T} M \log(P_u) + o(1). \quad (3.32)$$

We have used minimal training lengths for first training and second feedback intervals which are optimal at high SNR. Thus the uplink channel of FDD system, working under the proposed strategy, can give per symbol multiplexing gain of $M[1 - 2M/T]$ without any extra burden of training or feedback.

3.7 Conclusions

In this chapter, we have studied the high SNR capacity of a broadcast channel with no initial assumption of channel knowledge under two scenarios.

First, when the BS is not allowed any channel information, the capacity region was shown to be bounded by the capacity of a MISO point-to-point link, hence the pre-log of the sum rate becomes trivially known and the presence of multiple users give no gain at all.

When the BS may acquire channel information, we analyze separately the two cases depending upon whether the system operates under TDD mode or FDD mode. For a system operating in TDD mode, an elegant transmission scheme is proposed which gives a multiplexing gain of $M[1 - (M + 1)/T]$. A close upper bound to this multiplexing gain is also provided by letting all the users cooperate as multiple antennas of a single user.

When the broadcast channel is part of a system working under FDD mode, transmission scheme involves some further steps as compared to that for a TDD system. This scheme achieves a multiplexing gain of $M[1 - (2M + 1)/T]$ showing that the loss of multiplexing gain with respect to a broadcast channel with perfect CSI is almost twice of the corresponding loss for a TDD system. This is due to the simple fact that the information exchange required to provide CSIT for a system working under FDD mode is almost double of that required for a TDD system.

3.A MMSE Estimate for DL Training

We want to estimate $h_{k,k}$ in the equation below when known pilot symbols are transmitted with full power for T_2 channel uses.

$$y_k = \sqrt{\frac{PT_2}{M}} h_{k,k} + \sqrt{\frac{PT_2}{M}} \sum_{j \neq k} h_{k,j} + z_k \quad (3.33)$$

$h_{k,k}$ is Gaussian distributed with zero mean and unit variance and $h_{k,j}$ is zero mean Gaussian distributed with variance σ_h^2 . Based upon this received signal and known pilots, k -th user can form the MMSE estimate of the effective scalar channel $h_{k,k}$ which is given by

$$\hat{h}_{k,k} = \frac{\mathbb{E}[h_{k,k} y_k^\dagger]}{\mathbb{E}[y_k y_k^\dagger]} y_k. \quad (3.34)$$

The expectation $\mathbb{E}[h_{k,k} y_k^\dagger]$ would be

$$\mathbb{E}[h_{k,k} y_k^\dagger] = \sqrt{\frac{PT_2}{M}} \mathbb{E}[|h_{k,k}|^2] + \sqrt{\frac{PT_2}{M}} \sum_{j \neq k} \mathbb{E}[h_{k,k} h_{k,j}^\dagger] + \mathbb{E}[h_{k,k} z_k^\dagger]. \quad (3.35)$$

The expectations in the first and the third terms of the R.H.S. of the above expression are known and we handle the second term as follows

$$\begin{aligned} \mathbb{E}[h_{k,k} h_{k,j}^\dagger] &\stackrel{a}{=} \mathbb{E}[\mathbf{h}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger \tilde{\mathbf{h}}_k] \\ &\stackrel{b}{=} \mathbb{E}[\tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger \tilde{\mathbf{h}}_k] + \mathbb{E}[\hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger \tilde{\mathbf{h}}_k] \\ &\stackrel{c}{=} \mathbb{E}[\tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger \tilde{\mathbf{h}}_k] + \mathbb{E}[\hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger] \mathbb{E}[\tilde{\mathbf{h}}_k] \\ &\stackrel{d}{=} \mathbb{E}[\tilde{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger \tilde{\mathbf{h}}_k] + \mathbb{E}[\hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k \bar{\mathbf{v}}_j^\dagger] \mathbf{0} \\ &\stackrel{e}{=} \mathbb{E}[\bar{\mathbf{v}}_j^\dagger \tilde{\mathbf{h}}_k \hat{\mathbf{h}}_k^\dagger \bar{\mathbf{v}}_k] + 0 \\ &\stackrel{f}{=} \mathbb{E}[\bar{\mathbf{v}}_j^\dagger \mathbb{E}\{\tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^\dagger\} \bar{\mathbf{v}}_k] \\ &\stackrel{g}{=} \mathbb{E}[\bar{\mathbf{v}}_j^\dagger \sigma_h^2 \mathbf{I}_M \bar{\mathbf{v}}_k] \\ &\stackrel{h}{=} \sigma_h^2 \mathbb{E}[\bar{\mathbf{v}}_j^\dagger \bar{\mathbf{v}}_k]. \end{aligned} \quad (3.36)$$

In (b), we use $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$, (c) follows as $\tilde{\mathbf{h}}_k$ is independent of the estimate $\hat{\mathbf{h}}_k$ and the beamforming vectors, (d) follows as estimation error is of zero mean, (f) follows as estimation error is independent of the beamforming

vectors and (g) follows because elements of $\tilde{\mathbf{h}}_{\mathbf{k}}$ are i.i.d. So now we have to compute the expectation of the inner product of two ZF beamforming vectors which needs to be calculated over all the channel vectors. Without loss of generality, we can assume that $k = 1$ and $j = 2$, hence we want to compute $\mathbb{E}[\bar{\mathbf{v}}_2^\dagger \bar{\mathbf{v}}_1]$. Conditioned upon the estimates of the channel vectors $\hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4 \cdots \hat{\mathbf{h}}_M$, both of these vectors lie in a 2-D null space of these estimates. $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ can also be projected in this null space of other channel vectors. Now $\bar{\mathbf{v}}_1$ will be orthogonal to the projection of $\hat{\mathbf{h}}_2$ and $\bar{\mathbf{v}}_2$ will be orthogonal to the projection of $\hat{\mathbf{h}}_1$, both restricted in this null space. As $\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2$ and hence their projections in this 2-D null space are distributed in an independent and isotropic manner, the same is true for $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$. Hence conditioned upon $\hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4 \cdots \hat{\mathbf{h}}_M$, they are independent and isotropically distributed and the mean of an isotropically distributed vector is zero.

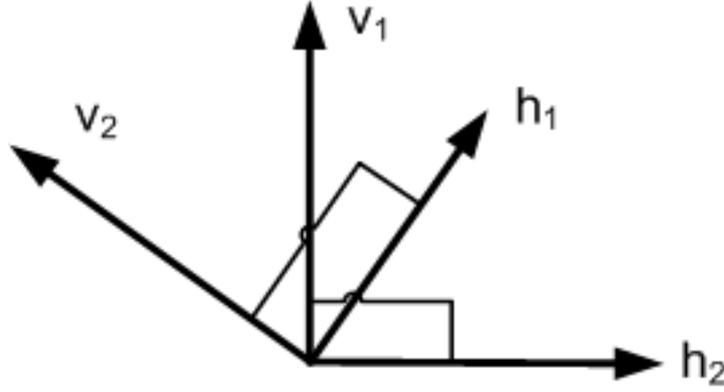


Figure 3.4: Channel Projections and corresponding ZF vectors for users 1 and 2 in 2-D null space of all other users' channel vectors.

$$\begin{aligned}
 \mathbb{E} \left[\bar{\mathbf{v}}_2^\dagger \bar{\mathbf{v}}_1 \right] &= \mathbb{E}_{\hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4, \dots, \hat{\mathbf{h}}_M} \left[\mathbb{E}_{\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2 | \hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4, \dots, \hat{\mathbf{h}}_M} \{ \bar{\mathbf{v}}_2^\dagger \bar{\mathbf{v}}_1 \} \right] \\
 &= \mathbb{E}_{\hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4, \dots, \hat{\mathbf{h}}_M} \left[\mathbb{E}_{\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2 | \hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4, \dots, \hat{\mathbf{h}}_M} \{ \bar{\mathbf{v}}_2^\dagger \} \mathbb{E}_{\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2 | \hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4, \dots, \hat{\mathbf{h}}_M} \{ \bar{\mathbf{v}}_1 \} \right] \\
 &= \mathbb{E}_{\hat{\mathbf{h}}_3, \hat{\mathbf{h}}_4, \dots, \hat{\mathbf{h}}_M} \left[\mathbf{0}^\dagger \mathbf{0} \right] \\
 &= 0
 \end{aligned} \tag{3.37}$$

Hence we conclude that

$$\mathbb{E}[h_{k,k}h_{k,j}^\dagger] = 0. \quad (3.38)$$

With this, $\mathbb{E}[h_{k,k}y_k^\dagger]$ becomes

$$\mathbb{E}[h_{k,k}y_k^\dagger] = \sqrt{\frac{PT_2}{M}}. \quad (3.39)$$

Once $\mathbb{E}[h_{k,k}h_{k,j}^\dagger]$ is known to be zero, the other expectation $\mathbb{E}[y_ky_k^\dagger]$ becomes very easy to compute.

$$\begin{aligned} \mathbb{E}[y_ky_k^\dagger] &= \frac{PT_2}{M}\mathbb{E}[|h_{k,k}|^2] + \frac{PT_2}{M}\sum_{j \neq k} \sum_{l \neq k} \mathbb{E}[h_{k,j}h_{k,l}^\dagger] + 1 \\ &= \frac{PT_2}{M} + \frac{PT_2}{M}\sum_{j \neq k} \mathbb{E}[|h_{k,j}|^2] + 1 \\ &= \frac{PT_2}{M} + \frac{PT_2}{M}(M-1)\sigma_h^2 + 1 \end{aligned} \quad (3.40)$$

Putting the values from eq. (3.39) and eq. (3.40) into eq. (3.34) gives the desired result.

Part II

CSIT Feedback : Optimization and Acquisition

Chapter 4

Feedback Optimization in MU TDD Systems

4.1 Introduction

4.1.1 Motivation

In the previous chapter, multiplexing gain of the multi-user broadcast channels with no initial CSI assumption was the focus of interest. This gain, also referred to as the DOF, becomes the capacity determining parameter at asymptotically high values of SNR and is a good performance metric for these extreme values of SNR. Unfortunately things can be drastically different at practically finite SNR values and DOF perspective may not portray the true performance. For a cognizant broadcast channel having a BS equipped with M transmit antennas and K ($K \geq M$) single antenna users, the multiplexing gain is M [17], [6] and it enjoys further a multi-user diversity [12] benefit of $M \log \log(K)$ in the sum rate [13]. The downside is that the CSIT feedback acquisition from K users could be highly burdensome, impacting badly the efficiency of the UL channel. The optimal transmission strategy for Gaussian broadcast channel has been shown to be DPC [31]. In [13], the authors gave an innovative scheme coined as Orthogonal Random Beam Forming (ORBF) where only a few bits of feedback are required from every user and the sum rate was shown to converge to the optimal

DPC sum capacity [14] for asymptotically large number of users.

The CSIT acquisition techniques and the feedback gains have been widely studied in literature but the fundamental issue, which is usually ignored, is the cost of obtaining feedback at the BS. Both the gain and the acquisition overhead increase with the amount of feedback but there is an optimal operating point (optimal amount of feedback) which maximizes the difference of gain and cost of feedback. This trade-off pushes the analysis of the **absolute gain of CSIT feedback** in general broadcast setting (with $K > M$) which can be defined as the gain in DL sum rate due to feedback taking into account the UL feedback load. A very simple example showing the importance of this absolute gain would be the ORBF transmission scheme which requires as few as $\log(M)$ bits of feedback plus a scalar from each user in the system, but considering the fact that ORBF requires the presence and the feedback from asymptotically large number of users, the absolute gain would become questionable.

The second fundamental aspect which often gets overlooked in the analysis of multi-user systems is the consideration of channel coherence time. The channels in practice have finite coherence times and when multi-user transmission strategies with multiple rounds of training, feedback and data are devised, there is possibility that channel has sufficiently changed during the preliminary training and feedback intervals and the channel information attained during these phases has become meaningless.

4.1.2 Contribution

In this chapter, we don't make any assumption of CSI, hence, initially the BS and the users are ignorant of the channel realization but they can estimate/feedback the CSI as is done in practice. To analyze the cost incurred and the benefit attainable of feedback in a meaningful and tractable fashion, the problem is simplified by selecting a TDD broadcast channel with perfect reciprocity. TDD reciprocal channels simplify the CSIT acquisition through UL pilot transmission [18], [19] contrary to the FDD systems where the users first estimate the DL channel and then send its quantized version in an UL slot. We restrict the CSIT acquisition through training only, possible due to TDD reciprocity [19], [18]. In the remaining parts of this chapter, we use the terms training and feedback synonymously due to our restriction of training based CSIT acquisition and the equivalence of training and analog feedback [19].

The fixed system bandwidth can be used for UL/DL data transmission or training/feedback. We assume that the users have no data to transmit in the

UL direction so the UL is solely reserved for channel training/feedback. Due to TDD mode of operation, any UL transmission will come at the expense of having no DL transmission during that interval, hence training/feedback gets properly accounted for. In this chapter, two transmission schemes are proposed. In the first scheme, the users, who feedback, are chosen independent of their channel realizations (hence termed as **oblivious users**). In the second scheme, the users first learn their channel information and decide to feedback based upon their channel realizations (hence termed as **informed users**). We derive a novel lower bound of the sum rate capturing the gains and the costs of CSIT acquisition which shows explicitly the rate loss w.r.t. a perfect CSI system. The simplified expressions obtained for these two schemes allow maximizing the DL sum rate achieving the cost-benefit trade-off of CSIT feedback.

4.1.3 The State of the Art

Caire et al. studied the achievable rates for multi-user MIMO DL removing all the assumptions of CSIR and CSIT for FDD systems in [48]. They gave transmission schemes incorporating all the necessary training and feedback stages and compared achievable rates for analog and quantized feedback schemes. This work was conducted under the assumption of extremely large channel coherence lengths (which permits to neglect the training and feedback overhead) and by restricting the number of users (K) equal to M . Later in [49], training and feedback parameters were optimized as a function of channel coherence length and SNR, although the number of users was still restricted to M .

In a recent work [50], the authors analyze the trade-off of multi-user diversity and the accuracy of quantized channel information at the BS. Under the restriction of a fixed number of feedback bits, they conclude that mostly accurate channel information is more important than having multi-user diversity.

The references [18] and [51] are related to our work as they treat TDD broadcast channel without any assumption of CSI. But there are major differences in the scope. They treat the case when the number of users in the system is less than the number of BS antennas, and try to exploit the channel hardening effect [52] due to large number of BS antennas, which eliminates the multi-user diversity gain completely. Moreover, in both of these references, the users are never trained about their effective channels and the data is transmitted on the expected value of the effective channel. Our analysis is for the systems with larger number of users than BS

transmit antennas because this setting is certainly more practical than its opposite counterpart. And in both of our transmission strategies (presented in following sections), the users are explicitly trained about their effective channels after precoding. The other major difference is in the achievable sum rate. Their sum rate is bounded in DL SNR, giving zero multiplexing gain even if DL and UL SNRs are of the same order, whereas our schemes achieve full multiplexing gain in this setting. A very recent related reference is [53] which is similar to [51] for the most part. Section VII of [53] gives a scheme similar to our scheme with oblivious users, given in section 4.3, but their sum rate lower bound, given in Theorem 3, involving four expectations, neither brings any insight whatsoever about the sum rate behavior nor seems amenable to any further analysis.

4.1.4 Organization

First the system model is described in section 4.2. Then in section 4.3, the transmission scheme with oblivious users is detailed and the novel lower bound of the sum rate is derived. Section 4.4 gives the same details about the scheme with informed users. The tightness of the sum rate lower bound and the accuracy of the approximate expressions are shown in section 4.5. The behavior of the sum rate for oblivious users strategy under various asymptotic regimes is investigated in section 4.6 followed by its counterpart for informed users in section 4.7. The results for optimal feedback load (optimal number of users) with finite system parameters have been explored in section 4.8. The conclusions of this work have been put together in section 4.9.

4.2 System Model

The system, we consider, consists of a BS having M transmit antennas and K ($K > M$) single-antenna user terminals. In the DL, the signal received by k -th user can be expressed as

$$y_k = \mathbf{h}_k^\dagger \mathbf{x} + z_k, \quad k = 1, 2, \dots, K \quad (4.1)$$

where $\mathbf{h}_1^\dagger, \mathbf{h}_2^\dagger, \dots, \mathbf{h}_K^\dagger$ are the channel vectors of user 1 through user K , $\mathbf{x} \in \mathbb{C}^{M \times 1}$ denotes the M -dimensional signal transmitted by the BS and z_1, z_2, \dots, z_K are independent complex Gaussian additive noise terms with zero mean and unit variances. We denote the concatenation of the channels by $\mathbf{H}_F^\dagger = [\mathbf{h}_1 \mathbf{h}_2 \cdots \mathbf{h}_K]$, so \mathbf{H}_F is $K \times M$ forward channel matrix with k -th row

(\mathbf{h}_k^\dagger) equal to the channel of k -th user. The channel input from the BS must satisfy a transmit power constraint of P i.e., $\mathbb{E}[|\mathbf{x}|^2] \leq P$. In this setting, the transmit power is equal to the true signal-to-noise ratio at each user due to normalized noise variances.

The channel is assumed to be block fading having coherence length of T symbol intervals where fading remains the same, with independent fading from one block to the next [23]. The entries of the forward channel matrix \mathbf{H}_F are i.i.d. complex Gaussian with zero mean and unit variance. Due to no CSI assumption, initially all the users and the BS are oblivious of the channel realization in each block.

For the power constraint on user terminals, we treat mainly the case of peak power constraint when the peak power per user per channel use is bounded by P_u . For the sake of completeness, we specialize the sum rate bounds for the average power constrained users in Appendix A.

4.3 Transmission Scheme with Oblivious Users

In this scheme, the feeding back users are unaware of their channel information. So they might be selected in a round-robin fashion or any other criteria independent of their channel realizations. For our block fading channel with coherence length of T symbol intervals, we divide this interval in three phases, 1) uplink training, 2) downlink training and 3) coherent data transmission. The first phase is the uplink training phase where a certain

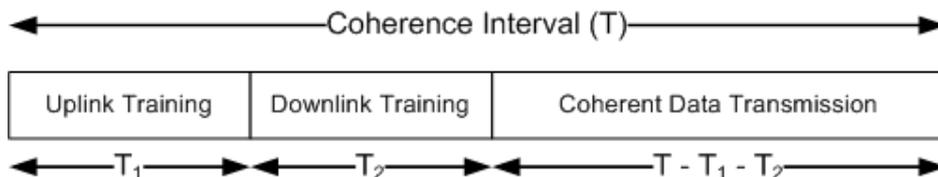


Figure 4.1: Transmission Phases for Oblivious Users

number of users train the BS about their forward channels and the BS makes an estimate of the forward channel matrix comprising of the channel vectors of these users. Based upon this channel information, the BS does the scheduling and chooses the transmit precoding which could be simple linear ZF, some non-linear strategy like vector perturbation or the optimal DPC. The second phase is the downlink training phase where the BS transmits pilots so that the scheduled users estimate their corresponding effective

channels. When this second phase ends, both sides of the broadcast channel have necessary CSI, albeit imperfect. Hence in the third data phase, the BS transmits simultaneously to the selected users who can decode the data coherently.

Below we give a detailed analysis of the three transmission phases and the necessary BS processing steps.

4.3.1 Uplink Training Phase

This training phase serves the purpose of furnishing the CSIT to the BS. In a TDD reciprocal broadcast channel, CSIT can be provided to the BS just by transmitting pilots from the users. BS estimates the users' UL channels and these are also the forward channels due to perfect reciprocity assumption. Suppose K^{obl} (superscript obl stands for oblivious users) of the K users transmit pilots, hence the length of this uplink training interval is $T_1 = \beta K^{obl}$ where $\beta \geq 1$ (β can be used when we want less users to transmit for more time and improve their channel estimates at the BS). If feedback users use orthogonal codes of length T_1 , they can transmit simultaneously with energy transmitted per user equal to $P_u T_1$. As each antenna at the BS receives transmitted code from a particular user through the channel coefficient which links this antenna to that user, the energy received for each CSIT coefficient would be $P_u T_1$. The BS, knowing the codes transmitted by the users, is able to separate them and hence the mean-square error (MSE) of CSIT for any channel coefficient when the BS employs MMSE estimation is given by

$$\sigma_h^2 = \frac{1}{P_u T_1 + 1} = \frac{1}{P_u \beta K^{obl} + 1}. \quad (4.2)$$

For k -th user with channel \mathbf{h}_k , the CSIT estimate is denoted as $\hat{\mathbf{h}}_k$ and the corresponding estimation error is $\tilde{\mathbf{h}}_k$ to give $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$. Due to MMSE channel estimation, both $\hat{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$ have zero mean i.i.d. Gaussian entries with variances $1 - \sigma_h^2$ and σ_h^2 , respectively.

4.3.2 Semi-Orthogonal User Selection and ZF Precoding

ZF linear precoding has been shown to behave quite optimally at asymptotically high SNR achieving full multiplexing gain of the sum rate [43]. Furthermore in [38], the authors showed that ZF preceded by semi-orthogonal user selection (SUS) algorithm achieves both the multiplexing gain and the multi-user diversity gain (with a large number of users). Due to simplicity,

analytical tractability and attractive performance, we choose SUS and ZF precoding as the BS transmission strategy.

