

# Reverse Channel Training in Multiple Antenna Time Division Duplex Systems

A Thesis

Submitted for the Degree of  
**Doctor of Philosophy**  
in the Faculty of Engineering

Submitted by  
Bharath B. N.



Electrical Communication Engineering  
Indian Institute of Science, Bangalore  
Bangalore – 560 012 (INDIA)

Jan. 2013

TO

My parents

*Smt. Shantha and Sri. Nagaraja*

and

my grandmother

*Smt. Janakamma*

# Acknowledgements

I sincerely thank my advisor Dr. Chandra R. Murthy for his continuous support and encouragement throughout my PhD work. He has not only taught me the way of doing research but also to present the results in an elegant fashion. He has been a source of inspiration to me. I thank the faculty members of the ECE, Mathematics, EE and CSA departments, where I have attended and enjoyed many fundamental courses. I also thank the members of the ECE office.

The journey of my life as a PhD student would not have been fruitful without the support of my friends. I take this opportunity to thank my lab-mates Sanjeev, Venugopalakrishna, Abhay, Ganesan, Ranjitha, Parthajit, Nirmal, Srinivas, Krishna, Santhosh, Sinchu, Karthik, Chandrashekar, Suma and Deepa from whom I have learnt a lot, both technical and nontechnical. I am grateful to Nagananda, Abhay, Sanjeev, Venugopalakrishna and Ranjitha for making my coffee time worthwhile. I would also like to thank Karthik of the Next Generation Wireless Lab for having interesting technical and non-technical discussions. I thank my IISc friends Harshan, Jithamitra, Raghava, Shubha, Pavan, Prasad, Haricharan, Vineeth, Nidhin and Nischal. Many thanks to my hostel mates Srinidhi, Naveen, Keshav, Sushruth, Premkumar, Pradeepa, Avinash, Sivarama, Shivananju, Prashanth, Manjesh, Bharath and Venkatesh. Especially, I would like to thank Srinidhi, Naveen, Keshav, Venu and Sushruth for spending the 10pm discussion with me, the subject matter of which ranged from philosophy to physics. My childhood friends Raghavendra, Vijay, Sreejith and Arvind have played a major role in my life. My thanks also go to my undergraduate friends Narayan, Karthik and Naveen.

My PhD dream would have remained as a dream without the love and encouragement of my family members. I begin by thanking my cousin Pramod for not only introducing me to the world of signal processing during my undergraduation but also

for his help in numerous circumstances. I would like to thank my cousins Srivatsa, and Ranjini. My special thanks go to Smt. Ganga, Sri. Prabhakar, Prasad, Vani, Prithvi, Deepthi, and Pranav. I will forever be indebted to them. I owe my sincere gratitude to Prof. Hirannaiah, Smt. Uma and Amogh. I would also like to thank all my paternal and maternal uncles and aunts, and their families. The acknowledgement would be incomplete without thanking Gopalakrishna, Chandrashekhara, Smt. Anitha, Nanjundaswami, Smt. Sharada, Smt. Vishali, Shresta, Smt. Seethamma and Ganeshaiiah for making my childhood days with them a memorable experience. May peace be with them. I thank all my cousins for their kind support. Thanks are also due to my maternal grandparents Smt. Gundamma and Sri. Nanjundaiah, and my maternal uncles Ramesh and Manjunath.

Finally, I would like to thank my parents Smt. Shantha and Sri. Nagaraja, and my paternal grandmother Smt. Janakamma for their unconditional love and encouragement. I am grateful to them for the support without which I could not have pursued my dreams. I dedicate my thesis to them.

# Abstract

Multiple-Input Multiple-Output (MIMO) communication using multiple antennas has received significant attention in recent years, both in the academia and industry, as they offer additional spatial dimensions for high-rate and reliable communication, without expending valuable bandwidth. However, exploiting these promised benefits of MIMO systems critically depends on fast and accurate acquisition of Channel State Information (CSI) at the Receiver (CSIR) and the Transmitter (CSIT). In Time Division Duplex (TDD) MIMO systems, where the forward channel and the reverse channel are the same, it is possible to exploit this reciprocity to reduce the overhead involved in acquiring CSI, both in terms of training duration and power. Further, many popular and efficient transmission schemes such as beamforming, spatial multiplexing over dominant channel modes, etc. do not require full CSI at the transmitter. In such cases, it is possible to reduce the Reverse Channel Training (RCT) overhead by only learning the part of the channel that is required for data transmission at the transmitter.

In this thesis, we propose and analyze several novel channel-dependent RCT schemes for MIMO systems and analyze their performance in terms of (a) the mean-square error in the channel estimate, (b) lower bounds on the capacity, and (c) the diversity-multiplexing gain tradeoff. We show that the proposed training schemes offer significant performance improvement relative to conventional channel-agnostic RCT schemes. The main take-home messages from this thesis are as follows:

- Exploiting CSI while designing the RCT sequence improves the performance.
- The training sequence should be designed so as to convey only the part of the CSI required for data transmission by the transmitter.
- Power-controlled RCT, when feasible, significantly outperforms fixed power RCT.