We adopt the SUS algorithm of [54], [38]. So among all the users, whose CSI is available at the BS, it chooses M best users following SUS. Suppose $\hat{\mathbf{H}}(\mathbf{S})$ denotes the BS estimate of the channel matrix of the selected users. Afterward, ZF beam selection is performed for this set of selected users. In ZF precoding, unit-norm beamforming vector for k -th selected user (denoted as $\bar{\mathbf{v}}_k$) is chosen such that it is orthogonal to the channel vectors of all other selected users, i.e., $\hat{\mathbf{h}}_j^\dagger \bar{\mathbf{v}}_k = 0$, where $j \neq k$ is the index of any other selected user. If $\mathbf{H}_{\text{inv}}(\mathbf{S})$ denotes the pseudo-inverse of $\hat{\mathbf{H}}(\mathbf{S})$

$$\mathbf{H}_{\text{inv}}(\mathbf{S}) = \hat{\mathbf{H}}(\mathbf{S})^\dagger \left[\hat{\mathbf{H}}(\mathbf{S}) \hat{\mathbf{H}}(\mathbf{S})^\dagger \right]^{-1}, \quad (4.3)$$

then the precoding matrix $\bar{\mathbf{V}} = [\bar{\mathbf{v}}_1 \bar{\mathbf{v}}_2 \cdots \bar{\mathbf{v}}_M]$ can be obtained from $\mathbf{H}_{\text{inv}}(\mathbf{S})$ by normalizing all of its columns. For ZF with perfect CSIT, each user receives only the beam directed to it and no multi-user interference is experienced. For the imperfect CSIT case, there is some residual interference. If \mathbf{u} represents the vector of information symbols (u_k intended for k -th user), the transmitted signal \mathbf{x} becomes $\mathbf{x} = \bar{\mathbf{V}}\mathbf{u}$ and the signal received by k -th selected user eq. (4.1) can be expressed as follows.

$$\begin{aligned} y_k &= \mathbf{h}_k^\dagger \bar{\mathbf{V}}\mathbf{u} + z_k \\ &= \mathbf{h}_k^\dagger \bar{\mathbf{v}}_k u_k + \sum_{j \neq k} \mathbf{h}_k^\dagger \bar{\mathbf{v}}_j u_j + z_k \end{aligned} \quad (4.4)$$

4.3.3 Downlink Training Phase

It was shown in [55] that only one symbol interval is sufficient to let the M selected users learn their effective scalar channels (inner product of channel and beamforming vector). In a very recent reference [56], the authors show that this minimal training also becomes optimal with joint pilot and data processing. As this DL training length has no relation with the number of users K present in the system or the number of BS antennas (M), we assume that the selected users are able to estimate their effective scalar channels perfectly even though we ignore the overhead of this phase. This simplifies the analysis without hampering the underlying cost-benefit trade-off of the feedback.

4.3.4 Coherent Data Phase

We adopt the strategy of independent data transmission to all users from the BS with equal power allocation. So k -th user input signal u_k is Gaussian i.i.d., i.e., $u_k \sim \mathcal{CN}(0, p)$, where p is the power allocated to k -th user data stream. The BS is bound to satisfy an average power constraint of P but it does not transmit during the entire coherence block due to initial UL training phase of length T_1 . Hence for the rest of the coherence block, the BS is able to transmit an average per symbol power of $P T/(T - T_1)$ instead of P . So the power p allocated to each stream would be

$$p = \frac{P}{M} \frac{T}{T - T_1}. \quad (4.5)$$

4.3.5 Sum Rate Lower Bound

We are interested in getting an expression for the achievable sum rate of this broadcast channel which captures the gain and the cost associated with feedback. The received signal from eq. (4.4) can be further written as

$$y_k = \hat{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}} u_k + \tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}} u_k + \sum_{j \neq k} \tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{j}} u_j + z_k. \quad (4.6)$$

This uses the fact that $\mathbf{h}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{j}} = \tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{j}}$ for $k \neq j$ due to ZF beamforming and by splitting the effective channel $\mathbf{h}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}}$ in two parts, where $\hat{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}}$ is perfectly known at the BS. The above equation can be written as

$$y_k = \hat{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}} u_k + \sum_{j=1}^M \tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{j}} u_j + z_k. \quad (4.7)$$

From this equation, where we have relegated the signal part $\tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}} u_k$ into interference and by treating all the interference terms as an additional source of Gaussian noise as in [57] and [44], a lower bound of the SINR of k -th user can be written as

$$\text{SINR}_k^{\text{obl}} = \frac{p |\hat{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{k}}|^2}{1 + p \sum_{j=1}^M \mathbb{E} |\tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{j}}|^2}. \quad (4.8)$$

The variance of each interference coefficient ($\tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \bar{\mathbf{v}}_{\mathbf{j}}$) can be computed based upon the fact that BS does MMSE estimation which makes estimation error $\tilde{\mathbf{h}}_{\mathbf{k}}$ (with variance σ_h^2 per channel entry) independent of any function of

channel estimates ($\hat{\mathbf{h}}_{\mathbf{k}}$) of which beamforming vectors are one particular example.

$$\mathbb{E}|\tilde{\mathbf{h}}_{\mathbf{k}}^\dagger \tilde{\mathbf{v}}_{\mathbf{j}}|^2 = \mathbb{E}\left[\tilde{\mathbf{v}}_{\mathbf{j}}^\dagger \mathbb{E}\left(\tilde{\mathbf{h}}_{\mathbf{k}} \tilde{\mathbf{h}}_{\mathbf{k}}^\dagger\right) \tilde{\mathbf{v}}_{\mathbf{j}}\right] = \sigma_h^2 \mathbb{E}\left[\tilde{\mathbf{v}}_{\mathbf{j}}^\dagger \tilde{\mathbf{v}}_{\mathbf{j}}\right] = \sigma_h^2 \quad (4.9)$$

Furthermore by using $\hat{\mathbf{h}}_{\mathbf{k}} = \sqrt{1 - \sigma_h^2} \mathbf{g}_{\mathbf{k}}$ where $\mathbf{g}_{\mathbf{k}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ in the numerator, the SINR becomes

$$\text{SINR}_k^{\text{obl}} = \frac{1 - \sigma_h^2}{1 + pM\sigma_h^2} p |\mathbf{g}_{\mathbf{k}}^\dagger \tilde{\mathbf{v}}_{\mathbf{k}}|^2, \quad (4.10)$$

where the coefficient of $p |\mathbf{g}_{\mathbf{k}}^\dagger \tilde{\mathbf{v}}_{\mathbf{k}}|^2$ represents the SINR loss factor w.r.t. a system with perfect CSIT and CSIR as $\mathbf{g}_{\mathbf{k}}$'s (and hence $\tilde{\mathbf{v}}_{\mathbf{k}}$'s) are perfectly known at the BS. So during the data phase, the lower bound (LB) of the per symbol sum rate can be written as

$$\text{LB} = \sum_{k=1}^M \mathbb{E}_{\mathbf{g}_{\mathbf{k}}} \log \left(1 + \frac{1 - \sigma_h^2}{1 + pM\sigma_h^2} p |\mathbf{g}_{\mathbf{k}}^\dagger \tilde{\mathbf{v}}_{\mathbf{k}}|^2 \right), \quad (4.11)$$

where the users being transmitted have been selected using SUS algorithm. Putting the value of p , we get

$$\text{LB} = \sum_{k=1}^M \mathbb{E}_{\mathbf{g}_{\mathbf{k}}} \log \left(1 + \frac{1 - \sigma_h^2}{1 + P \frac{T}{T-T_1} \sigma_h^2} \frac{T}{T-T_1} \frac{P}{M} |\mathbf{g}_{\mathbf{k}}^\dagger \tilde{\mathbf{v}}_{\mathbf{k}}|^2 \right). \quad (4.12)$$

If one deals with the same system (K users and M BS antennas) with perfect assumption of CSI ($\sigma_h^2 = 0, T_1 = 0$), the sum rate obtained through SUS and ZF beamforming would be

$$R_{\text{ZF}}(K, M, P) = \sum_{k=1}^M \mathbb{E}_{\mathbf{g}_{\mathbf{k}}} \log \left(1 + \frac{P}{M} |\mathbf{g}_{\mathbf{k}}^\dagger \tilde{\mathbf{v}}_{\mathbf{k}}|^2 \right). \quad (4.13)$$

And for large user regime, it was shown in [38] to be well-approximated by

$$R_{\text{ZF}}(K, M, P) \approx M \log \left(1 + \frac{P}{M} \log(K) \right). \quad (4.14)$$

So the lower bound of the sum rate (during the data phase) can be written in terms of the sum rate of a perfect CSI system as

$$\text{LB} = R_{\text{ZF}}(K^{\text{obl}}, M, P_m) \quad (4.15)$$

where P_m is the reduced SNR given by

$$P_m = \frac{(1 - \sigma_h^2) \frac{T}{T-T_1}}{1 + P \frac{T}{T-T_1} \sigma_h^2} P. \quad (4.16)$$

By taking into account the loss of coherence interval T due to feedback (training) interval of length $T_1 = \beta K^{obl}$, the per symbol sum rate lower bound for this oblivious scheme becomes

$$\text{LB}^{obl} = \frac{T - \beta K^{obl}}{T} \sum_{k=1}^M \mathbb{E}_{\mathbf{g}_k} \log \left(1 + \frac{(1 - \sigma_h^2) \frac{T}{T - \beta K^{obl}} P}{1 + P \sigma_h^2 \frac{T}{T - \beta K^{obl}}} \frac{P}{M} |\mathbf{g}_k^\dagger \bar{\mathbf{v}}_k|^2 \right). \quad (4.17)$$

The biggest virtue of this lower bound is that it gives the achievable sum rate of this scheme in terms of the sum rate of a perfect CSI system (employing SUS and ZF precoding) with loss appearing as an SNR reduction factor and as reduced multiplexing gain due to feedback interval.

We can separate the channel vector \mathbf{g}_k in its norm $\|\mathbf{g}_k\|$ and unit norm direction vector $\bar{\mathbf{g}}_k$ as $\mathbf{g}_k = \|\mathbf{g}_k\| \bar{\mathbf{g}}_k$ to get

$$\text{LB}^{obl} = \frac{T - \beta K^{obl}}{T} \sum_{k=1}^M \mathbb{E}_{\mathbf{g}_k} \log \left(1 + \frac{(1 - \sigma_h^2) \frac{T}{T - \beta K^{obl}} P}{1 + P \sigma_h^2 \frac{T}{T - \beta K^{obl}}} \frac{P}{M} \|\mathbf{g}_k\|^2 |\bar{\mathbf{g}}_k^\dagger \bar{\mathbf{v}}_k|^2 \right). \quad (4.18)$$

The factor $|\bar{\mathbf{g}}_k^\dagger \bar{\mathbf{v}}_k|^2$ has a limited effect because when there are only M users, this factor is beta distributed with parameters $\beta(1, M - 1)$ [21], [17] whose average value is $\frac{1}{M}$. Contrary to this, when the BS has CSIT of infinite number of users, this factor is still upper bounded by 1.

The factor $\|\mathbf{g}_k\|^2$ (chi-square distributed with $2M$ degrees of freedom) was shown to grow logarithmically with the number of users whose channel information is available at the BS (see eq. (A10) in [13]) using the results from order statistics for asymptotically large number of users. With large number of users, this scaling was shown to hold for each selection stage of SUS algorithm in [38] and the simulations showed this scaling to hold even when the number of users is of the same order as that of the BS transmit antennas. This permits us to approximate $\|\mathbf{g}_k\|^2$ by $\log(K^{obl})$ in the above expression. Theoretically this scaling kicks in when K^{obl} is sufficiently large but we show in section 4.5 that this holds very well even for K^{obl} close to M . Using this approximation¹ and putting the value of σ_h^2 from eq. (4.2),

¹There is another way to get to eq. (4.19). In SUS version of [38], where the users are forced to be ϵ -orthogonal, they showed that, for large user regime, the dominant term in the expansion of the squared norm of the effective channel (cascade of channel and beamforming vector) is logarithmic in the number of users, i.e., $\log(K^{obl})$.

the above sum rate will become

$$\text{SR}^{obl} = \frac{T - \beta K^{obl}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - \beta K^{obl})} \frac{P_u \beta K^{obl}}{P_u \beta K^{obl} + 1} \log(K^{obl})}{1 + P \frac{T}{(T - \beta K^{obl})} \frac{1}{P_u \beta K^{obl} + 1}} \right). \quad (4.19)$$

Due to the approximation made at this final step, this sum rate expression is not necessarily a lower bound but it closely follows the lower bound and the true sum rate of the system.

4.4 Transmission Scheme with Informed Users

This scheme also consists of the transmission phases through which the BS and a subset of the users get necessary CSI. We call this the scheme with “informed users” as the users who feedback are no more randomly selected. These users are selected based upon their channel realizations in a manner to be described shortly. First we describe this scheme and then characterize its sum rate in the next subsections.

4.4.1 Transmission Scheme

This scheme divides the coherence length of T symbol intervals in four phases, 1) initial downlink training, 2) uplink training, 3) downlink training and 4) coherent data transmission.

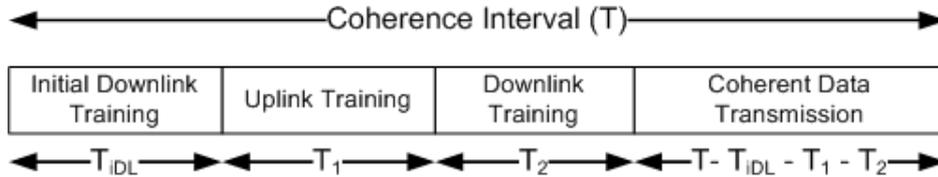


Figure 4.2: Transmission Phases for Informed Users

In the first phase, termed as initial downlink training, the BS transmits DL pilots based upon which all the users estimate their corresponding channel vectors. As the BS has M antennas, this training interval length is lower bounded by M , giving $T_{iDL} \geq M$ and is independent of the number of users K . As the BS can transmit with sufficient power to provide good estimates, we don't take into account the estimation error during this phase but only subtract $T_{iDL} = M$ from the coherence length. Another point is that when

the selected users transmit pilots so that the BS obtains the CSIT, this CSIT does not depend upon the users' estimates. Hence even if users have slightly noisy estimates of their channels, this imperfection doesn't propagate to CSIT.

Once the users have acquired the information regarding their respective channels, there could be plenty of criteria to prioritize some of the users depending upon their channel realizations but we restrict ourselves to the simple scheme where K^{inf} best users w.r.t. the channel norm are selected for feedback. Hence the BS receives the CSIT from K^{inf} of the K users which have the largest channel norms for the current channel coherence block.

The next three transmission phases are exactly similar as those for the transmission scheme with oblivious users. In the second phase of uplink training, K^{inf} users with the largest channel norms feedback their channel information to the BS which boils down to UL pilot transmission due to TDD system. Based upon this channel information, the BS uses SUS algorithm to further select M best users and computes corresponding ZF beamforming vectors. In the third phase of downlink training, BS transmits through these ZF beamforming vectors so that the selected users estimate their corresponding effective scalar channels. We neglect the overhead of this phase for the reasons mentioned in the scheme with oblivious users. The last phase is the coherent data transmission phase with equal power allocated to all M independent users' streams.

Important Remark. In this transmission scheme involving informed users, we select just the strongest users (having largest channel norms) who train the BS about their channels. Strictly speaking, this is impractical as how can the users know about being the strongest or not with only the information about their own channels. But the underlying idea is to evaluate how much feedback load (how many users) should be there to maximize the DL sum rate if good users feedback. Then, in practice, those many users can be made to feedback, on the average, by intelligent selection of a threshold with which users compare their channel strengths locally as detailed in [58] and decide to feedback or not, and by designing a proper UL channel access protocol. This threshold will be a function of the total number of users, their channel statistics, the number of BS antennas and the optimal number of users who should feedback.

4.4.2 Sum Rate Lower Bound

We'll be quite brief here as the treatment resembles a lot as done for the oblivious users' setting. If every user is constrained with a peak per symbol

power constraint of P_u and K^{inf} users transmit pilots in the UL direction, the feedback length would be $T_1 = \beta K^{inf}$ where $\beta \geq 1$. The MSE of CSIT per channel coefficient at the BS is given by

$$\sigma_h^2 = \frac{1}{P_u \beta K^{inf} + 1}. \quad (4.20)$$

Following the same steps as done for the oblivious users' case, we can write the rate expression for a single selected user. Again the BS is simultaneously transmitting data to M users. For this scheme with informed users, we have an initial step of DL training so the length of the data phase reduces to $T - M - \beta K^{inf}$, where the additional M factor appears due to initial DL training and βK^{inf} denotes the length of the feedback phase. Thus the DL sum rate for informed scheme with peak power constrained users is given by

$$\text{SR}^{inf} = \frac{T - M - \beta K^{inf}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - \beta K^{inf})} \frac{P_u \beta K^{inf}}{P_u \beta K^{inf} + 1} \log(K)}{1 + P \frac{T}{(T - \beta K^{inf})} \frac{1}{P_u \beta K^{inf} + 1}} \right). \quad (4.21)$$

One striking difference in the sum rate of this scheme is the channel strength factor (multi-user diversity factor) of $\log(K)$ where K is the number of users in the system, which came out to be $\log(K^{obl})$ for the scheme with oblivious users. This difference arises due to the fact that only the best users, with respect to the strength of their channel norms, feedback in this scheme.

4.5 Accuracy of the Approximate Sum Rate Expressions

To obtain the sum rate expressions, we derived the novel sum rate lower bound and then used the large user regime approximation at the final step. If we want to see how closely these sum rate expressions capture the true sum rate behavior of the two schemes, it's sufficient to show the accuracy for one of the rate expressions because the same approximation is made for both.

We choose the sum rate expression for oblivious peak power constrained users. To see how closely it captures the behavior of the sum rate with different system parameters, we compare it with true sum rate. We use eq. (4.19) in the form as below

$$\text{SR} = M \log \left(1 + \frac{\frac{P}{M} \frac{P_u \beta K^{obl}}{P_u \beta K^{obl} + 1} \log(K^{obl})}{1 + P \frac{1}{P_u \beta K^{obl} + 1}} \right). \quad (4.22)$$

For fixed values of T , β and K^{obl} , we absorbed the constant factor of $T/(T - \beta K^{obl})$ in SNR both in numerator and denominator of SINR and leaving the constant multiplying factor of $(T - \beta K^{obl})/T$ outside of the logarithm. So in this form it shows the behavior of the sum rate expression for any coherence length T . To obtain the true sum rate corresponding to different system parameters, we use Monte-Carlo simulations. All the steps in the transmission strategy e.g. feedback, SUS scheduling and ZF beam formation are replicated and then SINR at each user is evaluated. The rates thus obtained are compared with those obtained from the approximate analytical expression.

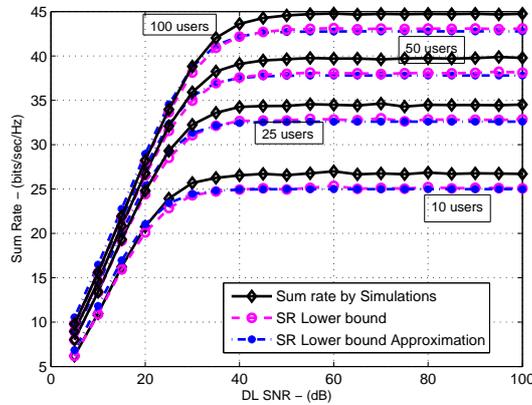


Figure 4.3: Sum Rate versus SNR. Approximation expression captures closely the true sum rate vs. SNR behavior for a broad range of number of users feeding back.

Fig. 4.3 shows the plots of the sum rate versus DL SNR. Uplink power constraint for each user has been fixed to 10 dB and the BS is equipped with 4 antennas. The curves for the true sum rate, the lower bound from eq. (4.17) and the approximation of the LB have been plotted when 10, 25, 50 and 100 users feedback their channel information to the BS. Approximate expression captures very closely the true sum rate behavior for any SNR, even at saturation. This saturation of the sum rate is caused by the imperfect CSIT based upon which ZF beamforming vectors are computed (shown for quantized FB in [17] and for analog feedback in [55]).

Fig. 4.4 shows the plots of the sum rate with varying number of feeding back users K^{obl} . Again 4-antenna BS has users available with power con-

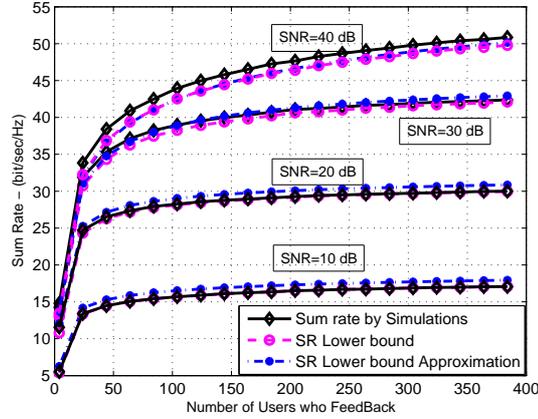


Figure 4.4: Sum Rate versus Number of Users. Approximate expression shows good match with the behavior of the true sum rate vs. number of feeding back users for many different values of SNR.

straint of 10 dB each. The curves have been plotted at DL SNR of 10, 20, 30 and 40 dB. These curves show the tightness of the lower bound of eq. (4.17) and the approximate expression in capturing the multi-user diversity benefit.

Based upon the above plots of the true sum rate, the lower bound and its approximation, it becomes evident that the lower bound is sufficiently tight and the approximate expression fully captures the true sum rate behavior for any SNR and for any number of users in the system.

4.6 Asymptotic Analysis with Oblivious Users

In this section, we analyze how the sum rate of oblivious users behaves in different asymptotic regimes. Although this analysis is asymptotic, yet it gives valuable insight about the optimal amount of feedback and its utilization.

4.6.1 Noise Limited Regime

For noise limited regime, the power available to the BS is very limited, i.e., $P \rightarrow 0$. In this regime, the noise in each user's received signal fully

dominates the interference coming from the beams of other selected users.

$$\text{SR}^{obl} = \frac{T - \beta K^{obl}}{T} M \log \left(1 + \frac{P}{M} \frac{T}{(T - \beta K^{obl})} (1 - \sigma_h^2) \log(K^{obl}) \right). \quad (4.23)$$

Using the approximation of $\log(1 + x) \approx x$ for very small x , the above sum rate becomes

$$\text{SR}^{obl} = P(1 - \sigma_h^2) \log(K^{obl}). \quad (4.24)$$

At very low DL SNRs, this sum rate can be obtained by receiving feedback from K^{obl} users and transmitting only to the strongest user in a peaky manner. This asymptote shows that, at low SNR, multiplexing gain is lost and multi-user diversity gives logarithmic instead of double logarithmic gain.

4.6.2 Interference Limited Regime

Due to imperfect CSIT at the BS, the interference power scales up with the increase in DL power. So when DL SNR is very large, the interference completely dominates the noise and the sum rate becomes

$$\text{SR}^{obl} = \frac{T - \beta K^{obl}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - \beta K^{obl})} (1 - \sigma_h^2) \log(K^{obl})}{P \frac{T}{(T - \beta K^{obl})} \sigma_h^2} \right), \quad (4.25)$$

which reduces to

$$\text{SR}^{obl} = \frac{T - \beta K^{obl}}{T} M \log \left(1 + \frac{\log(K^{obl})}{M} \left(\frac{1}{\sigma_h^2} - 1 \right) \right). \quad (4.26)$$

Basically this equation gives the sum rate where saturation occurs due to imperfect CSIT and the rate grows no more with the increase in DL SNR. This equation shows that the sum rate saturation point can be pushed further in SNR by improving the CSIT quality, i.e., by reducing σ_h^2 .

4.6.3 Asymptotically Large Number of Users

For the oblivious scheme with peak power constrained users, the sum rate expression is completely independent of the number of users present in the system and depends only upon the number of users who actually feedback.

4.7 Asymptotic Analysis with Informed Users

In this section, we analyze how the sum rate of the scheme with informed users behaves in different asymptotic regimes.

4.7.1 Noise Limited Regime

In noise limited regime, the noise completely dominates the interference at each active user.

$$\text{SR}^{inf} = \frac{T - M - \beta K^{inf}}{T} M \log \left(1 + \frac{P}{M} \frac{T}{(T - \beta K^{inf})} (1 - \sigma_h^2) \log(K) \right) \quad (4.27)$$

Again by using the approximation of $\log(1 + x) \approx x$ for very small x , the above sum rate becomes

$$SR^{inf} = \left(P \frac{T - M - \beta K^{inf}}{T - \beta K^{inf}} \right) (1 - \sigma_h^2) \log(K). \quad (4.28)$$

4.7.2 Interference Limited Regime

When $P \rightarrow \infty$, the interference resulting from the imperfect CSIT completely dominates the noise power at each user (which is being served) and the sum rate becomes

$$\text{SR}^{inf} = \frac{T - M - \beta K^{inf}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - \beta K^{inf})} (1 - \sigma_h^2) \log(K)}{P \frac{T}{(T - \beta K^{inf})} \sigma_h^2} \right), \quad (4.29)$$

which reduces to

$$\text{SR}^{inf} = \frac{T - M - \beta K^{inf}}{T} M \log \left(1 + \frac{\log(K)}{M} \left(\frac{1}{\sigma_h^2} - 1 \right) \right). \quad (4.30)$$

This expression has the same form as that obtained for oblivious scheme. Hence, the sum rate saturation point can be found by plugging in the MSE value of CSIT and can eventually be pushed further by refining the CSIT quality.

4.7.3 Asymptotically Large Number of Users

For the informed scheme, for any power constraint imposed on user terminals, the effective signal strength increases with $\log(K)$ as only strong users feedback to the BS and hence get scheduled. For this reason, the sum rate shows unbounded growth with the number of users present in the system.

4.8 Feedback Load Optimization

As we have shown in section 4.5 that the approximate sum rate expressions match closely the true sum rate, they can be used to optimize over how many users should feedback the BS about their channel information. We had introduced two parameters namely the number of users who feedback and the β factor. If we look carefully the final sum rate expressions eq. (4.19) in section 4.3 and eq. (4.21) in section 4.4, we note that β always appears in product form with the number of users who feedback. For both of these expressions, β can always be selected to be 1 without any loss of optimality of the sum rate. One must admit that in certain regimes, significant savings in the total UL energy (used for training/feedback) can be attained by optimizing over β but for the DL sum rate maximization perspective, $\beta = 1$ does not involve any loss. We adapt this for the simplicity of the presentation and hence the amount of feedback load appears as the number of users who feedback. We formulate the problem for the scheme with oblivious users. The sum rate for this scheme was developed to be

$$\text{SR}^{obl} = \frac{T - K^{obl}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - K^{obl})} \frac{P_u K^{obl}}{P_u K^{obl} + 1} \log(K^{obl})}{1 + P \frac{T}{(T - K^{obl})} \frac{1}{P_u K^{obl} + 1}} \right). \quad (4.31)$$

This sum rate which is a function of K^{obl} (the number of users who feedback) can be optimized over this parameter. If the optimal number of feeding back users which maximizes the sum rate is denoted by K^{obl*} , the objective function can be written as

$$K^{obl*} = \arg \max_{K^{obl}} \frac{T - K^{obl}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - K^{obl})} \frac{P_u K^{obl}}{P_u K^{obl} + 1} \log(K^{obl})}{1 + P \frac{T}{(T - K^{obl})} \frac{1}{P_u K^{obl} + 1}} \right). \quad (4.32)$$

The solution for K^{obl*} can be found by solving $\text{SR}'^{obl} = 0$ where SR'^{obl} denotes the first derivative of SR^{obl} with respect to K^{obl} . Unfortunately it does not assume closed form expression but the optimal value of feeding back users for both of the schemes can be found by trivially simple numerical optimization.

4.8.1 Optimal Users vs DL SNR

First we see how the optimal number of users (feeding back) scales with SNR. We plot the curves of the optimal number of users versus SNR in Fig. 4.5 and plot corresponding sum rates achieved by using that optimal number

of users for each value of SNR in Fig. 4.6. We take $T=1000$ symbol intervals, there are 200 users in the system with per user peak power constraint of 5 dB and the BS is equipped with $M = 4$ antennas. It's evident that the gains with optimal feedback are undeniable as the sum rate with only feedback from M users is much less than the sum rate with optimal number of users. The saturation of the sum rate because of imperfect CSIT depicted in Fig. 4.6 has already been investigated in [55] and [17].

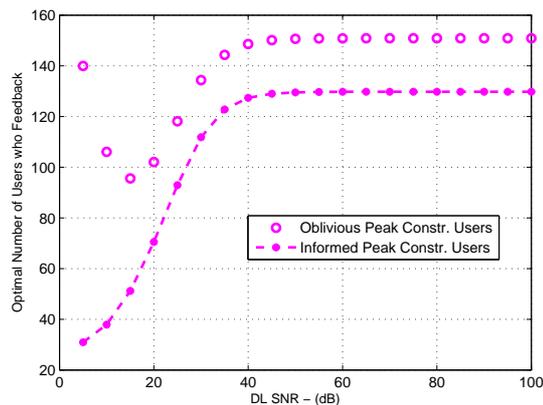


Figure 4.5: Optimal Users versus SNR

The behavior of the curves of optimal number of users feeding back for two schemes versus SNR is not very straight forward. At high SNR (interference limited regime, see eqs. (4.26) and (4.30)), both schemes require very good quality CSIT and due to peak power constrained users with $\beta = 1$, it translates to obtaining feedback from each user for longer intervals which comes out to be a lot of users transmitting feedback (users have orthogonal codes and hence can be separated).

At low SNR both curves show very different behavior. The reason is at low SNR, the system is basically noise limited and the multi-user diversity factor is very important hence the users with very strong channels should be scheduled (see eq. (4.24) and (4.28)). In informed users scheme, only the strong users feedback so it requires feedback from small number of users and as SNR increases and CSIT quality starts becoming more important, more users start feeding back (this is the way to improve CSIT quality for peak-power constrained users). The scheme with oblivious users (where users feedback independent of their channel realizations) requires feedback

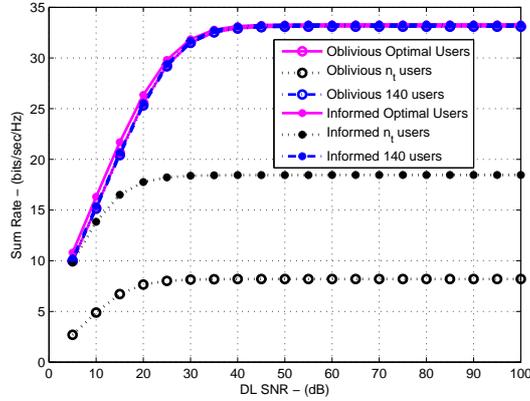


Figure 4.6: Sum Rate with Optimal Users versus SNR

from a large number of users initially to enjoy multi-user diversity but that consumes a lot of coherence time in feedback so the number of users who feedback decreases further and later starts increasing again to provide high quality CSIT.

Although the optimal feeding back users in two schemes differ significantly from lower to medium SNR values, the corresponding sum rates overlap completely. At low to medium SNR values, informed user strategy gives slightly higher rates but this difference is minor. We have also plotted the sum rate when 140 users (this is the number of users at high DL SNR) feedback in each coherence interval for both schemes. These curves also overlap fully the sum rates of two schemes with optimal feedback load (Fig. 4.6). It indicates that for a fixed channel coherence length, a fixed reasonable value of feeding back users (normally much larger than M) can achieve significantly the cost-benefit trade-off of feedback. In other words, the sum rate as a function of SNR is not very sensitive to the number of users who feedback.

4.8.2 Optimal Users vs Channel Coherence Time

We now analyze how the optimal number of users behaves with the change in channel coherence time. We plot two figures, one showing the optimal number of users versus coherence interval in Fig. 4.7 and the other showing the sum rate corresponding to the optimal number of users versus coherence interval in Fig. 4.8. Here BS has $M = 4$ antennas, its power constraint is

20 dB and there are 500 users in the system with each user restricted to a peak power constraint of 5 dB.

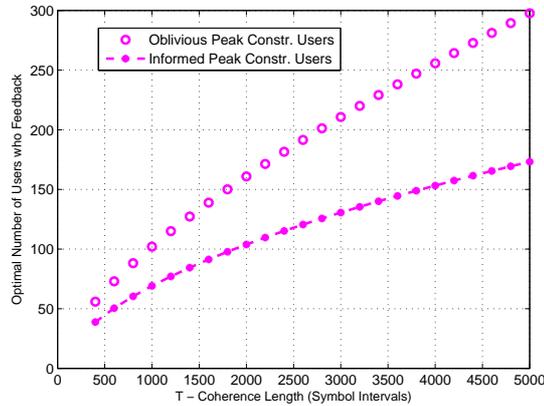


Figure 4.7: Optimal Users versus Coherence Length

The curves of the optimal number of users versus channel coherence time show almost linear increase. For smaller values of the coherence interval, small number of users is optimal so that not a lot of coherence interval gets consumed in feedback. For very large values of the coherence interval, feedback from a large number of users is optimal so as to select the good users with good quality CSIT. Thus the number of users, who feedback, scales up with the increase in channel coherence time. The optimal number of users, who feedback, is always more for the scheme with oblivious users than in the case of informed users. This behavior can be anticipated from Fig. 4.5 which shows that from low to medium DL SNR values, the optimal number of users in the oblivious scheme is more than that in the informed scheme.

The sum rate curves for optimal users have been plotted at $P = 20$ dB so informed user scheme performs better as can be guessed from Fig. 4.6. Sum rate curves have also been plotted for fixed number of users (200) feeding back but contrary to the sum rate versus SNR curves where a single suitable number of users feeding back captures the gain of optimal feedback, here it is not possible to find one such number of users capturing the sum rate gains as of with optimal feedback load. So the sum rate as a function of T is relatively sensitive to the number of users who feedback.

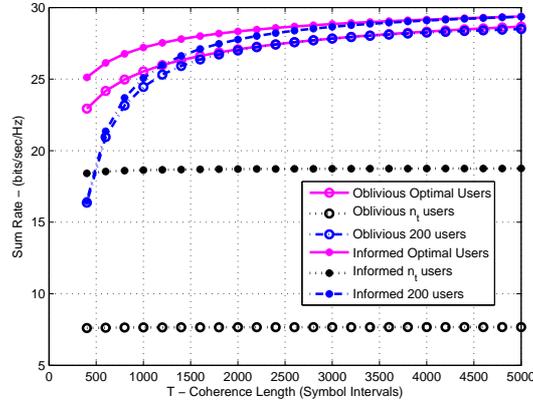


Figure 4.8: Sum Rate with Optimal Users versus Coherence Length

4.9 Conclusions

In this chapter, we have studied the problem of determining the optimal amount of feedback for sum rate maximization of the broadcast channel with no initial assumption of CSI. We introduced two transmission strategies for providing the CSIT to the BS and derived a novel tight lower bound which clearly shows the rate loss w.r.t. a perfect CSI system. The corresponding simplified sum rate expressions, incorporating the gains of the feedback and the cost of exchange of information, allow us to determine the optimal amount of feedback for any set of system parameters. Moreover, the asymptotic analysis carried out for both schemes gives us insight into the amount and the split of the optimal feedback between obtaining multi-user diversity and accurate channel information for better inter-user interference cancellation. The optimal split involves in which region of the sum rate, system is operating. The noise limited regime demands the use of feedback to harness fully the multi-user diversity benefit whereas the interference limited regime requires the use of feedback resources to get fine quality CSIT because the MSE of CSIT is the principal factor to determine the saturation point of the sum rate versus SNR. In between these two regimes, the feedback split depends upon the contribution of multi-user diversity gain, the importance of CSIT quality in the sum rate and the fraction of coherence interval used for feedback.

4.A Average Power Constrained Users

We treated the case when the users in the system are peak power constrained. For average power constrained users, the feedback behavior will change as the MSE of CSIT changes for average power constrained users. We keep the discussion to minimum as we believe this power constraint to be unrealistic and impractical. If there are K users in the system having channel coherence length of T and each constrained to an average power P_{avg} per channel use, the total UL energy available in each coherence block is $P_{avg}KT$. Now if K^{obl} users feedback, each one of these can transmit an energy of $P_{avg}KT/K^{obl}$. Here the use of orthogonal codes is not necessary because, due to weaker power constraint, the users can transmit all of their available power in short intervals. Hence with this energy transmitted for every channel coefficient, the MSE of CSIT at the BS will be

$$\sigma_h^2 = \frac{1}{\frac{P_{avg}KT}{K^{obl}} + 1}. \quad (4.33)$$

Although the feeding back users will be able to transmit pilots with larger energy (if $K \gg K^{obl}$), yet they will be transmitting only occasionally, the probability of which will reduce with more users in the system and hence long term average power constraint will be satisfied. Such users' power constraint for the transmission in the UL direction was employed in [47]. The sum rate expressions for the two schemes when the users are average power constrained can be obtained by plugging in the MSE of CSIT from eq. (4.33).

$$\text{SR}^{obl} = \frac{T - \beta K^{obl}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - \beta K^{obl})} \frac{\frac{P_{avg}KT}{K^{obl}}}{\frac{P_{avg}KT}{K^{obl}} + 1} \log(K^{obl})}{1 + P \frac{T}{(T - \beta K^{obl})} \frac{1}{\frac{P_{avg}KT}{K^{obl}} + 1}} \right) \quad (4.34)$$

$$\text{SR}^{inf} = \frac{T - M - \beta K^{inf}}{T} M \log \left(1 + \frac{\frac{P}{M} \frac{T}{(T - \beta K^{inf})} \frac{\frac{P_{avg}KT}{K^{inf}}}{\frac{P_{avg}KT}{K^{inf}} + 1} \log(K)}{1 + P \frac{T}{(T - \beta K^{inf})} \frac{1}{\frac{P_{avg}KT}{K^{inf}} + 1}} \right) \quad (4.35)$$

Then these sum rate expressions can be optimized to evaluate the amount of feedback (the number of users) to maximize the sum rate.

Chapter 5

Novel CSIT Acquisition for Reciprocal Channels

5.1 Introduction

5.1.1 Motivation

Multi-antenna transmitters and receivers are instrumental to optimizing the performance of bandwidth and power limited wireless communication systems which has mainly been the focus of interest in previous chapters. In the DL, in particular, the communication between a multiple-antenna enabled BS and one or more users with either a single or multiple antenna each can be significantly enhanced through the use of scheduling, beamforming and power allocation algorithms, be it in single user or multi-user mode (spatial division multiplexing). This requires the availability of CSIT at the BS [6], [15], except when the number of users reaches an asymptotic (large) regime in which case random opportunistic beamforming scheme can be exploited [59], [13]. Not only the presence of CSIT is required for harnessing the gains of DL channel, the quality of CSIT also plays a very important role. It has been shown that for reasonable DL performance, the quality of CSIT needs to be refined with DL SNR otherwise the multiplexing gain is lost [17]. Hence good quality CSIT is desired at the expense of minimum resource consumption. This has motivated the proposal of many techniques

for providing the CSIT in an efficient manner. Proposals for how to provide CSIT roughly fall in two categories depending upon the chosen duplexing scheme for the considered wireless network. In the case of TDD systems, it was always assumed that CSIT should exploit the reciprocity of UL and DL channels, so as to avoid the use of any resource consuming feedback channel [18], [19]. The way reciprocity is exploited in the current TDD systems, is through the use of a training sequence sent by the user on the UL, based on which the BS first builds an estimate of the UL channel which in turn serves as an estimate for the DL channel in the next DL slot [18]. In FDD systems, UL and DL portions of the bandwidth are normally quite apart and hence the channel realizations can be safely assumed to be independent of each other. This lack of channel reciprocity motivates instead the use of a dedicated feedback link in which the user conveys the information, about the estimated DL channel, back to the BS. Recently, several interesting strategies have been proposed for how to best use a limited feedback channel and still provide the BS with exploitable CSIT (see [20], [17], [21], [22] and the references therein for further details).

5.1.2 Contribution

In this chapter, we focus on the problem of CSIT acquisition in a TDD system. We show that the traditional approach using training sequences exclusively fails to fully exploit the channel reciprocity. The key shortcoming is as follows: when sending a training sequence in the UL of a traditional TDD system, the user allows the BS to estimate the channel by a classical channel estimator (it can be a least-square (LS) estimator or minimum mean square error (MMSE) based, just to name a few). However, note that the user itself has the knowledge of the channel coefficients (obtained during the current DL frame or from the DL synchronization sequence or other control signals or even from the previous DL frames if the channel is correlated in time) but, regrettably, does not exploit that knowledge in order to facilitate the CSIT acquisition by the BS. Interestingly, by contrast, in FDD systems, the user exploits its DL channel knowledge by quantizing the channel and sending the result over a dedicated feedback link. In FDD case, UL training is used by the BS solely for UL data detection as this UL training cannot give any direct information to the BS about the DL channel.

We point out that in TDD systems there is a unique opportunity to combine the advantages of both forms of CSIT acquisition. In doing so, we obtain a new CSIT acquisition scheme mixing the classical channel estimation using training with the quantized limited channel feedback of the

same channel. This gives us a framework for fully utilizing the channel reciprocity in a TDD setup and it improves the classical trade-off between the CSIT quality and the amount of training/feedback resource used. The optimal CSIT acquisition structure is characterized under this novel framework. This hybrid CSIT acquisition setup gives rise to a very interesting joint estimation and detection problem for which we propose two iterative algorithms. We further propose a sub-optimal outage rate based approach which helps us to optimize the fixed resource partitioning between training and quantized feedback phases. We adapt this optimization framework to use it with practical constellations like QSPK and 16-QAM. The results obtained confirm our intuition and clearly demonstrate the benefit of this hybrid (mix of training and quantized feedback) approach for upcoming TDD systems. The quality of CSIT acquired through this new hybrid combining is much better than that of CSIT obtained through classical training under a global training and feedback resource constraint.

5.1.3 The State of the Art

In previous work, Caire et al. studied the achievable rates for multi-user MIMO DL removing all the assumptions of CSIR and CSIT for FDD systems in [48]. They gave transmission schemes incorporating all the necessary training and feedback stages and compared achievable rates either with analog feedback or with quantized feedback. The reference [60] studies the decay rate of the feedback distortion versus SNR with analog and digital quantized feedback for FDD systems. A very recent work [61] studies combining the analog and digital feedback for FDD systems. All of these works fundamentally differ from this work as there is no channel reciprocity in FDD systems and hence there is no point in combining the UL training and the quantized feedback of the DL channel.