# Glossary

AWGN	:	Additive White Gaussian Noise
BER	:	Bit Error Rate
<b>BS</b>	:	Base Station
BF	:	Beamforming
CSI	:	Channel State Information
CSIR	:	CSI at the Receiver
CSIT	:	CSI at the Transmitter
CSIB	:	CSI at <i>node B</i>
CSIA	:	CSI at <i>node A</i>
CLB	:	Capacity Lower Bound
CSCG	:	Circularly Symmetric Complex Gaussian
DMT	:	Diversity-Multiplexing Gain Tradeoff
i.i.d.	:	Independent and Identically Distributed
MIMO	:	Multiple-Input Multiple-Output
MISO	:	Multiple-Input Single-Output
ML	:	Maximum Likelihood
MMSE	:	Minimum Mean Square Error
PCRCT	:	Power-Controlled RCT
RCT	:	Reverse Channel Training
SIMO	:	Single-Input Multiple-Output
SM	:	Spatial Multiplexing
SISO	:	Single-Input Single-Output
SNR	:	Signal-to-Noise Ratio
SVD	:	Singular Value Decomposition
TDD	:	Time Division Duplex

# Notation

$n_A$	: Number of antennas at <i>node A</i> or the transmitter
$n_B$ or $r$	: Number of antennas at <i>node B</i> or the receiver
$M$	: Number of downlink users
$P_{B,\tau}$	: Training power at <i>node B</i>
$P_{A,\tau}$	: Training power at <i>node A</i>
$P_{A,d}$	: Data power at <i>node A</i>
$L_{A,d}$	: Data duration corresponding to <i>node A</i>
$L_{A,\tau}$	: Training duration corresponding to <i>node A</i>
$L_{B,\tau}$	: Training duration corresponding to <i>node B</i>
$L_c$	: Coherence time of the channel
$\Re(\cdot)$	: Real part of the complex argument
$\Im(\cdot)$	: Imaginary part of the complex argument
$(\cdot)^T$	: Transposition
$(\cdot)^H$	: Hermitian transposition
$(\cdot)^*$	: Complex conjugation
$\mathbb{E}[\cdot]$	: Expectation operator
$(\cdot)^+$	: Signum function
$ \cdot $	: Absolute value of a complex number or the determinant of a matrix or the cardinality of a set, depending on the context
$\ \cdot\ _2$	: Euclidean norm of a vector
$\ \cdot\ _F$	: Frobenius norm of a matrix
$I_n$	: $n \times n$ identity matrix
$\lfloor c \rfloor$	: Largest integer less than $c$
$\mathcal{C}$	: Field of complex numbers
$\mathcal{R}$	: Field of real numbers
$\mathcal{R}^+$	: Field of non-negative real numbers
$\mathcal{CN}(\mu, \sigma^2)$	: Circularly symmetric complex Gaussian distribution with mean $\mu$ and $\sigma^2$ variance
<b>Boldface lower case letters</b>	: Vectors
<b>Upper case letters</b>	: Matrices

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Glossary</b>	<b>iv</b>
<b>Notation</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Wireless Channel Model . . . . .	2
1.2 Channel Knowledge and Its Acquisition . . . . .	4
1.3 Outline of the Thesis . . . . .	7
1.4 List of Publications . . . . .	10
<b>2 On the DMT of TDD-SIMO Systems with Channel-Dependent Reverse Channel Training</b>	<b>12</b>
2.1 Introduction . . . . .	12
2.2 System Model . . . . .	17
2.3 DMT Analysis with Perfect CSIR . . . . .	20
2.3.1 Power-Controlled Data Transmission from <i>Node A</i> to <i>Node B</i> . . . . .	20
2.3.2 Achievable DMT Analysis . . . . .	24
2.4 Three Way Training . . . . .	25
2.4.1 Power Control Scheme . . . . .	25
2.4.2 Power-Controlled Forward Link Training Scheme . . . . .	26
2.4.3 Power-Controlled Data Transmission from <i>Node A</i> to <i>Node B</i> . . . . .	27
2.4.4 Achievable DMT Analysis . . . . .	28

2.5	Discussion . . . . .	28
2.6	Simulation Results . . . . .	31
2.7	Conclusions . . . . .	31
<b>3</b>	<b>PCRCT Achieves an Infinite Diversity Order in a TDD-SIMO System with Perfect CSIR</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	System Model . . . . .	38
3.3	PCRCT and Data Transmission Protocol . . . . .	39
3.4	Outage Analysis . . . . .	41
3.5	An Achievable Coding Scheme . . . . .	43
3.6	Numerical Results . . . . .	44
3.7	Conclusions . . . . .	45
<b>4</b>	<b>Channel Training Signal Design for Reciprocal Multiple Antenna Systems with Beamforming</b>	<b>48</b>
4.1	Introduction . . . . .	48
4.2	System Model . . . . .	53
4.3	RCT Signal Design With Perfect CSIB . . . . .	55
4.3.1	Approximate Capacity Lower Bound . . . . .	58
4.3.2	Optimal Training Duration . . . . .	59
4.3.3	Optimal Sharing of Resources . . . . .	60
4.4	Forward-Link Training With Perfect CSIA . . . . .	63
4.5	<i>Node A</i> Initiated Training Versus <i>Node B</i> Initiated Training . . . . .	68
4.5.1	<i>Node A</i> Initiated Training . . . . .	69
4.5.2	<i>Node B</i> Initiated Training . . . . .	69
4.6	Simulation Results and Discussion . . . . .	70
4.6.1	<i>Node A</i> Initiated Training . . . . .	71
4.6.2	<i>Node B</i> Initiated Training . . . . .	76
4.6.3	<i>Node A</i> Initiated Training Vs. <i>Node B</i> Initiated Training . . . . .	76
4.7	Conclusions . . . . .	77