Some other contributions [18], [51], [55], [53] and [62] analyze the sum rate of TDD systems starting without any assumption of CSI but restrict the CSIT acquisition through training only. [19] does a comparison of TDD systems versus FDD systems in terms of CSIT acquisition accuracy. [63] studies the diversity-multiplexing trade-off [64] of two-way SIMO channels when TDD is the mode of operation. All of these references treat no-CSI TDD systems but all acquire CSIT through training only. Hence this work becomes the first attempt of CSIT combining through training and quantized feedback for reciprocal channels.

5.1.4 Organization

The system model and the mode of CSIR acquisition is given in section 5.2, followed by the classical CSIT acquisition for FDD and TDD systems in section 5.3. The optimal CSIT acquisition strategy combining training and feedback is outlined in section 5.4. Two iterative and one non-iterative algorithms for the joint estimation and detection have been proposed in section 5.5. The simplified outage-rate based framework to optimize the resource split appears in section 5.6 followed by its adaption for practical constellations in section 5.7. The simulation results have been provided in section 5.8, followed by the conclusions in section 5.9.

5.2 System Model and CSIR Acquisition

We consider the two way communication in a cell between a single BS, equipped with M antennas, and a single antenna mobile user. The DL channel $\mathbf{h} \in \mathbb{C}^M$ is assumed to be flat-fading with independent complex Gaussian zero-mean unit-variance entries. We assume block fading channel so each channel realization stays constant for T channel uses [23] which can be accordingly partitioned between UL and DL data transmissions.

The goal of this work is to provide a reliable estimate of the DL channel to the BS, which in turn can be used for scheduling/beamforming/precoding purposes as was the case in previous two chapters. However in this chapter, we focus on the acquisition issue of the channel knowledge and not about its use in MIMO transmission schemes.

In the DL, the received signal at the user for T_{DL} channel uses is given by

$$\mathbf{y}_{\text{dl}} = \mathbf{X}_{\text{dl}}\mathbf{h} + \mathbf{z}_{\text{dl}}, \quad (5.1)$$

where $\mathbf{X}_{\text{dl}} \in \mathbb{C}^{T_{\text{DL}} \times M}$ is the signal transmitted by the BS for T_{DL} channel uses (satisfying BS power constraint), $\mathbf{z}_{\text{dl}} \in \mathbb{C}^{T_{\text{DL}}}$ is the complex Gaussian noise with independent zero-mean unit-variance entries and $\mathbf{y}_{\text{dl}} \in \mathbb{C}^{T_{\text{DL}}}$ is the observation sequence during this T_{DL} -length interval.

If the user has to use the above DL system equation for DL channel estimation, for identifiability of M -dimensional channel at the user's side, the length T_{DL} of the transmitted data (the training sequence in this case) should be larger than M . Based upon the knowledge of the training sequence \mathbf{X}_{dl} and the observed signal \mathbf{y}_{dl} , the user can estimate the DL channel \mathbf{h} using various techniques. The LS estimate, denoted as $\check{\mathbf{h}}_{\text{LS}}$, would be [44]

$$\check{\mathbf{h}}_{\text{LS}} = \left(\mathbf{X}_{\text{dl}}^\dagger \mathbf{X}_{\text{dl}} \right)^{-1} \mathbf{X}_{\text{dl}}^\dagger \mathbf{y}_{\text{dl}}. \quad (5.2)$$

The user can employ the MMSE estimation criteria and in this case the estimate is given by

$$\check{\mathbf{h}}_{\text{MMSE}} = \left(\mathbf{X}_{\text{dl}}^\dagger \mathbf{X}_{\text{dl}} + \mathbf{I}_M \right)^{-1} \mathbf{X}_{\text{dl}}^\dagger \mathbf{y}_{\text{dl}}. \quad (5.3)$$

5.3 Classical CSIT Acquisition

We now briefly review the classical approaches for acquiring CSIT at the BS in FDD and TDD systems. We shall build upon the equations below in order to present our ideas later.

5.3.1 FDD Systems

A typical UL frame for FDD systems is shown in Fig. 5.1 where the initial T_{fb} channel uses are reserved for feedback. For the BS to be able to decode

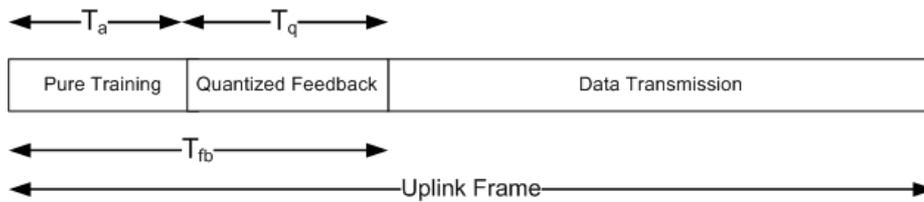


Figure 5.1: Uplink frame structure: Total feedback length is divided between UL training and quantized feedback phases.

the feedback properly (sent as UL payload), it should first know/estimate the UL channel (denoted as $\mathbf{h}_{\mathbf{u}} \in \mathbb{C}^M$). If the user sends a normalized training sequence $\mathbf{x}_{\mathbf{a}} \in \mathbb{C}^{1 \times T_a}$ of length T_a in the UL direction, the signal received at the BS for T_a channel uses is given by

$$\mathbf{Y}_{\mathbf{a}} = \sqrt{P_u} \mathbf{h}_{\mathbf{u}} \mathbf{x}_{\mathbf{a}} + \mathbf{Z}_{\mathbf{a}}, \quad (5.4)$$

where $\mathbf{Z}_{\mathbf{a}} \in \mathbb{C}^{M \times T_a}$ represents the spatio-temporally white Gaussian noise with zero-mean unit-variance entries and $\mathbf{Y}_{\mathbf{a}} \in \mathbb{C}^{M \times T_a}$ is the received signal at M antennas of the BS during this T_a -length training interval. P_u represents the user's peak power constraint which is equal to the UL SNR at every BS antenna due to the normalized noise variances. After observing $\mathbf{Y}_{\mathbf{a}}$, the BS can make an estimate $\hat{\mathbf{h}}_{\mathbf{u}}$ of the UL channel $\mathbf{h}_{\mathbf{u}}$, knowing the training sequence $\mathbf{x}_{\mathbf{a}}$. Estimation techniques like LS or MMSE as described in the previous section can be applied.

In FDD systems, the mobile station obtains the DL channel estimate $\check{\mathbf{h}}$ from the DL frame as described in the previous section. If Q denotes the quantization function, then for the DL channel estimate $\check{\mathbf{h}}$, its quantized version (the index of the closest codeword in the codebook) is given by $Q(\check{\mathbf{h}})$. Afterward user maps this index (sequence of bits) into a sequence of constellation symbols, using the mapping function denoted by S . Let the finite cardinality set of all mapped codewords be denoted by \mathcal{CB} . Hence the feedback of the DL channel would be

$$\mathbf{x}_{\mathbf{q}} = S(Q(\check{\mathbf{h}})), \quad (5.5)$$

where $\mathbf{x}_{\mathbf{q}} \in \mathbb{C}^{1 \times T_q}$ is the T_q dimensional row vector of the normalized constellation symbols. This parameter T_q determines the size of the codebook. The signal received at the BS upon transmission of $\mathbf{x}_{\mathbf{q}}$ is

$$\mathbf{Y}_{\mathbf{q}} = \sqrt{P_u} \mathbf{h}_{\mathbf{u}} \mathbf{x}_{\mathbf{q}} + \mathbf{Z}_{\mathbf{q}}, \quad (5.6)$$

where $\mathbf{Y}_{\mathbf{q}}$ and $\mathbf{Z}_{\mathbf{q}}$ are $M \times T_q$ matrices of the received signal and the noise respectively at M antennas of the BS during this T_q length feedback interval. Based upon the estimate $\hat{\mathbf{h}}_{\mathbf{u}}$ of the UL channel $\mathbf{h}_{\mathbf{u}}$ and the received feedback $\mathbf{Y}_{\mathbf{q}}$, BS tries to recover the DL channel feedback (quantized version, $\mathbf{x}_{\mathbf{q}}$) using the optimum (although relatively complex) maximum likelihood (ML) sequence estimation technique.

$$\hat{\mathbf{h}} = \arg \min_{\check{\mathbf{h}}} \|\mathbf{Y}_{\mathbf{q}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{u}} S(Q(\check{\mathbf{h}}))\|^2 \quad (5.7)$$

The search space will be restricted to the codebook, hence the BS, at best, can estimate the quantized version of the channel.

5.3.2 TDD Systems

If the communication system is operating under TDD mode, DL and UL channels are reciprocal, hence $\mathbf{h}_{\mathbf{u}} = \mathbf{h}$. So if a user transmits pilot sequence on the UL (like eq. (5.4)), the simple (UL) channel estimation at the BS furnishes CSIT due to UL and DL channel reciprocity. In the past, this has been the classical way of getting CSIT in TDD systems [18], [19].

5.4 Optimal Training and Feedback Combining in TDD Systems

The classical training based CSIT acquisition for TDD systems ignores the fact that user knows the DL channel and the CSIT acquisition based only on

the quantized feedback for FDD systems cannot use the channel reciprocity whereas in TDD systems both can be exploited at the same time.

We propose a novel hybrid two stage CSIT acquisition strategy which exploits the channel reciprocity and user's channel knowledge at the same time. We assume perfect channel knowledge at the user's side for ease of exposition¹ and later, in section 5.8.4, we show results removing this perfect CSIR assumption. Working under a constraint of fixed resource available for CSIT acquisition (T_{fb} channel uses and user's power constraint of P_u), our strategy consists of dividing this interval in two phases as shown in Fig. 5.1 for the UL frame of FDD, whereas the classical strategy will use all this resource of T_{fb} channel uses for pilot sequence transmission. The first stage of this hybrid approach, termed as "pure training", is the transmission of training sequence from the user to the BS for T_a channel uses and the received signal will be

$$\mathbf{Y}_a = \sqrt{P_u} \mathbf{h} \mathbf{x}_a + \mathbf{Z}_a. \quad (5.8)$$

(See eq. (5.4) for the dimensions of all parameters.)

The optimal training based estimate, denoted as $\hat{\mathbf{h}}_a$, based upon the observed signal \mathbf{Y}_a and knowing \mathbf{x}_a will be

$$\hat{\mathbf{h}}_a = \arg \min_{\mathbf{h}} \|\mathbf{Y}_a - \sqrt{P_u} \mathbf{h} \mathbf{x}_a\|^2. \quad (5.9)$$

The second stage, termed as "quantized feedback", consists of the transmission of quantized channel, already known at the user, for T_q channel uses and the received signal will be

$$\mathbf{Y}_q = \sqrt{P_u} \mathbf{h} \mathbf{x}_q + \mathbf{Z}_q, \quad (5.10)$$

(See eq. (5.6) for the dimensions of all parameters.)

where $\mathbf{x}_q = S(Q(\mathbf{h})) \in \mathcal{CB}$. This equation reveals the intriguing aspect that the BS needs to acquire \mathbf{h} which appears both as the channel and the transmitted feedback \mathbf{x}_q . The BS can try to decode only the quantized channel information based upon the knowledge of $\hat{\mathbf{h}}_a$ (obtained as in eq. (5.9) making use of pure training \mathbf{x}_a)

$$\hat{\mathbf{h}}_q = \arg \min_{\mathbf{x}_q \in \mathcal{CB}} \|\mathbf{Y}_q - \sqrt{P_u} \hat{\mathbf{h}}_a \mathbf{x}_q\|^2. \quad (5.11)$$

The optimal CSIT will be obtained by the joint estimation and detection (of \mathbf{h} and \mathbf{x}_q respectively) based upon the observation of \mathbf{Y}_a and \mathbf{Y}_q ,

¹In general, the CSIR quality at the users' side is much better. Firstly the DL pilots are global (they are not transmitted per user contrary to the UL pilots) and secondly, the BS can surely pump larger power as compared to small hand-held mobile devices.

knowing \mathbf{x}_a and assuming an optimal split between the training and the quantized feedback phases where the total CSIT acquisition interval length T_{fb} is constrained as $T_a + T_q = T_{fb}$.

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \|\mathbf{[Y}_a \mathbf{Y}_q] - \sqrt{P_u} \mathbf{h}[\mathbf{x}_a \ S(Q(\mathbf{h}))]\|^2 \quad (5.12)$$

The optimal solution requires a double minimization and does not seem to bear a closed form expression for $\hat{\mathbf{h}}$.

5.5 Joint Channel Estimation and Feedback Detection Algorithms

We give three algorithms in this section which separately solve the estimation and the detection problem of the joint minimization of eq. (5.12). The first two algorithms are iterative which separately solve the estimation and detection problems and iterate till convergence. These algorithms have been closely inspired by [65] which proposes similar algorithms for joint blind estimation and detection for signal separation. We have made modifications for our requirements where data aided channel estimation after the initialization step and the presence of channel as “data” (feedback) make them unique and suitable for the concerned objective. The third algorithm is just the single-shot solution of the joint estimation and detection. Owing to its simplicity, it allows us to further optimize the resource split between training and quantized feedback in the next section.

5.5.1 Iterative Estimation and Detection

We describe below our algorithm.

Step 1) Initial channel estimation based upon the pilots only

$$\hat{\mathbf{h}}_a^0 = \arg \min_{\mathbf{h}} \|\mathbf{Y}_a - \sqrt{P_u} \mathbf{h} \mathbf{x}_a\|^2, \quad (5.13)$$

which is a simple least squares problem with the solution

$$\hat{\mathbf{h}}_a^0 = \mathbf{Y}_a \mathbf{x}_a^\dagger (\mathbf{x}_a \mathbf{x}_a^\dagger)^{-1} \frac{1}{\sqrt{P_u}}. \quad (5.14)$$

$$i = 1 \quad (5.15)$$

Superscript denotes the iteration number.

Step 2) At iteration i , do enumeration over all the codes in the codebook assuming that the channel $\hat{\mathbf{h}}_{\mathbf{a}}^{i-1}$ is perfectly known.

$$\hat{\mathbf{x}}_{\mathbf{q}}^i = \arg \min_{\mathbf{x}_{\mathbf{q}} \in \mathcal{CB}} \|\mathbf{Y}_{\mathbf{q}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\mathbf{q}}\|^2 \quad (5.16)$$

Step 3) Regenerate extended pilot sequence \mathbf{x}_{ext} (pilots and detected feedback)

$$\mathbf{x}_{\text{ext}}^i = [\mathbf{x}_{\mathbf{a}} \hat{\mathbf{x}}_{\mathbf{q}}^i], \quad \mathbf{Y}_{\text{ext}} = [\mathbf{Y}_{\mathbf{a}} \mathbf{Y}_{\mathbf{q}}]. \quad (5.17)$$

Step 4) Channel estimation based upon the extended pilots (i.e. knowing $\mathbf{x}_{\text{ext}}^i$)

$$\hat{\mathbf{h}}_{\mathbf{a}}^i = \arg \min_{\mathbf{h}} \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \mathbf{h} \mathbf{x}_{\text{ext}}^i\|^2 \quad (5.18)$$

$$\hat{\mathbf{h}}_{\mathbf{a}}^i = \mathbf{Y}_{\text{ext}} \mathbf{x}_{\text{ext}}^i \dagger (\mathbf{x}_{\text{ext}}^i \mathbf{x}_{\text{ext}}^i \dagger)^{-1} \frac{1}{\sqrt{P_u}} \quad (5.19)$$

Step 5) If $\hat{\mathbf{x}}_{\mathbf{q}}^i \neq \hat{\mathbf{x}}_{\mathbf{q}}^{i-1}$ or $\hat{\mathbf{h}}_{\mathbf{a}}^i \neq \hat{\mathbf{h}}_{\mathbf{a}}^{i-1}$, $i = i + 1$ and go to Step 2.

The final channel estimate $\hat{\mathbf{h}}$ is the channel vector corresponding to $\hat{\mathbf{x}}_{\mathbf{q}}^i$ in the codebook.

The formal convergence statement and its proof for this iterative algorithm appear in the following theorem.

Theorem 2 (Convergence for Iterative Estimation and Detection Algorithm). *Let $\hat{\mathbf{h}}_{\mathbf{a}}^i$ be the estimated channel and $\hat{\mathbf{x}}_{\mathbf{q}}^i$ be the detected feedback, both at i -th iteration of the iterative estimation and detection algorithm. Let the residual function $f(\hat{\mathbf{h}}_{\mathbf{a}}, \mathbf{x}_{\text{ext}}; \mathbf{Y}_{\text{ext}}) \triangleq \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}} \mathbf{x}_{\text{ext}}\|^2$ be selected as the descent function for this algorithm. Then there exists some positive integer j such that for any $i \geq j$, $\hat{\mathbf{x}}_{\mathbf{q}}^i = \hat{\mathbf{x}}_{\mathbf{q}}^j$ and $\hat{\mathbf{h}}_{\mathbf{a}}^i = \hat{\mathbf{h}}_{\mathbf{a}}^j$.*

Proof. The residual descent function $f(\hat{\mathbf{h}}_{\mathbf{a}}, \mathbf{x}_{\text{ext}}; \mathbf{Y}_{\text{ext}}) = \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}} \mathbf{x}_{\text{ext}}\|^2$ is non-negative and continuous. Considering the residual function at i -th it-

eration:

$$\begin{aligned}
f\left(\hat{\mathbf{h}}_{\mathbf{a}}, \mathbf{x}_{\text{ext}}^i; \mathbf{Y}_{\text{ext}}\right) &\stackrel{a}{=} \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^i \mathbf{x}_{\text{ext}}^i\|^2 \\
&\stackrel{b}{=} \min_{\mathbf{h}} \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \mathbf{h} \mathbf{x}_{\text{ext}}^i\|^2 \\
&\stackrel{c}{\leq} \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\text{ext}}^i\|^2 \\
&\stackrel{d}{=} \|\mathbf{Y}_{\mathbf{a}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\mathbf{a}}\|^2 + \|\mathbf{Y}_{\mathbf{q}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \hat{\mathbf{x}}_{\mathbf{q}}^i\|^2 \\
&\stackrel{e}{=} \|\mathbf{Y}_{\mathbf{a}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\mathbf{a}}\|^2 + \min_{\mathbf{x}_{\mathbf{q}} \in \mathcal{CB}} \|\mathbf{Y}_{\mathbf{q}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\mathbf{q}}\|^2 \\
&\stackrel{f}{\leq} \|\mathbf{Y}_{\mathbf{a}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\mathbf{a}}\|^2 + \|\mathbf{Y}_{\mathbf{q}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \hat{\mathbf{x}}_{\mathbf{q}}^{i-1}\|^2 \\
&\stackrel{g}{=} \|\mathbf{Y}_{\text{ext}} - \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1} \mathbf{x}_{\text{ext}}^{i-1}\|^2 \\
&\stackrel{h}{=} f\left(\hat{\mathbf{h}}_{\mathbf{a}}^{i-1}, \mathbf{x}_{\text{ext}}^{i-1}; \mathbf{Y}_{\text{ext}}\right)
\end{aligned} \tag{5.20}$$

Equalities d and g make use of the property of the Frobenius norm [66]. The set of equations above shows that each single iteration of the algorithm over estimation and detection causes to monotonically reduce the residual function unless iterates converge. This monotonic reduction of the descent function, its non-negativity and the fact that $\mathbf{x}_{\mathbf{q}}$ belongs to a finite set (codes of the codebook) and hence corresponding iterates of the estimation sub-problem are also finite prove the convergence of this algorithm to the locally optimal solution in a finite number of steps. The globally optimal solution is achieved by having a good initial point which depends upon the training part as confirmed by our simulations. \square

5.5.2 Simplified Iterative Estimation and Detection

This algorithm is very similar to the previous algorithm in essence but the difference arises at the detection step. The second step of the previous algorithm, the ML detection of the quantized code from the codebook, is computationally quite onerous, especially for codebooks with large cardinality. So we replace this enumeration step with least squares detection followed by mapping on the codebook. So the Step 2 of the previous algorithm gets replaced by two sub-steps.

Step 2-A) At iteration i , do LS detection of the quantized feedback assuming $\hat{\mathbf{h}}_{\mathbf{a}}^{i-1}$ as the perfectly known channel

$$\hat{\mathbf{x}}_{\text{LS}}^i = (\hat{\mathbf{h}}_{\mathbf{a}}^{\dagger i-1} \hat{\mathbf{h}}_{\mathbf{a}}^{i-1})^{-1} \hat{\mathbf{h}}_{\mathbf{a}}^{\dagger i-1} \mathbf{Y}_{\mathbf{q}} \frac{1}{\sqrt{P_u}}. \tag{5.21}$$

Step 2-B) Do hard detection on the constellation symbols which will map the LS channel estimate to the nearest code in the codebook.

$$\hat{\mathbf{x}}_{\mathbf{q}}^i = \text{HardDetection}(\hat{\mathbf{x}}_{\text{LS}}^i) \quad (5.22)$$

This helps to significantly reduce the computational complexity. Later results show that this does not involve any discernible performance degradation.

5.5.3 Single-Shot Estimation and Detection

Among the proposed algorithms, this is the simplest and the fastest algorithm for the joint estimation and detection problem where the channel estimation and the feedback detection are performed (separately) only once. Step 1) Channel estimation based only upon the pilots

$$\hat{\mathbf{h}}_{\mathbf{a}} = \arg \min_{\mathbf{h}} \|\mathbf{Y}_{\mathbf{a}} - \sqrt{P_u} \mathbf{h} \mathbf{x}_{\mathbf{a}}\|^2. \quad (5.23)$$

We can employ either the LS or the MMSE estimation technique.

Step 2) Detection of the feedback $\mathbf{x}_{\mathbf{q}}$ assuming channel $\hat{\mathbf{h}}_{\mathbf{a}}$ is perfectly known. This detection problem can be solved either by enumerating all the codewords like the first algorithm or by simple LS like the second algorithm or even by applying MMSE filter.

5.6 Outage Based Training and Feedback Partitioning

5.6.1 Definitions and Initial Setup

The solution for the optimal CSIT estimate, $\hat{\mathbf{h}}$ in eq. (5.12), requires joint estimation and detection. Furthermore, the fixed resource (T_{fb} channel uses) needs to be optimally split between the training length T_a and the quantized feedback length T_q . Even if, as a simplification, we focus separately on training based estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ (given in eq. (5.9)) and digital feedback based estimate $\hat{\mathbf{h}}_{\mathbf{q}}$ (given in eq. (5.11)), two questions arise: i) how the fixed CSIT acquisition interval T_{fb} should be split between training and feedback?, and ii) how the two estimates should be combined to get the final estimate?

We use the minimization of the mean-square error (MSE) of the final CSIT (formally defined below) as the criterion for the optimal resource split, thus answering the first question for which we give the proper framework in

the next subsection. It has been shown in [17] that the principal factor in the DL sum rate loss due to imperfect CSI is the MSE of CSIT. Hence the minimization of the MSE of CSIT is equivalent to the maximization of the system wide sum rate, the most commonly adopted system performance metric. Furthermore, we propose to use the quantized feedback based estimate $\hat{\mathbf{h}}_{\mathbf{q}}$ as the final CSIT estimate $\hat{\mathbf{h}}$ due to better channel diversity exploitation properties of digital transmission as an answer to the second question. It may give the impression that the training based estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ goes wasted but in reality quantized feedback $\mathbf{x}_{\mathbf{q}}$, which provides $\hat{\mathbf{h}}_{\mathbf{q}}$, is decoded based upon this training based estimate $\hat{\mathbf{h}}_{\mathbf{a}}$.

This optimization framework consists of first providing a training based estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ to the BS in the training interval of T_a channel uses. In the second interval of T_q channel uses, the user sends the quantized version of its unit-norm channel direction information (CDI) vector which we assume to be perfectly known at the user. As the channel stays constant for each acquisition interval, this feedback transmission is equivalent to the transmission over slow fading channels for which deep channel fades (causing outage) are the typical error events [3]. We define the “outage” as an event when the channel realization and the quality of the training based estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ (a function of T_a) don’t allow the BS to successfully decode the feedback information. Let $\epsilon(T_a, b)$ be the outage probability in the quantized feedback phase when transmitting b bits per channel use on the UL feedback channel. Thus b is the $\epsilon(T_a, b)$ -outage rate [3] of the UL channel in the quantized feedback phase. So the user can send a total of $B = bT_q$ feedback bits at $\epsilon(T_a, b)$ outage. Although the constellations used in practice have 2^b points where b must be a positive integer, for the time being we relax this restriction and allow positive real values for b .

We define the squared CDI error as the sine squared of the angle (θ) between the true channel direction vector $\bar{\mathbf{h}}$ and the BS estimated direction vector $\hat{\mathbf{h}}$, denoted as $\sigma^2(\mathbf{h}, \hat{\mathbf{h}})$.

$$\sigma^2(\mathbf{h}, \hat{\mathbf{h}}) \triangleq \sin^2(\theta) = 1 - \cos^2(\theta) = 1 - |\bar{\mathbf{h}}^\dagger \hat{\mathbf{h}}|^2 \quad (5.24)$$

Further the MSE of CSIT is defined to be the expected value of the squared CDI error at the transmitter and denoted as σ^2 . Although it’s a slight abuse of notation but it has been shown that the CDI plays a vital role both for single-user and multi-user scenarios [17].

For the quantization of M -dimensional unit-norm CDI at the user, we employ random vector quantization (RVQ). For RVQ, the exact expression

for the mean-square quantization error σ_q^2 has been given in [67], [17] as

$$\sigma_q^2 = 2^B \beta \left(2^B, \frac{M}{M-1} \right), \quad (5.25)$$

where B is the total number of feedback bits (i.e. the codebook consists of 2^B codes) and β represents the beta function which is defined in terms of the Gamma function as $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. However it turns out that a simple and tight upper bound given in reference [17] suffices:

$$\sigma_q^2 \leq 2^{\frac{-B}{M-1}}. \quad (5.26)$$

5.6.2 Resource Split between Training and Feedback

Theorem 3 (The minimization of the MSE of CSIT). *Under the training and feedback combining strategy, the MSE of CSIT σ^2 is minimized as a result of the following optimization governing the fixed resource (T_{fb}) split between the training T_a and the quantized feedback interval T_q and the outage rate b :*

$$\sigma^{2*} = \min_{T_a, b} \left[2^{\frac{-b(T_{fb}-T_a)}{M-1}} + \epsilon(T_a, b) \right] \quad (5.27)$$

The constraints for this minimization are:

$$1 \leq T_a \leq T_{fb} \quad \text{and} \quad 0 \leq b \quad (5.28)$$

The outage probability in the feedback interval $\epsilon(T_a, b)$ and the outage rate b are linked by the relation:

$$b = \log \left(1 + \frac{P_u^2 T_a}{2(P_u + P_u T_a + 1)} F^{-1}(\epsilon(T_a, b)) \right), \quad (5.29)$$

where P_u is the user's power constraint and $F^{-1}(\cdot)$ is the inverse of the standard cumulative distribution function (CDF) of χ_{2M}^2 distributed variable.

Proof. The proof consists of two parts. First we show the argument of minimization to be an upper bound on the MSE of CSIT and in the second part, the relation between $\epsilon(T_a, b)$ and b is derived.

Upper bound on the MSE of CSIT:

During the feedback phase, when the channel is not in outage and the BS is able to decode the feedback correctly, there is only quantization error in

the final CSIT estimate. On the other hand, when the channel is in outage (happens with probability $\epsilon(T_a, b)$), the BS cannot decode the feedback information. Hence the MSE of CSIT σ^2 can be written as

$$\begin{aligned}\sigma^2 &= (1 - \epsilon(T_a, b)) \sigma_q^2 + \epsilon(T_a, b) \mathbb{E} \sigma_{\mathbf{h} \neq \hat{\mathbf{h}}}^2(\mathbf{h}, \hat{\mathbf{h}}) \\ &\leq (1 - \epsilon(T_a, b)) \sigma_q^2 + \epsilon(T_a, b) \\ &\leq \sigma_q^2 + \epsilon(T_a, b),\end{aligned}\tag{5.30}$$

where σ_q^2 is the mean-square quantization error and $\sigma_{\mathbf{h} \neq \hat{\mathbf{h}}}^2(\mathbf{h}, \hat{\mathbf{h}})$ represents the MSE of CSIT when the channel is in outage (which means a feedback error occurs). The first inequality is obtained as $\mathbb{E} \sigma_{\mathbf{h} \neq \hat{\mathbf{h}}}^2(\mathbf{h}, \hat{\mathbf{h}})$ is upper-bounded by 1. Putting the value of σ_q^2 from eq. (5.26) using $B = bT_q$ and $T_{fb} = T_a + T_q$ in eq. (5.30), we get the desired upper bound of the MSE of CSIT as

$$\sigma^2 \leq 2^{\frac{-b(T_{fb}-T_a)}{M-1}} + \epsilon(T_a, b),\tag{5.31}$$

which concludes the first part of our proof.

The Interplay of Training and Quantized Feedback: The MSE bound of the CSIT eq. (5.31) is the desired performance metric. Its minimization gives us the optimal values for T_a , T_q and b (the number of feedback bits per channel use - this parameter governs the constellation size and hence the quantization error) for a fixed resource T_{fb} . This bound shows us the basic trade-off involved. If the total number of feedback bits $B = bT_q$ is made large (either by choosing a large rate b per channel use in the feedback channel or by making T_q large), it will allow the user to select a larger codebook (with 2^B codewords) and hence the quantization error will be negligible. But this strategy will plague the final CSIT estimation error by introducing a lot of outage (due to large b or poor channel estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ caused by small $T_a = T_{fb} - T_q$). On the other hand for a small number of total feedback bits B , the degradation due to outage probability will fade away, but there will be fewer codewords in the codebook and hence a large quantization error.

The relation of b and $\epsilon(T_a, b)$:

Pilot sequence transmission from the user to the BS for an interval of length T_a , given in eq. (5.8), can be equivalently written in a simplified form as

$$\mathbf{y}_{\mathbf{a}} = \sqrt{P_u T_a} \mathbf{h} + \mathbf{z}_{\mathbf{a}},\tag{5.32}$$

where P_u is the user's power constraint and $\mathbf{y}_{\mathbf{a}}$, \mathbf{h} , $\mathbf{z}_{\mathbf{a}}$ are the received signal, the channel vector and the noise respectively, all column vectors of dimension

M . The BS can make MMSE estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ of the channel \mathbf{h} as

$$\hat{\mathbf{h}}_{\mathbf{a}} = \frac{\sqrt{P_u T_a}}{P_u T_a + 1} \mathbf{y}_{\mathbf{a}}. \quad (5.33)$$

As the i.i.d. channel entries are standard Gaussian, the MMSE estimation error $\tilde{\mathbf{h}}_{\mathbf{a}} = \mathbf{h} - \hat{\mathbf{h}}_{\mathbf{a}}$ has also Gaussian i.i.d. entries as $\tilde{\mathbf{h}}_{\mathbf{a}} \sim \mathcal{CN}(\mathbf{0}, \sigma_a^2 \mathbf{I}_M)$ and the MSE per channel coefficient σ_a^2 is given by

$$\sigma_a^2 = \frac{1}{P_u T_a + 1}. \quad (5.34)$$

Similarly the estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ has Gaussian i.i.d. entries and is distributed as $\hat{\mathbf{h}}_{\mathbf{a}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{P_u T_a}{P_u T_a + 1} \mathbf{I}_M\right)$.

Now we focus our attention on the quantized feedback interval of the CSIT acquisition, given in eq. (5.10). The signal received during one symbol interval of this phase is given by

$$\mathbf{y}_{\mathbf{q}} = \sqrt{P_u} \mathbf{h} x_q + \mathbf{z}_{\mathbf{q}}, \quad (5.35)$$

where x_q represents the scalar feedback symbol transmitted by the user and $\mathbf{y}_{\mathbf{q}}, \mathbf{h}, \mathbf{z}_{\mathbf{q}}$ are M -dimensional column vectors representing respectively the observed signal, the channel and the noise for this particular symbol interval. To decode this information, the BS uses the estimate $\hat{\mathbf{h}}_{\mathbf{a}}$ that it developed during the training phase. The above equation can be written as

$$\mathbf{y}_{\mathbf{q}} = \sqrt{P_u} \hat{\mathbf{h}}_{\mathbf{a}} x_q + \sqrt{P_u} \tilde{\mathbf{h}}_{\mathbf{a}} x_q + \mathbf{z}_{\mathbf{q}}. \quad (5.36)$$

The average effective signal-to-noise-ratio (denoted as SNR_{eff}) at the BS during the feedback interval relegating the signal part which comes associated with the channel estimation error $\tilde{\mathbf{h}}_{\mathbf{a}}$ into noise and treating $\hat{\mathbf{h}}_{\mathbf{a}}$ as the perfectly known channel is given by:

$$\text{SNR}_{\text{eff}} = \frac{P_u \|\hat{\mathbf{h}}_{\mathbf{a}}\|^2}{P_u \sigma_a^2 + 1}. \quad (5.37)$$

Plugging in the value of σ_a^2 from eq. (5.34), SNR_{eff} will become

$$\text{SNR}_{\text{eff}} = \frac{P_u \|\hat{\mathbf{h}}_{\mathbf{a}}\|^2}{\frac{P_u}{P_u T_a + 1} + 1}. \quad (5.38)$$

We can do a small change of variable as $\frac{2(P_u T_a + 1)}{P_u T_a} \|\hat{\mathbf{h}}_{\mathbf{a}}\|^2$ represents a standard chi-square random variable having $2M$ degrees of freedom (DOF), denoted as χ_{2M}^2 . So the SNR_{eff} becomes

$$\text{SNR}_{\text{eff}} = \frac{P_u^2 T_a}{2(P_u + P_u T_a + 1)} \chi_{2M}^2. \quad (5.39)$$

The outage probability $\epsilon(T_a, b)$ during this feedback interval corresponding to the outage rate b bits per channel use can be written as

$$\begin{aligned} \epsilon(T_a, b) &= \mathbb{P}[\log(1 + \text{SNR}_{\text{eff}}) \leq b] \\ &= \mathbb{P}\left[\log\left(1 + \frac{P_u^2 T_a}{2(P_u + P_u T_a + 1)} \chi_{2M}^2\right) \leq b\right], \end{aligned} \quad (5.40)$$

where \mathbb{P} denotes the probability of an event. This relation can be inverted to obtain the outage rate b corresponding to the outage probability $\epsilon(T_a, b)$, as given below

$$b = \log\left(1 + \frac{P_u^2 T_a}{2(P_u + P_u T_a + 1)} F^{-1}(\epsilon(T_a, b))\right), \quad (5.41)$$

where $F^{-1}(\cdot)$ is the inverse of the CDF of χ_{2M}^2 distributed variable. This concludes the proof.

The analytical solution to the minimization in Theorem 3 does not bear closed form expression but its numerical optimization is quite trivial. \square

5.7 Optimization Setup with Practical Constellations

In the previous optimization procedure, we had relaxed the restriction of practical constellations and allowed any positive real values for the outage rate b bits per channel use. But this is not true for the practical communication systems as the constellations used always have number of points equal to an integer power of 2, i.e., b can only take an integer value. We propose two simple strategies in the following sub-sections to handle this issue which arises due to this limitation of practical constellations.

5.7.1 Resource Split for Fixed Constellations

We can optimize the MSE of CSIT for a fixed constellation, i.e. for a fixed outage rate b . In this case, the outage rate based optimization setup, built

in the previous section, remains operational except that b is no more an optimization variable but a fixed parameter corresponding to the chosen constellation. Thus b will assume the values of 2 and 4 for QPSK and 16-QAM, respectively, although any other constellation can be chosen. The minimization of the MSE of CSIT will give the optimal resource split tailored for the particular constellation chosen. Hence the objective function for a fixed constellation (fixed value of b) becomes:

$$\min_{T_a} \left[2^{\frac{-b(T_{fb}-T_a)}{M-1}} + \epsilon(T_a, b) \right] \quad (5.42)$$

where $T_{fb} = T_a + T_q$ and b are fixed, and b and $\epsilon(T_a, b)$ are related as in Theorem 3. The constraint for this minimization is:

$$1 \leq T_a \leq T_{fb} \quad (5.43)$$

This minimization gives the optimal value of training length T_a which should be used to get the minimum MSE of CSIT for this particular constellation (fixed b) under fixed values of M , P_u and T_{fb} . This restriction of fixed constellation brings in some limitations. For example, the use of smaller constellation like QSPK at very high SNR will not be beneficial as CSIT error will stay bounded due to the fixed cardinality of the codebook (hence quantization error will be non-diminishing as a function of SNR) even for asymptotically large values of SNR.

5.7.2 Resource Split with Continuous Rates & Parity Bits

The other way to resolve the issue of discrete practical constellations is through the use of channel coding. This allows us to use positive real values for b , obtained from the original optimization setup. The only restriction, we impose, is that B should take an integer value which can be obtained by using ceiling or floor operation on the product bT_q . Now this B governs the cardinality of the codebook. The actual constellation, which is used to send feedback, is the one larger than that dictated by b , among the available constellations. Let the rate of that constellation be denoted by b_c . Hence the number of total bits, which will be sent in the feedback phase, is $B_c = b_c T_q$ where $B_c > B$ as $b_c > b$. All the extra bits $B_c - B$ in the feedback phase are used as parity bits. So one can employ either linear block codes or convolutional codes with an appropriate rate so as to convert B information (true channel feedback) bits into B_c coded bits. One advantage of using convolutional codes is that puncturing can give more flexibility for

rate matching. Now these B_c bits are sent in the digital feedback phase. As the outage rate b is less than the rate b_c of the constellation chosen, the use of larger constellation will give rise to increase in the number of erroneous coded bits. The number of errors will grow large in direct proportion to the difference $B_c - B$. On the other hand, all the extra feedback bits $B_c - B$ are the parity bits and when decoding will be performed at the BS, the capability of this coding/decoding operation to combat the channel errors (introduced in the quantized feedback) is also proportional to this difference, hence compensating the negative impact of using larger constellation.

5.8 Simulation Results

Our simulation environment consists of a BS with $M = 4$ antennas and a single user with a single antenna. The channel model is the same as described in Section 5.2. The feedback interval T_{fb} is fixed to 20 channel uses for all simulations.

5.8.1 Continuous Constellations

First we present the results when the outage rate b is not constrained to be an integer and can assume any positive real value. The optimization of the objective function, given in section 5.6, gives us the values for the optimal training length T_a and the optimal outage rate b for various values of user's power constraint, which is equal to the UL SNR as the noise at every BS antenna has been normalized to have unit variance. Knowing the values of $\epsilon(T_a, b)$ and T_q , computed based upon the optimal values of T_a and b , allows us to compute the upper bound of the final CSIT error eq. (5.31). These values have been plotted in dB scale in Fig. 5.2. For comparison purpose, we have also plotted the MSE of CSIT with classical training based estimation. This plot clearly shows the interest for our hybrid two-staged CSIT acquisition strategy as, from medium to large SNR values, CSIT error incurred by this scheme is much less than the error obtained by training based only CSIT acquisition. Only at very low SNR values, this two stage scheme performs worse than the classical training scheme. This happens because we have restricted our final estimate to come from the digital feedback. Here the total feedback resource (SNR and T_{fb}) does not allow transmission of sufficient number of bits through the channel so quantization error is quite large. This gets aggravated due to the poor training based estimate based upon which these bits are decoded, further degrading the performance. This degradation can be avoided by selecting

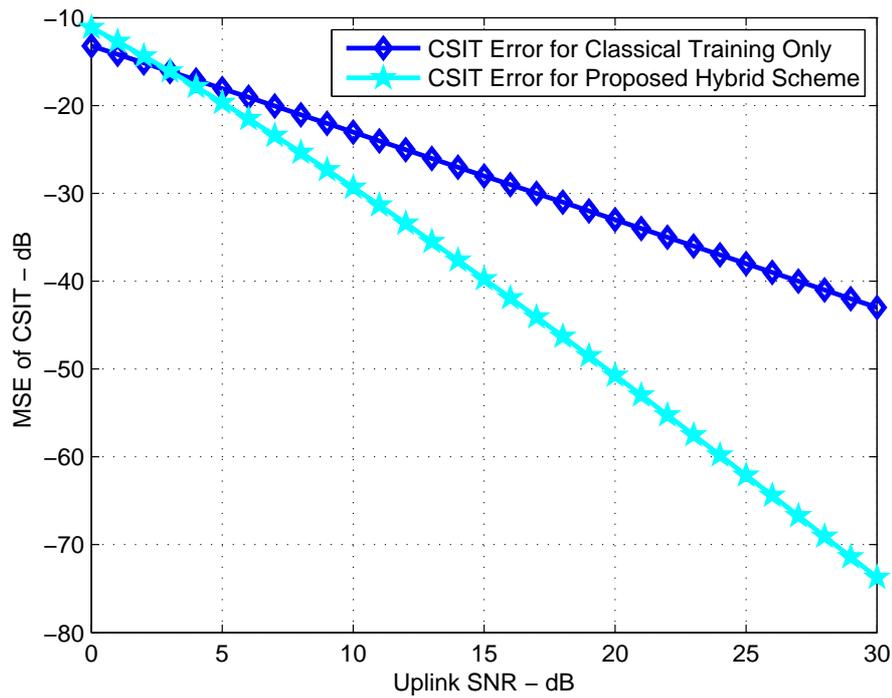


Figure 5.2: Mean-Square CSIT Errors: $T_{fb} = 20$ and $M = 4$. The novel hybrid scheme performs much better than the classical training based CSIT acquisition. Gains are significant even with naive use of practical constellations without any coding.

an SNR threshold below which traditional training based scheme should be employed.

To see the optimal split between training and quantized feedback, we have plotted the optimal values of training length T_a , corresponding values of quantized feedback interval T_q and the optimal outage rate b in Fig. 5.3.

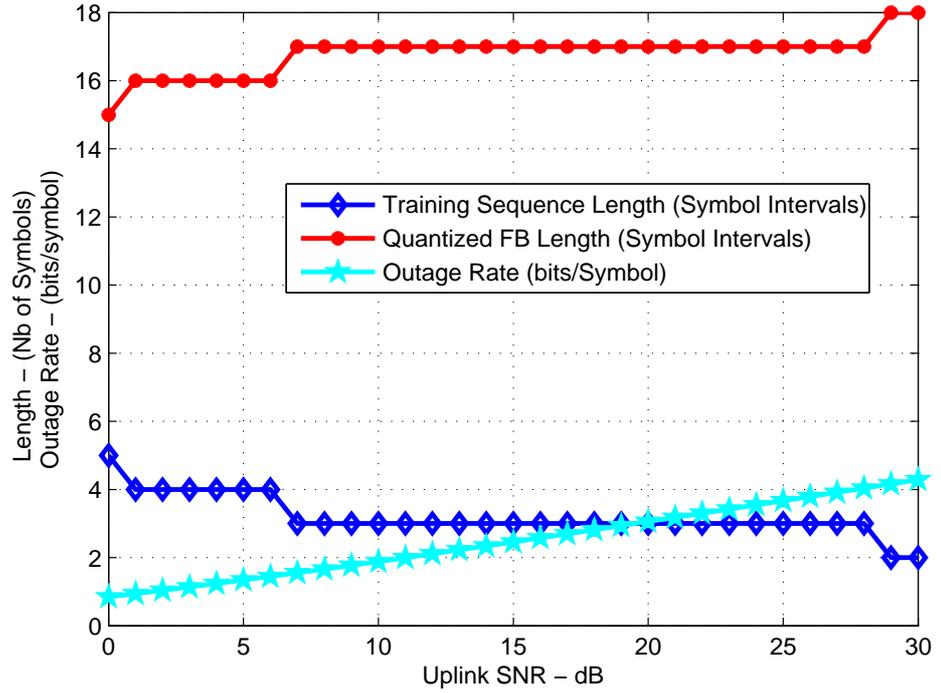


Figure 5.3: Optimal Lengths and Outage Rate: $T_{fb} = 20$ and $M = 4$. With increase in SNR, both the length of the quantized feedback interval T_q and the outage rate b increase gradually.

5.8.2 Fixed Discrete Constellations

In this section, we present the simulation results when the fixed constellations like QPSK and 16-QAM are used for quantized feedback transmission. So the outage rate b becomes fixed corresponding to the fixed constellation (2 for QPSK and 4 for 16-QAM) and the optimization is carried only over the

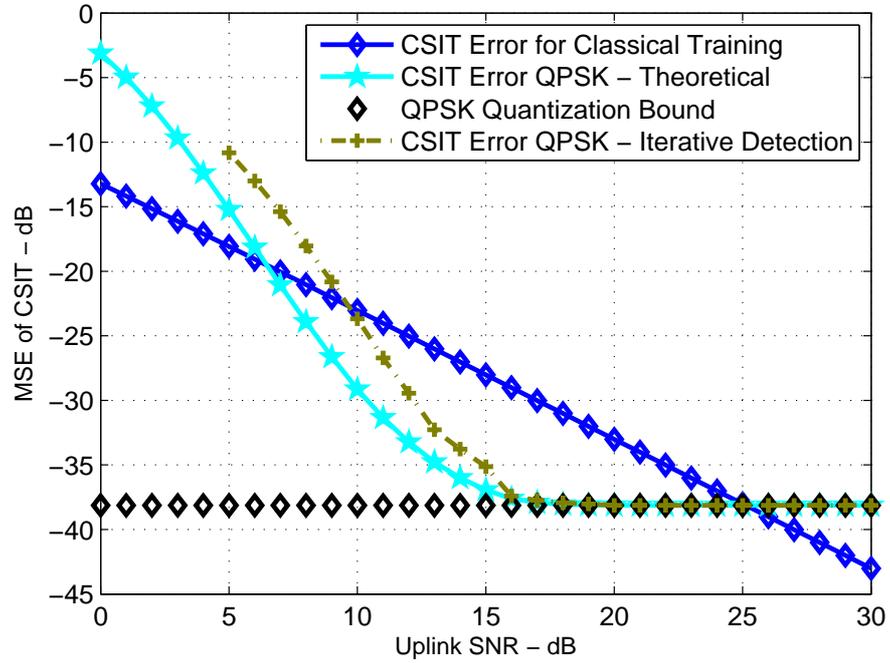
resource split between training and quantized feedback as described in section 5.7.1. The curves for the MSE of CSIT obtained theoretically, by doing the simulations with actual constellations and the corresponding quantization bound for that constellation have been plotted in Fig. 5.4. Quantization bound gives the quantization error when maximal $(T_{fb} - 1)$ symbols are used for quantized feedback part. Hence, it gives the lower bound on the MSE of CSIT (performance upper bound) for that particular constellation. For comparison purpose, we have also plotted the MSE of CSIT for classical training scheme. This figure shows that from low to medium SNR values, the novel scheme with QPSK gives CSIT error below that of the classical training approach but 16-QAM is not attractive in this range due to many incorrect detection events. At high SNR values, hybrid scheme with QPSK suffers from performance degradation due to its bounded quantization error but 16-QAM behaves much better than the classical scheme. At very high values of SNR, even the 16-QAM will show bounded performance for the same reason that its rate does not increase with SNR but then one needs to switch to further larger constellations.

In Fig. 5.4, both for QPSK and 16-QAM, we have plotted the MSE of CSIT using our proposed iterative estimation and detection algorithms from section 5.5. A surprising fact about the two proposed iterative algorithms is their similar performance. One would expect the iterative estimation and detection algorithm (with ML detection) to perform much better than the simplified iterative estimation and detection algorithm (which uses the simple LS detection), but extensive simulations show that the performance difference between the two algorithms is negligible. In all our simulations, both algorithms show very rapid convergence and they were always converging in second or third iteration. There were extremely rare instances (less than one in ten million) when convergence was not achieved in three iterations.

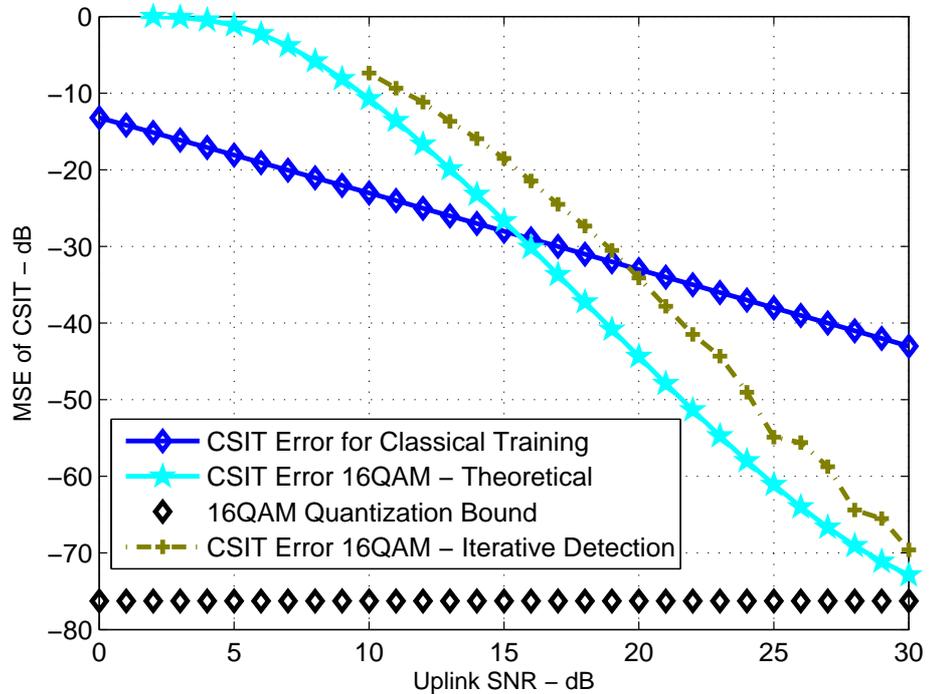
We don't plot the optimal training and quantized feedback interval lengths but they show the same behavior as displayed in Fig. 5.3, i.e., the optimal quantized feedback interval gets larger with the increase in SNR for both constellations.

5.8.3 Discrete Constellations and Coding

Now we plot the results of the MSE of CSIT when quantized feedback is sent using discrete constellations and the rate matching is performed using convolutional codes as explained in section 5.7.2. The code rates and the puncturing patterns need to be selected carefully. First of all, convolutional



(a)



(b)

Figure 5.4: Mean-Square CSIT Errors: $T_{fb} = 20$ and $M = 4$ (a) QPSK and (b) 16-QAM. The novel hybrid scheme with QPSK performs better than the classical one from 9 to 25 dB of SNR, but 16-QAM outperforms both after 21 dB.

codes of all desired rates are not available. Secondly, although puncturing can help a lot to reach to the desired rate still it needs to be selected carefully as random choice of puncturing pattern may destroy the code structure and hence ultimately its performance.

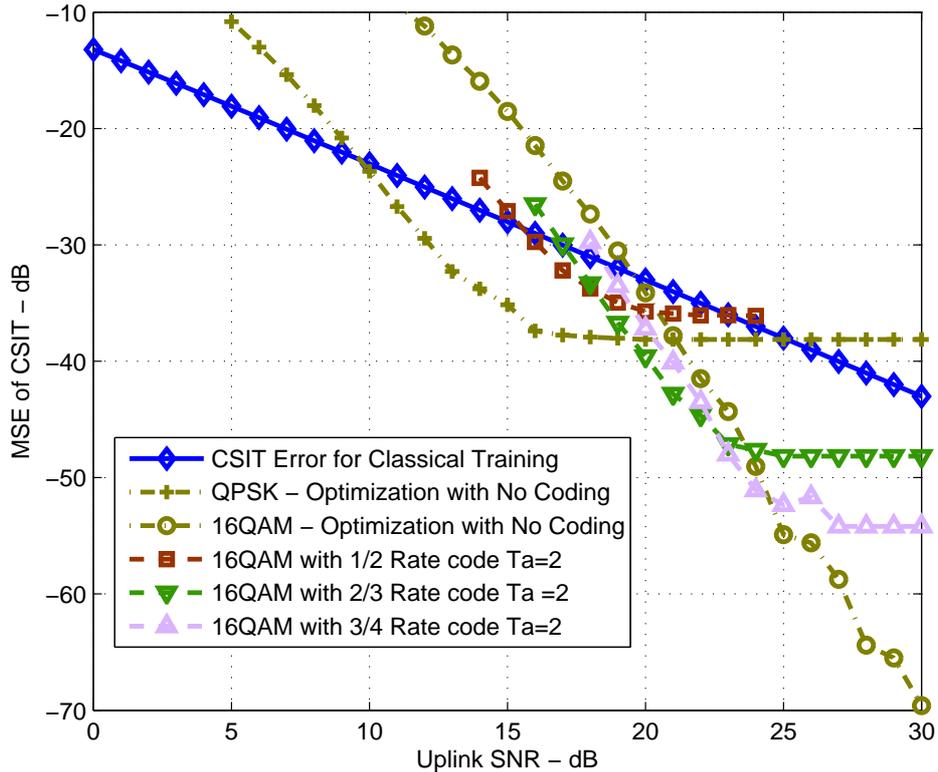


Figure 5.5: Mean-Square CSIT Errors with Convolutional Coding: $T_{fb} = 20$ and $M = 4$. At certain SNR intervals, coding strategy performs better than no coding optimal resource split outcome.

We plot the results obtained using three different codes (1/2 rate code, 2/3 rate code and 3/4 rate code) in Fig. 5.5. All of these codes have been used with 16-QAM (4 bits per channel use). Hence the number of actual information (feedback) bits are 2, 2.67 and 3 per channel use for 1/2, 2/3 and 3/4 rate code respectively. For comparison purpose, the plot shows the

MSE of CSIT obtained by using QPSK and 16-QAM constellations without any coding and through classical training scheme.

For 1/2 rate code, the generator matrix is $[171\ 133]_8$ and trace back length is 30. It performs better than classical training from 16 to 23 dB of SNR but QPSK without any coding performs better than this curve. For 2/3 rate code, the generator matrix is $[4\ 5\ 17; 7\ 4\ 2]_8$ with trace back length of 20. From 17 dB onward, it performs better than classical training. It performs even better than 16-QAM (without coding) before 24 dB of SNR. For 3/4 rate code, we use the 1/2 rate base code (same as before) and use the puncturing pattern of $[111001]$ to get the final rate of 3/4.

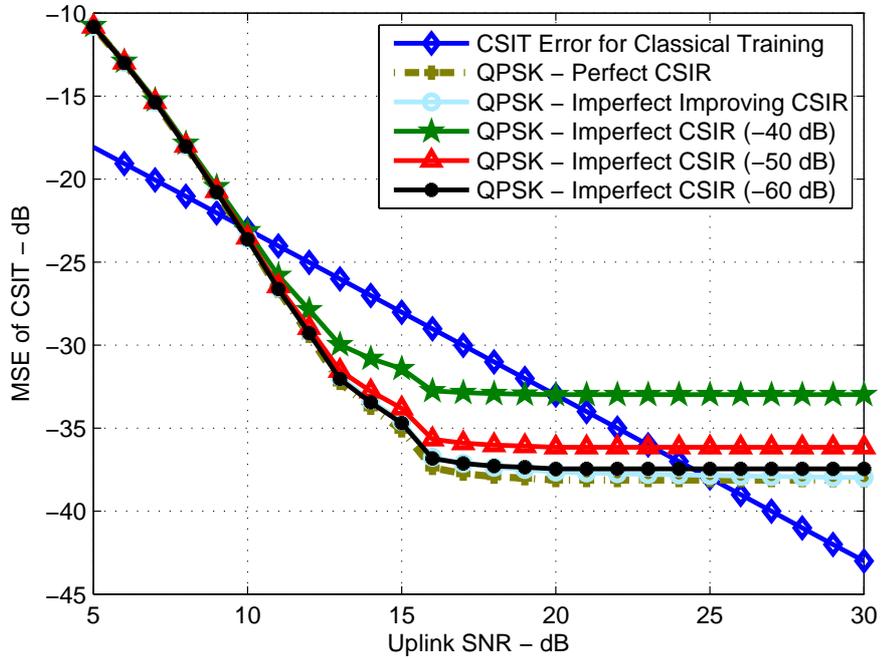
5.8.4 Imperfect CSIR Analysis

All the previous results have been obtained working under the assumption of perfect CSIR which is certainly too good to be true. Here we remove this perfect CSIR assumption and analyze how the MSE of CSIT with novel scheme behaves with imperfect CSIR.

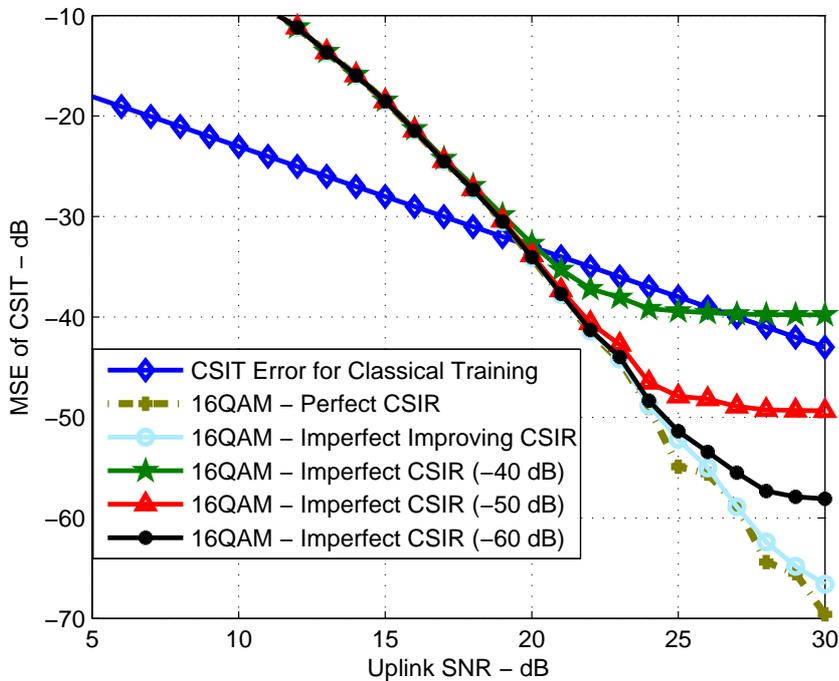
The curves, when quantized feedback is transmitted using QPSK and 16-QAM, have been plotted in Fig. 5.6. We have plotted these curves under two scenarios. First, when the CSIR quality varies and improves with the increase in UL SNR which is quite logical as, due to reciprocity, the link quality improves in both directions and the BS can surely pump more power as compared to a small hand-held mobile unit. For this case, we take the MSE of CSIR 30 dB less than the classical training only CSIT curve. The second scenario is when CSIR quality is held fixed independent of the UL SNR. For this, we plot the MSE of CSIT when the MSE of CSIR is kept fixed at -40 , -50 and -60 dB. We believe this scenario to be of relatively less importance. We remark that when CSIR quality improves with UL SNR, hybrid approach performs very close to the perfect CSIR curve. For the other case when CSIR quality is kept fixed, it may become the performance limit of the MSE of CSIT (if not of proper quality).

5.9 Conclusions

Traditional CSIT acquisition in reciprocal systems relying exclusively on the use of training sequences ignores the shared knowledge of an identical channel between the BS and the user. We presented a novel approach of CSIT acquisition at the BS for the DL transmission in a reciprocal MIMO communication system combining the use of a training sequence together



(a)



(b)

Figure 5.6: Mean-Square CSIT Errors with Imperfect CSIR: $T_{fb} = 20$ and $M = 4$ (a) QPSK and (b) 16-QAM. For an imperfect CSIR of reasonable quality, the novel scheme performs much better than the classical scheme and the performance approaches to the perfect CSIR case for a good enough CSIR.

with quantized channel feedback. We characterized the optimal CSIT acquisition setup and proposed two iterative algorithms for the resulting joint estimation and detection problem and provided a convergence proof. The novel outage-rate based approach allows the optimal resource partitioning between the training and the quantized feedback. We proposed two strategies to overcome the limitation of practical constellation availability with integer number of bits per channel use either by optimizing the resource split for a particular constellation or by the use of channel coding for rate matching. The novel combining scheme shows superior performance due to better exploitation of the reciprocity principle and the trade-off between the CSIT quality and the resource utilization improves significantly. It is further shown that with an imperfect CSIR of reasonable quality, performance gains comparable to the perfect CSIR case are achievable.

Part III

Transmit Power Minimization with User Selection

Chapter 6

Transmit Power Minimization with User Selection

6.1 Introduction

6.1.1 Motivation

In all the previous chapters, the objective has been either to determine the fundamental rate limits (both for SU and MU systems) or to devise new transmission/feedback schemes to achieve higher rates under fixed power constraints. In this chapter, we change the perspective completely and study a problem which is dual to the sum rate maximization under fixed power constraint, namely, transmit power minimization under fixed rate constraints. We consider a system with non-elastic traffic. The users in the system have certain SINR constraints which can be directly translated to rate or quality constraints. Thus if a user is selected for transmission in a particular slot (frame), it should achieve SINR at least equal to its specified threshold. The transmit power minimization corresponding to achieving a certain set of QoS is a very important metric for service providers as it is equivalent to the cost of the service which they would surely like to minimize without compromising the quality.

6.1.2 The State of the Art

The problem of the minimization of downlink transmit power required to meet users' SINR constraints by joint optimization of transmit beamforming (BF) and power allocation was solved in [33] and [34]. They showed the interesting duality of UL and DL channels for this problem. Exploiting this UL-DL duality, they gave iterative algorithms to find the optimal beamforming matrices and the optimal power assignments to users and showed the convergence of these algorithms to the optimal solution. For Gaussian multi-user channels (either UL or DL), they showed that the problem of the minimization of transmit power corresponding to certain SINR targets bears a relatively simple solution due to the added structure which may be exploited by successive interference cancellation (SIC) in the UL and by DPC based encoding for known interference in the DL channels and the results were presented in [35], [36] and [34]. The optimal BF turns out to be the MMSE solution where no interference arises from the already encoded users and treating the interference of unencoded users as extra noise, and power allocation is done to raise the SINR level to the target SINR. Actually the DL problem is solved by first solving the dual UL problem due to its relatively simple structure.

The performance of different user selection algorithms for transmit power minimization was studied in [37]. The Gaussian multi-user systems were analyzed without exploiting the extra system structure through SIC or DPC. For the case of 2 users transmitted simultaneously, analytical expressions were obtained for minimum average transmit power required for guaranteed rates with norm-based user selection (NUS) and angle-based user selection (AUS).

6.1.3 Contribution

We study the problem of average transmit power minimization to meet users' SINR constraints in conjunction with user scheduling. In this Gaussian multi-user system, we make use of SIC in the UL channel or DPC based encoding in the DL channel. As the channel information is already required at the BS for beamforming and power allocation assignments, this extra processing does not require any extra information. This problem formulation gives twofold advantage over [37]: first no iterations are required to compute the optimal BF vectors and power allocation scalars, and second less average power is required at the transmitter to satisfy the same SINR constraints at the users' side. For the case of two users transmitted simultaneously,

we derive analytical expressions for the minimum average transmit power required with semi-orthogonal user selection (SUS), NUS and AUS. To find the minimum average transmit power to achieve certain SINR constraints when users are selected through SUS is one of the novelties of this work. We compare the performance of these user selection algorithms in terms of minimum average transmit power required to satisfy users' SINR constraints. It turns out that NUS and AUS are strictly sub-optimal when compared with SUS.

6.1.4 Organization

This chapter is organized as follows. Section 6.2 describes the system model. Section 6.3 gives a brief overview of the problem of transmit power minimization without user selection. In section 6.4, certain user selection algorithms are reviewed for which later we analyze the performance. The main results of the chapter, the analytical expressions for the minimum average transmit power when two users are simultaneously transmitted, are presented in section 6.5 along with performance comparison. The concluding remarks appear in section 6.6.

6.2 System Model

The system, we consider, consists of a BS having M transmit antennas and K single-antenna user terminals. In the DL, the signal received by k -th user can be expressed as

$$y_k = \mathbf{h}_k^\dagger \mathbf{x} + z_k, \quad k = 1, 2, \dots, K \quad (6.1)$$

where $\mathbf{h}_1^\dagger, \mathbf{h}_2^\dagger, \dots, \mathbf{h}_K^\dagger$ are the channel vectors of users 1 through user K with $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$, $\mathbf{x} \in \mathbb{C}^{M \times 1}$ denotes the signal transmitted by the BS and z_1, z_2, \dots, z_K are independent complex Gaussian additive noise terms with zero mean and variance σ^2 . We denote the concatenation of the channels by $\mathbf{H}_F^\dagger = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_K]$, so \mathbf{H}_F is the $K \times M$ forward channel matrix with k -th row equal to the channel of k -th user (\mathbf{h}_k^\dagger). The channel is assumed to be block fading having coherence length of T symbol intervals. The entries of the forward channel matrix \mathbf{H}_F are i.i.d. complex Gaussian with zero mean and unit variance. We make the simplifying assumption of the presence of perfect CSIT so as to focus completely on the performance of different user selection algorithms.

We suppose that each user has the same SINR constraint of γ . If K_s out of K users are selected for transmission during each coherence interval, the

channel input \mathbf{x} can be written as $\mathbf{x} = \overline{\mathbf{V}}\mathbf{P}^{1/2}\mathbf{u}$, where $\overline{\mathbf{V}} \in \mathbb{C}^{M \times K_s}$ denotes the beamforming matrix with normalized columns, \mathbf{P} is $K_s \times K_s$ diagonal power allocation matrix with positive real entries and $\mathbf{u} \in \mathbb{C}^{K_s \times 1}$ is the vector of zero-mean unit-variance Gaussian information symbols. Hence, $\mathbb{E}[\text{Tr}(\mathbf{P})]$ is the average transmit power which can be minimized by optimizing over the beamforming matrix $\overline{\mathbf{V}}$ and the power allocation matrix \mathbf{P} to achieve the SINR target γ for all selected users. We select this minimum average transmit power as the performance metric and study the performance of various user selection algorithms when users' SINR constraints (γ) have to be satisfied.

6.3 Overview of Transmit Power Minimization Problem

The signal received by k -th user can be written as

$$\begin{aligned} y_k &= \mathbf{h}_k^\dagger \overline{\mathbf{V}}\mathbf{P}^{1/2}\mathbf{u} + z_k, \quad k = 1, 2, \dots, K_s \\ &= \sqrt{p_k} \mathbf{h}_k^\dagger \overline{\mathbf{v}}_k u_k + \sum_{\substack{j=1 \\ j \neq k}}^{K_s} \sqrt{p_j} \mathbf{h}_k^\dagger \overline{\mathbf{v}}_j u_j + z_k, \end{aligned} \quad (6.2)$$

where p_k represents the power allocated to the stream of k -th user. The second term in the expression represents the interference contribution at k -th user due to beams meant for other selected users. Based upon this received signal, the SINR of k -th user can be written as

$$\text{SINR}_k = \frac{p_k |\mathbf{h}_k^\dagger \overline{\mathbf{v}}_k|^2}{\sum_{\substack{j=1 \\ j \neq k}}^{K_s} p_j |\mathbf{h}_k^\dagger \overline{\mathbf{v}}_j|^2 + \sigma^2}. \quad (6.3)$$

Without user selection, the problem of optimization of beamforming vectors and power allocation matrix was solved in [34] and [33] using the UL-DL duality (see Section 4.3 and 5.2 in [34] for details). They gave iterative algorithms to obtain the optimal beamforming vectors and the optimal power allocation for each user. The optimal beamforming vectors corresponding to a particular (sub-optimal) power allocation are obtained, then power allocations are updated corresponding to these beamforming vectors. This process is repeated till both converge to their optimal values. Unfortunately

general closed form expressions for the transmit power required to achieve SINR targets don't exist due to intricate inter-dependence of beamforming vectors and power allocations, as is evident from eq. (6.3).

For Gaussian multi-user systems (the case of interest), the extra structure allows the use of SIC in UL or DPC based encoding in the DL. This permits to obtain the optimal beamforming vectors and power assignments using back substitution without any iteration. Although iterations are not required in this scenario, yet beamforming vector and power allocation of one user depend upon the BF vectors and power assignments of already treated users, hence closed form results are possible only when two users are transmitted simultaneously. If both of the users have the same SINR target γ (for relatively large γ), the minimum instantaneous transmit power required is given by the following expression taken from Section 5.2 of [34].

$$p_{\text{tx}}(\mathbf{h}_1, \mathbf{h}_2) = \sigma^2 \gamma \left(\frac{1}{\|\mathbf{h}_1\|^2} + \frac{1}{\|\mathbf{h}_2\|^2 \sin^2(\theta_{12})} \right), \quad (6.4)$$

where θ_{12} is the angle between the channel vectors of the two users.

6.4 Review of User Selection Algorithms

In this section, we briefly state how different user selection algorithms operate.

6.4.1 Norm-Based User Selection (NUS)

In NUS, users are selected based only upon their channel strengths. So K users are sorted in descending order of their channel norm values, and first K_s users are selected for transmission.

6.4.2 Angle-Based User Selection (AUS)

The user selection criterion in AUS is the mutual orthogonality of their channel vectors. The first user is selected which has the largest channel norm. The second user is selected as the one which is the most orthogonal to this user, without any regard to its channel strength. The third selected user is the one whose channel vector is the most orthogonal to the subspace formed by the two already selected users' channels. This process is repeated till K_s users have been selected.

6.4.3 Semi-Orthogonal User Selection (SUS)

The user selection metric for SUS is the combination of the channel strength and its spatial orthogonality with respect to the other users. The first chosen user is the one with the largest channel norm. The second chosen user is the one whose projection on the null space of the first user has the largest norm. Similarly, the third chosen user will be the one whose projection on the null space of the subspace formed by the channel vectors of the first two users has the largest norm. This process is repeated till K_s users get selected. Interested readers can find the details of this algorithm in [38] or [21].

6.5 Transmit Power with User Selection

In this section, first we give the main results of this chapter, some analytical expressions for the minimum average transmit power required to achieve certain SINR targets at users' side when these users have been selected following different user selection algorithms. Later, we compare the performance of these user selection algorithms.

6.5.1 Main Results

We select users obeying different user selection algorithms as detailed in Section 6.4 and in the second step, we compute the optimal beamforming vectors and power assignments following the steps outlined in Section 6.3.

Theorem 4 (Minimum Average Transmit Power). *Consider a DL system having a BS equipped with M transmit antennas and K single antenna users, each having an SINR constraint of γ , and $K_s = 2$ users are selected for simultaneous transmission in each coherence block. If user selection is done through NUS, the minimum average transmit power is given by*

$$p_N = \gamma\sigma^2 \left(K\alpha_{M,K-1} - \left(K - 2 - \frac{1}{M-2} \right) \alpha_{M,K} \right). \quad (6.5)$$

If user selection is performed using AUS, the minimum average transmit power required is given by

$$p_A = \gamma\sigma^2 \left(\frac{1}{K-1} \left(\frac{K}{M-1} - \alpha_{M,K} \right) + \frac{(M-1)(K-1)\alpha_{M,K}}{(M-1)(K-1)-1} \right). \quad (6.6)$$

Now if SUS is employed at the BS to select the users, the upper bound of the average transmitted power required (the performance lower bound) is given

by

$$p_S = \gamma\sigma^2 (\alpha_{M,K} + K\alpha_{M-1,K-1} - (K-1)\alpha_{M-1,K}), \quad (6.7)$$

where $\alpha_{M,K}$ is a constant solely governed by M and K and is given by

$$\alpha_{M,K} = \int_0^\infty K \frac{e^{-x}x^{M-2}}{\Gamma(M)} G(M,x)^{K-1} dx, \quad (6.8)$$

where $G(M,x)$ is the regularized Gamma function [68].

For the case of two selected users, a very useful lower bound on the average transmit power required to achieve SINR targets (performance upper bound) can be easily derived and is given by

$$p_L = \gamma\sigma^2 \left(K\alpha_{M,K-1} - (K-1) \left(1 - \frac{M-1}{(M-1)(K-1)-1} \right) \alpha_{M,K} \right). \quad (6.9)$$

Proof. The proof outline is given in Appendix 6.B using the distribution results from Appendix 6.A. □

6.5.2 Performance Comparison

In this subsection, we compare the performance of already described user selection algorithms when the metric of interest is the average transmit power required to satisfy users' SINR constraints.

The case of $K_s = 2$ Selected Users

The plot of minimum average transmit power required to attain specific SINR targets γ versus the number of antennas at the BS appears in Fig. 6.1 for the considered user selection algorithms. We remark that SUS performs better than the other user selection schemes but with the increase in the number of transmit antennas, NUS also performs very well. The similar behavior was observed in [37] and the reason comes from the fact that with the increase in the number of transmit antennas, users' channels start becoming (close to) spatially orthogonal and furthermore, due to difference $(M - K_s)$ very good beamforming vectors can be chosen to cause very small interference for other users.

Fig. 6.2 plots the curves of the minimum average transmit power versus the number of users for a fixed number of transmit antennas. SUS again performs very close to the optimal (obtained by exhaustive search) but we

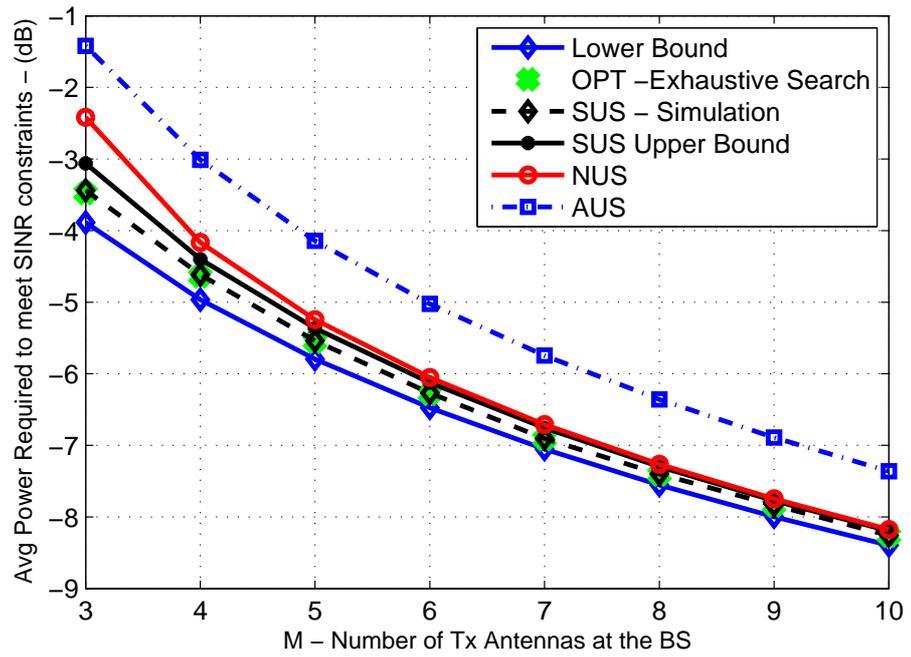


Figure 6.1: Min. Avg. Transmit Power vs. M for $K = 10$, $K_s = 2$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS is the best strategy and follows closely the power lower bound. NUS also performs close to SUS with increasing number of transmit antennas.

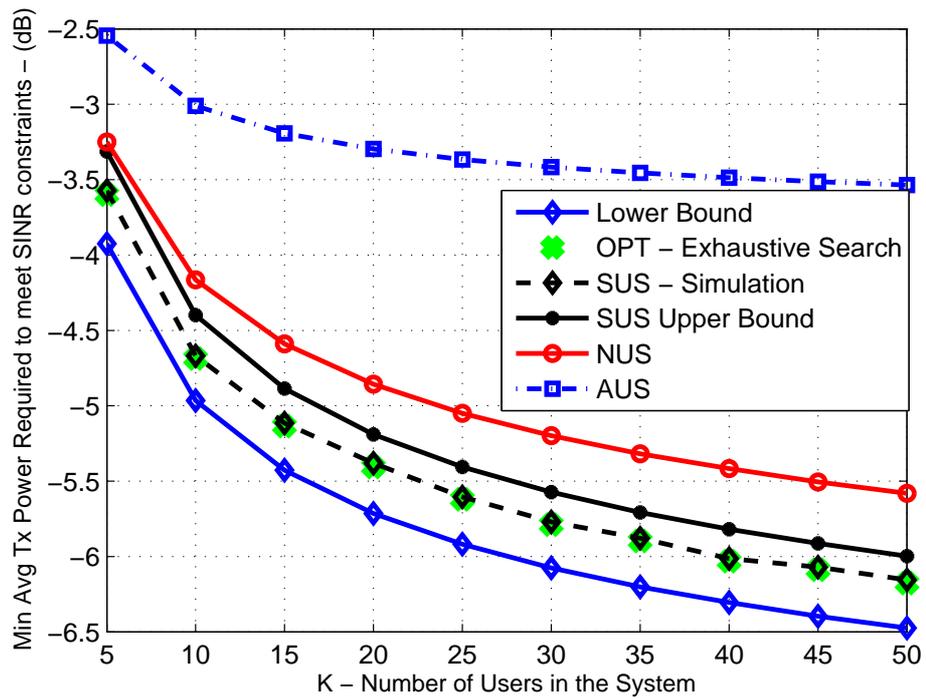


Figure 6.2: Min. Avg. Transmit Power vs. Nb. of Users for $M = 4$, $K_s = 2$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS performs the best and NUS becomes sub-optimal when number of users increases.

remark that NUS does not behave very well in this scenario because it just chooses users with good channel norms without paying any attention to their spatial orthogonality which may affect significantly the interference observed by the selected users.

The Case of $K_s > 2$ Selected Users

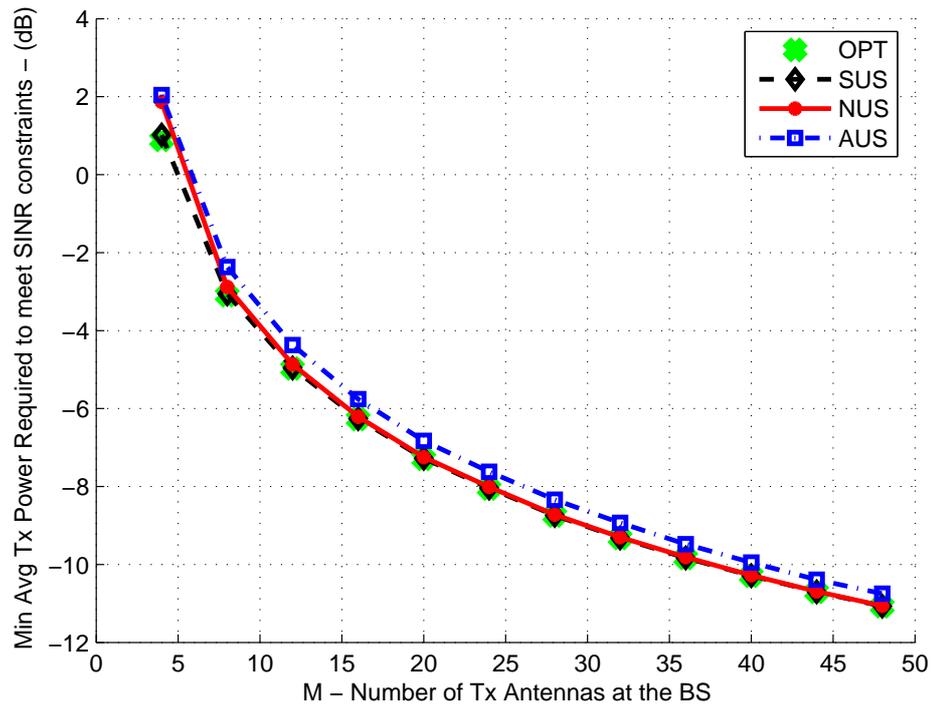


Figure 6.3: Min. Avg. Transmit Power vs. M for $K = 8$, $K_s = 4$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS is the best strategy and follows closely the power lower bound. NUS also becomes optimal for a reasonably large number of transmit antennas.

We have plotted the minimum average transmit power required to achieve certain SINR targets versus the number of transmit antennas and versus the number of system users in Fig. 6.3 and Fig. 6.4 respectively, for the user selection algorithms of interest. For both of these plots, the number of selected users is 4. We observe the same behavior as observed in the case

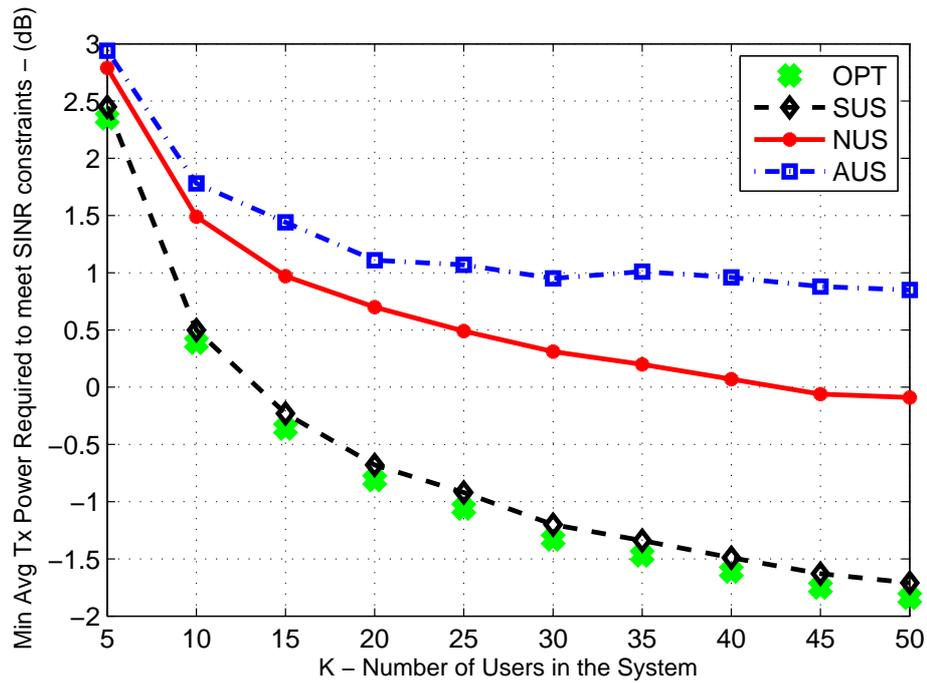


Figure 6.4: Min. Avg. Transmit Power vs. Nb. of Users for $M = 4$, $K_s = 4$, $\gamma = 10$ dB, $\sigma^2 = 0.1$. Curves show that SUS performs the best and NUS becomes strictly sub-optimal when number of users increases.

of 2 selected users. For large number of transmit antennas, both SUS and NUS perform very close to the optimal, even AUS achieves a reasonable performance.

On the other hand, for a fixed number of transmit antennas at the BS when the number of users present in the system increases, the performance of NUS degrades substantially. The reason is that NUS captures the raw aspect of multi-user diversity which governs only the self signal power but pays no attention to the inter-user spatial separation which might have a larger impact on the interference power. The worst performance of AUS is expected as it pays no attention to the strength of the selected users which is quite important for power minimization objective. Moreover, SUS performs very close to the optimal, for any set of system parameters. The reason is the selection criterion of SUS where both the channel strength and the spatial orthogonality of the users are properly taken care of.

6.6 Conclusions

In this chapter, we have studied the performance of various user selection algorithms in terms of the minimization of average transmit power required to satisfy specific SINR targets at users' side. Closed form expressions of the average transmit power for the three user selection algorithms, namely SUS, NUS and AUS, were derived when only two users are selected for simultaneous transmission. SUS, which has been shown to behave close to optimal for the sum rate maximization objective under fixed power constraint, shows very attractive performance in this problem setting of transmit power minimization to achieve hard SINR targets and gives better results than NUS and AUS algorithms. For a fixed number of users and increasing number of transmit antennas, NUS performs very close to SUS. In the complementary setting of fixed number of BS transmit antennas and an increasing number of system users, NUS shows substantial performance degradation but SUS still performs very close to the optimal.

6.A Some Useful Distributions

In this appendix, we give some useful cumulative distribution functions (CDF) for which probability density functions (PDF) can be computed by simple differentiation. Most of these are known relations, others have been computed using the tools from order statistics [69] and were also given in [37].

As users' channels are spatially i.i.d. Gaussian, the CDF of $\|\mathbf{h}_i\|^2$ for any i is χ^2 distributed with $2M$ degrees of freedom whose CDF is

$$F_i(M; x) = G(M, x), \quad (6.10)$$

where G denotes the regularized Gamma function [68], and is defined as

$$G(M, x) = \frac{1}{\Gamma(M)} \int_0^x e^{-t} t^{M-1} dt. \quad (6.11)$$

The PDF corresponding to the CDF $F_i(M; x)$ is given by

$$f_i(M; x) = \frac{e^{-x} x^{M-1}}{\Gamma(M)}. \quad (6.12)$$

Below we give the CDFs for the largest and the second largest order statistics. The CDF of the user having the largest channel norm among K i.i.d. users distributed as $F_i(M; x)$ is

$$F_1(M, K; x) = F_i(M; x)^K. \quad (6.13)$$

The CDF of the user having the second largest channel norm among K i.i.d. users is

$$F_2(M, K; x) = K F_i(M; x)^{K-1} - (K-1) F_i(M; x)^K. \quad (6.14)$$

Similarly from [37], the distribution of any random user which does not have the largest norm can be specified as

$$F_i(M, K; x) = \frac{K}{K-1} F_i(M; x) - \frac{1}{K-1} F_i(M; x)^K. \quad (6.15)$$

The CDF of the angle between the channel vectors of any two randomly distributed M -dimensional channel vectors is given by

$$F_{\theta_i}(M; x) = [\sin(x)]^{2M-2}. \quad (6.16)$$

This can be computed based upon the fact that the squared cosine of the angle between two random vectors is β distributed. Similarly the distribution of the largest angle between one user and any other user in a system of

K users is the maximum of $K - 1$ i.i.d. angles, each of whom is individually distributed as $F_{\theta_i}(M; x)$, hence the CDF for this largest angle statistic is given by

$$F_{\theta_1}(M, K; x) = F_{\theta_i}(M; x)^{K-1} = [\sin(x)]^{2(M-1)(K-1)}. \quad (6.17)$$

6.B Proof of Theorem 4

In the proof of this theorem, we make extensive use of some useful CDFs which have been grouped together in appendix 6.B so we encourage the readers to go through the previous appendix for better understanding.

The instantaneous transmit power required to meet SINR constraint of γ for the two selected users, by using optimal beamforming vectors and power allocation assignments, given in eq. (6.4), is

$$p_{\text{tx}}(\mathbf{h}_1, \mathbf{h}_2) = \sigma^2 \gamma \left(\frac{1}{\|\mathbf{h}_1\|^2} + \frac{1}{\|\mathbf{h}_2\|^2 \sin^2(\theta_{12})} \right). \quad (6.18)$$

θ_{12} represents the angle between the channel vectors of two selected users. The objective is to compute the average transmit power when these two users have been selected using various user selection algorithms, namely NUS, AUS and SUS.

6.B.1 Norm-Based User Selection

For NUS, the users are chosen as described in section 6.4. Hence the squared norm of the first selected user is distributed as $F_1(M, K; x)$ and that of the second user as $F_2(M, K; x)$. As these users are selected based only upon their channel norms, the angle between their channel vectors θ_{12} is distributed as the angle between any two random vectors and as a result has the CDF of $F_{\theta_i}(M; x)$.

For the case of two users, we have two ways to perform SIC (considering UL) or DPC based encoding (considering DL). It's known that for the objective of the minimization of transmit power, the weaker user should be the one which gets decoded without interference [3]. For this optimal ordering, \mathbf{h}_1 should be the channel of the weaker user (2nd strongest, distributed as $F_2(M, K; x)$ and gets decoded with no interference) and \mathbf{h}_2 should be the channel of the strongest user distributed as $F_1(M, K; x)$ facing some

interference. Hence average transmit power is given by

$$p_N = \sigma^2 \gamma \left(\mathbb{E}_{F_2(M, K; x)} \frac{1}{\|\mathbf{h}_1\|^2} + \mathbb{E}_{F_1(M, K; x)} \frac{1}{\|\mathbf{h}_2\|^2} \mathbb{E}_{F_{\theta_i}(M; x)} \frac{1}{\sin^2(\theta_{12})} \right). \quad (6.19)$$

Computing all these expectations, we get the result.

6.B.2 Angle-Based User Selection

For AUS, the first selected user is the strongest user whose squared norm is distributed as $F_1(M, K; x)$ and the second user is the one making the largest angle with the first user. The squared norm of the second selected user is distributed as the squared norm of the any random user which is not the user with the largest norm and hence the CDF is $F_i(M, K; x)$. The angle between the channel vectors of two selected users assumes the distribution of the largest order angle statistic among K users and hence the CDF is given by $F_{\theta_1}(M, K; x)$. Now following the optimal ordering, \mathbf{h}_2 is the strongest user distributed as $F_1(M, K; x)$. The average transmit power for this user selection is given by

$$p_A = \sigma^2 \gamma \left(\mathbb{E}_{F_i(M, K; x)} \frac{1}{\|\mathbf{h}_1\|^2} + \mathbb{E}_{F_1(M, K; x)} \frac{1}{\|\mathbf{h}_2\|^2} \mathbb{E}_{F_{\theta_1}(M, K; x)} \frac{1}{\sin^2(\theta_{12})} \right). \quad (6.20)$$

This will give the result given in Theorem 4.

6.B.3 Semi-Orthogonal User Selection

In SUS, the first user is selected with the largest channel norm but the second selected user is the one whose projection on the null space of the first user has the largest norm. Let's assume that user 1 having channel \mathbf{h}_1 is the user with the largest norm, whose squared norm is distributed as $F_1(M, K; x)$. Now $\mathbf{h}_2 \sin(\theta_{12})$ is the projection of the vector \mathbf{h}_2 on the null space of \mathbf{h}_1 . To compute the expectation over this projected squared norm, we make use of [70, Lemma 3], which was also used in [38, Appendix III]. The term $\|\hat{\mathbf{h}}_2\|^2 \triangleq \|\mathbf{h}_2\|^2 \sin^2(\theta_{12})$ is the maximum of $K - 1$ channel norms orthogonalized w.r.t. \mathbf{h}_1 . Following [70], we can orthogonalize all the channel vectors w.r.t. an arbitrary vector so for each of them the squared norm is χ^2 distributed with $2(M - 1)$ degrees of freedom and each has the distribution which is given by $F_i(M - 1; x)$. Let us denote the projection of \mathbf{h}_i on the null space of that arbitrary vector by $\hat{\mathbf{h}}_i$ and same for others.

Now the second largest norm of these orthogonalized vectors will be

$$\|\hat{\mathbf{h}}_2\|^2 = 2^{\text{nd}} \max \|\hat{\mathbf{h}}_i\|^2, i = 1, \dots, K \quad (6.21)$$

whose distribution is given by $F_2(M-1, K; x)$. Lemma 3 in [70] shows that statistically $\|\hat{\mathbf{h}}_2\|^2$ is smaller than $\|\mathbf{h}_2\|^2$. Hence for SUS, an upper bound of the average transmit power can be computed by replacing \mathbf{h}_2 with $\hat{\mathbf{h}}_2$ as

$$p_S = \sigma^2 \gamma \left(\mathbb{E}_{F_1(M, K; x)} \frac{1}{\|\mathbf{h}_1\|^2} + \mathbb{E}_{F_2(M-1, K; x)} \frac{1}{\|\hat{\mathbf{h}}_2\|^2} \right). \quad (6.22)$$

Once the expectations computed in this expression, we get the SUS result of Theorem 4.

6.B.4 Performance Upper Bound

To compute a lower bound on the minimum average transmit power required to satisfy SINR targets of γ , the performance upper bound, we assume that the two selected users have the two largest norms as in NUS with CDFs as $F_1(M, K; x)$ and $F_2(M, K; x)$ and the angle between their channel vectors is the largest angle possible as in AUS, distributed as $F_{\theta_1}(M, K; x)$. Hence with optimal ordering, the lower bound on the average transmit power can be obtained by computing the expectations in the following expression

$$p_L = \sigma^2 \gamma \left(\mathbb{E}_{F_2(M, K; x)} \frac{1}{\|\mathbf{h}_1\|^2} + \mathbb{E}_{F_1(M, K; x)} \frac{1}{\|\mathbf{h}_2\|^2} \mathbb{E}_{F_{\theta_1}(M, K; x)} \frac{1}{\sin^2(\theta_{12})} \right). \quad (6.23)$$

Chapter 7

Conclusions and Future Perspectives

7.1 Conclusions

This thesis has mainly dealt with three research directions which become the three parts of the thesis. The first part focuses on the high SNR analysis of the communication channels and the determination of the DOF for SU and MU channels is the aim of this part as at extreme values of SNR, the capacity is described by its DOF. The second part principally deals with CSIT issues and here we have tried to determine how much CSIT feedback is optimal for sum rate maximization if the communication resources utilized for CSIT acquisition are completely accounted for. In this part, we also propose a novel CSIT acquisition strategy for reciprocal channels which improves the trade-off of resource utilization for CSIT acquisition and its quality. The third part of this thesis is dedicated to the study of the joint problem of transmit power minimization and user scheduling. In the following, we briefly discuss the main results and the conclusions of this thesis on a per chapter basis.

- **DOF for SISO Doubly Selective Channels**

In this chapter, the pre-log expression for doubly selective SISO underspread channels is derived and is shown that the loss in pre-log as

compared to the corresponding coherent channels is equal to the channel spread factor. For doubly selective overspread channels, a scheme based upon zero-padding is proposed which shows the achievability of the logarithmic scaling for such channels which is the first result in literature showing the existence of this regime for overspread channels.

- **MU MIMO: DOF with no CSI**

In this chapter, it is shown that for a DL channel with M -antenna BS and K ($K \geq M$) single-antenna users with no CSIT ergodic capacity is equal to the capacity of one single link in this DL, hence no gains due to the presence of multiple users are achievable. For the same system working under TDD mode, the achievability of $M[1 - (M + 1)/T]$ DOF is shown to be achievable hence CSIT availability may give enormous gains in DOF even if its resource utilization is accounted for.

- **Feedback Optimization in MU TDD Systems**

In this chapter, the sum rate maximization problem is treated for a DL channel removing all the assumptions about the presence of CSI and the optimization is carried over the amount of CSI feedback where all the resources used for CSI acquisition are taken into account. A novel sum rate lower bound is derived which captures all the benefits coming with CSIT (interference cancellation and multi-user diversity gain) and the associated costs (channel coherence reduction and reduced multi-user diversity gain). This bound and its approximation allow the determination of optimal amount of feedback for sum rate maximization under any set of system parameters.

- **Novel CSIT Acquisition for Reciprocal Channels**

In reciprocal systems, the CSIT is acquired by using simple training sequences and hence the shared knowledge of the channel gets ignored. In this chapter, we propose a novel CSIT acquisition scheme for reciprocal DL channels which combines the reciprocity exploitation through training sequences and the user's channel knowledge through quantized channel feedback. It is shown that with a judicious choice of the resource split between training and quantized feedback, a CSIT of much better quality can be obtained as compared to the traditional scheme for the same global resource utilization.

- **Transmit Power Minimization with User Selection**

In this chapter, the dual problem to the sum rate maximization namely the minimization of transmit power to achieve certain QoS at users is

treated in conjunction with user scheduling. Closed form expressions of minimum average transmit power are derived for various user selection schemes. It is further shown that semi-orthogonal greedy user selection performs much better than norm-based and angle-based selection schemes.

7.2 Future Perspectives

This thesis has been an attempt to solve many relevant problems in wireless communications and communications theory. As is the usual case with research, the pursuit of the solution for one problem reveals multiple open problems and further research directions and this thesis has certainly been no exception in this regard. In the following, we describe some possible interesting research directions which are either the ideas we came across while working for this thesis or they are the possible future extensions of the work conducted in this thesis.

- For single-user SISO doubly selective overspread channels, we have given an achievability result which shows the existence of logarithmic scaling of the capacity with SNR for such channels but the characterization of exact DOF of capacity still remains an open problem.
- For the broadcast problem with reciprocal channels at high SNR, we have derived lower and upper bounds of the DOF taking into account all the CSI resource utilization. Although the bounds are pretty close yet they are not matched. Derivation of an upper bound matched to the lower bound or vice versa will characterize the exact DOF of such non-coherent multi-user channels.
- The problem of sum rate maximization over the amount of CSI feedback was treated in this thesis for DL reciprocal channels. The reciprocity eases the CSIT acquisition and gives a nice setup for this problem formulation. To study the same problem for an FDD system is very interesting and useful as currently most of the wireless standards use FDD mode.
- We proposed a novel CSIT acquisition scheme for reciprocal channels combining training and quantized feedback. The proposed novel scheme holds verbatim in the case of multiple users. In the first phase of “pure training”, the users should use orthogonal training signals so that the BS gets an initial estimate of the channel. Then during the

second “quantized feedback” phase, the UL channel should be used as MIMO-MAC. The optimization of resources remains however an open problem in this setting. In this scenario, the resource optimization will depend heavily upon the BS transmission strategy, e.g., the optimal resource split could be extremely different for TDMA or SDMA. The presence of more users in the system larger than the BS transmit antennas and subsequently required user scheduling would add an extra twist to this problem.

- About the hybrid CSIT acquisition scheme combining training and quantized feedback for reciprocal channels, there are different ways to treat the fully general case of multiple users with multiple antennas where even a single user can be transmitted multiple streams. It adds an extra level of complexity to the open problem of multiple single-antenna users. For the users with multiple antennas, a simplifying strategy could be to do antenna combining as in [71] to minimize the quantization error. This scheme is promising as it reduces the feedback requirement by converting the MIMO channel into a vector channel and in a direction of minimal quantization error. Hence effectively it will become the multiple single-antenna user extension of our work. But the fully general case remains an open problem.

Bibliography

- [1] C. Shannon, “A mathematical theory of communications,” *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 1948.
- [2] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1991.
- [3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K. Cambridge Univ. Press, 2005.
- [4] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, pp. 585–595, November 1999.
- [5] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.
- [6] G. Caire and S. S. (Shitz), “On the achievable throughput of a multi-antenna Gaussian broadcast channel,” *IEEE Trans. on Information Theory*, vol. 49, pp. 1691–1706, July 2003.
- [7] P. Viswanath and D. Tse, “Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality,” *IEEE Trans. on Information Theory*, vol. 49, pp. 1912–1921, August 2003.
- [8] W. Yu and J. M. Cioffi, “Sum capacity of Gaussian vector broadcast channels,” *IEEE Trans. on Information Theory*, vol. 50, pp. 1875–1892, September 2004.
- [9] R. G. Gallager, “An inequality on the capacity region of multiple access multipath channels,” *Communications and Cryptography: Two sides of one tapestry*, Kluwer, pp. 129–139, 1994.

-
- [10] R. S. Cheng and S. Verdu, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. on Information Theory*, vol. 39, pp. 773–785, May 1993.
- [11] D. Tse and S. Hanly, "Multi-access fading channels: Part i: Polymatroidal structure, optimal resource allocation and throughput capacities," *IEEE Trans. on Information Theory*, vol. 44, pp. 2796–2815, November 1998.
- [12] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. International Conference on Communications, Seattle, USA*, June 1995, pp. 331–335.
- [13] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Transactions on Information Theory*, vol. 51, pp. 506–522, February 2005.
- [14] —, "A comparison of time-sharing, DPC and beamforming for MIMO broadcast channels with many users," *IEEE Transactions on Communications*, vol. 55, pp. 11–15, January 2007.
- [15] D. Gesbert, M. Kountouris, J. R. W. Heath, C. B. Chae, and T. Salzer, "From single user to multiuser communications: Shifting the MIMO paradigm," *IEEE Sig. Proc. Magazine*, 2007.
- [16] T. Cover, "Broadcast channels," *IEEE Trans. on Information Theory*, vol. 18, pp. 2–14, January 1972.
- [17] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. on Information Theory*, vol. 52, pp. 5045–5060, November 2006.
- [18] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA*, November 2006, pp. 359–363.
- [19] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. on Signal Processing*, vol. 54, pp. 1268–1278, April 2006.
- [20] D. J. Love, R. W. H. Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, pp. 54–59, October 2004.

-
- [21] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, September 2007.
- [22] D. J. Love *et al.*, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 1341–1365, October 2008.
- [23] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communications link in Rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 45, pp. 139–157, January 1999.
- [24] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, February 2002.
- [25] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 3095–3110, December 2004.
- [26] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 2426–2467, October 2003.
- [27] ———, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. on Information Theory*, vol. 52, pp. 437–453, February 2006.
- [28] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. on Information Theory*, vol. 51, pp. 437–446, February 2005.
- [29] R. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. on Information Theory*, vol. 52, pp. 1576–1608, April 2006.
- [30] N. Jindal and A. Goldsmith, "Dirty paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. on Information Theory*, vol. 51, pp. 1783–1794, May 2005.
- [31] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Information Theory*, vol. 52, pp. 3936–3964, September 2006.

-
- [32] U. Salim and D. Slock, "Capacity pre-log achieving schemes for stationary frequency-flat multi-antenna channels," in *Proc. Allerton Conf. on Communication, Control, and Computing*, 2007.
- [33] M. Schubert and H. Boche, "Solution of the multi-user downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 18–28, January 2004.
- [34] M. Schubert, "Power-aware spatial multiplexing with unilateral antenna cooperation," Ph.D. dissertation, TU Berlin, 2003.
- [35] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proc. IEEE Vehicular Technology Conference Fall*, September 2002.
- [36] M. Schubert and H. Boche, "Joint 'dirty paper pre-coding and downlink beamforming," in *Proc. ISSSTA*, September 2002, pp. 536–540.
- [37] X. Zhang, E. Jorswieck, and B. Ottersten, "User selection schemes in multiple antenna broadcast channels with guaranteed performance," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*, 2007.
- [38] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, March 2006.
- [39] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Transactions on Information Theory*, vol. 52, pp. 2572–2583, September 2004.
- [40] T. Koch and A. Lapidath, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proc. IEEE International Symposium on Information Theory*, vol. 52, September 2005, pp. 437–453.
- [41] M. Guillaud and D. Slock, "Channel Modeling and Associated Inter-Carrier Interference Equalization for OFDM Systems with High Doppler Spread," in *Proc. Int'l Conf. Acoustics, Speech and Sig. Proc. (ICASSP)*, Hong Kong, Apr. 6-10 2003.
- [42] T. Koch and A. Lapidath, "Multipath channels of bounded capacity," in *Proc. IEEE Information Theory Workshop*, May 2008, pp. 6–10.

-
- [43] N. Jindal, "A high SNR analysis of MIMO broadcast channels," in *Proc. IEEE Int. Symp. Information Theory, Adelaide, Australia*, 2005, pp. 2310–2314.
- [44] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, pp. 2058–2080, April 2003.
- [45] D. Samardzija and N. B. Mandayam, "Unquantized and uncoded channel state information feedback in multiple antenna multiuser systems," *IEEE Transactions on Communications*, vol. 54, pp. 1335–1345, July 2006.
- [46] G. Caire and K. Narayanan, "On the distortion SNR exponent of hybrid digital-analog space time coding," *IEEE Trans. on Information Theory*, vol. 53, pp. 2867–2878, August 2007.
- [47] S. Murugesan, E. Uysal-Biyikoglu, and P. Schniter, "Optimization of training and scheduling in the non-coherent MIMO multiple-access channel," *IEEE Journal on Selected Areas in Communication*, vol. 25, pp. 1446–1456, September 2007.
- [48] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO downlink made practical : Achievable rates with simple channel state estimation and feedback schemes," *Arxiv preprint cs.IT/0710.2642*.
- [49] M. Kobayashi, G. Caire, N. Jindal, and N. Ravindran, "How much training and feedback are needed in MIMO broadcast channels?" in *Proc. IEEE Int. Symp. Information Theory*, 2008.
- [50] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel feedback for MIMO broadcast channels," in *Proc. IEEE International Conference on Communications, Beijing, China*, 2008, pp. 3684–3688.
- [51] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO TDD systems," in *Proc. IEEE International Conference on Communications, Beijing, China*, 2008, pp. 4100–4105.
- [52] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Information Theory*, vol. 50, pp. 1893–1909, September 2004.

-
- [53] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and precoding in multi-user multiple antenna time division duplex systems," *Arxiv preprint <http://arxiv.org/abs/0812.0621v1>*.
- [54] Z. Tu and R. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Communication Letters*, vol. 7, pp. 370–372, August 2003.
- [55] U. Salim and D. Slock, "Broadcast channel: Degrees of freedom with no CSIR," in *Proc. Allerton Conf. on Communication, Control, and Computing*, 2008.
- [56] N. Jindal, A. Lozano, and T. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *Proc. IEEE Int. Symp. Information Theory*, 2009.
- [57] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, pp. 933–946, May 2000.
- [58] D. Gesbert and M. S. Alouini, "How much feedback is multi-user diversity really worth?" in *Proc. IEEE International Conference on Communications, Paris, France*, 2004.
- [59] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, June 2002.
- [60] P. Tejera and W. Utschick, "Feedback of channel state information in wireless systems," in *Proc. International Conference on Communications*, 2007, pp. 908–913.
- [61] K. R. Kumar and G. Caire, "Channel state feedback over the MIMO-MAC," in *Proc. IEEE Int. Symp. Information Theory*, 2009.
- [62] U. Salim and D. Slock, "How many users should inform the BS about their channel information?" in *Proc. International Symposium on Wireless Communication Systems*, 2009.
- [63] C. Steger and A. Sabharwal, "Single-input two-way SIMO channel: diversity-multiplexing tradeoff with two-way training," *IEEE Trans. on Wireless Communications*, vol. 7, pp. 4877–4885, December 2008.
- [64] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

-
- [65] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array-Part 1: Algorithms," *IEEE Trans. on signal processing*, vol. 44, pp. 1184–1197, May 1996.
- [66] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, United Kindom: Cambridge University Press, 1985.
- [67] C. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. on Wireless Communications*, vol. 6, pp. 458–462, February 2007.
- [68] M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1964.
- [69] H. A. David, *Order Statistics*. New York: Wiley, 1980.
- [70] M. A. Maddah-Ali, M. A. Sadrabadi, and A. K. Khandani, "Broadcast in MIMO systems based on a generalized QR decomposition: Signaling and performance analysis," *IEEE Trans. on Information Theory*, vol. 54, pp. 1124–1138, March 2008.
- [71] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Trans. on Wireless Communications*, vol. 10, pp. 3834–3844, October 2008.