<b>5</b>	<b>Reverse Channel Training in a Multi-User TDD-MIMO SM System</b>	<b>82</b>
5.1	System Model . . . . .	87
5.2	Design and Optimization of RCT for the Scheduled User . . . . .	90
5.2.1	MSE Optimal $D$ and $\phi_c$ . . . . .	93
5.2.2	Data Transmission and Capacity Lower Bound . . . . .	96
5.2.3	Capacity Lower Bound Optimal $D$ , $\phi_c$ and $L_{B,\tau}$ . . . . .	98
5.3	Multiuser Scenario . . . . .	100
5.4	Simulation Results . . . . .	102
5.4.1	Mean Square Error . . . . .	103
5.4.2	Data Rate with a Single User . . . . .	104
5.4.3	Data Rate with Multiple Users . . . . .	106
5.5	Conclusions . . . . .	106
<b>6</b>	<b>Conclusions and Future Work</b>	<b>113</b>
6.1	Conclusions . . . . .	113
6.2	Future Work . . . . .	116
<b>A</b>	<b>Appendix for Chapter 2</b>	<b>118</b>
A.0.1	Useful Lemmas . . . . .	118
A.0.2	Proof of Lemma 1 . . . . .	119
A.0.3	Proof of Theorem 1 . . . . .	122
A.0.4	Proof of Lemma 2 . . . . .	124
A.0.5	Proof of Theorem 2 . . . . .	125
<b>B</b>	<b>Appendix for Chapter 3</b>	<b>129</b>
B.0.6	Proof of Theorem 4 . . . . .	129
<b>C</b>	<b>Appendix for Chapter 4</b>	<b>133</b>
C.0.7	Constrained Cramér-Rao Lower Bound . . . . .	133
C.0.8	Derivation of (4.12) . . . . .	135
C.0.9	Proof of Theorem 5 . . . . .	135
C.0.10	Proof of Theorem 6 . . . . .	136
C.0.11	Derivation of (4.25) . . . . .	136
C.0.12	Proof of Theorem 7 . . . . .	138

---

C.0.13 Proof of Theorem 8 . . . . .	139
<b>D Appendix for Chapter 5</b>	<b>140</b>
D.0.14 Proof of Theorem 9 . . . . .	140
D.0.15 Proof of Lemma 4 . . . . .	142
D.0.16 Proof of Theorem 10 . . . . .	143
D.0.17 Proof of Theorem 11 . . . . .	146
D.0.18 Proof of Theorem 12 . . . . .	146
D.0.19 Proof of Theorem 13 . . . . .	147
<b>Bibliography</b>	<b>148</b>

# List of Figures

1.1	System model of a reciprocal MIMO considered in this thesis. . . . .	3
1.2	MSE versus training power in the reverse-link in a TDD-SISO system. Here, <i>conventional</i> refers to the channel-agnostic training scheme, whereas <i>improved</i> refers to the proposed channel-dependent training sequence. . . . .	7
2.1	System model for reverse channel training with perfect CSIR used in Section 2.3. . . . .	22
2.2	The achievable DMT with the training and power control scheme proposed in Sec. 2.3, compared with the performance of the orthogonal RCT and the data power control proposed in [1,2] (and appropriately accounting for the training duration overhead and switching off antennas to achieve higher values of $g_m$ ). The plot corresponds to a SIMO system with $r = 5$ antennas, with coherence time $L_c = 20$ symbols, and reverse training duration of $L_{B,\tau} = 1$ symbol. . . . .	29
2.3	Outage probability versus the average data power $\bar{P}$ for the fixed-power training scheme proposed in Sec. 2.3, with the data power control scheme given by (2.10) with $s = 1$ . Here, $r = 3, L_c = 40$ and $L_{B,\tau} = 1$ . With $g_m = 0.8$ , the target data rate was set as $R_{\bar{P}} = 4 + g_m \log \bar{P}$ to facilitate the comparison of the curves. . . . .	33
2.4	Outage probability versus the average data power $\bar{P}$ for the fixed-power training scheme proposed in Sec. 2.3, with the data power control scheme given by (2.10) with $s = r$ . Here, $r = 3, L_c = 40$ and $L_{B,\tau} = 1$ . . . . .	34
3.1	System model for PCRCT with perfect CSIR. . . . .	39

3.2	Upper bound on the outage probability versus data power for different values of $\gamma$ , with $r = 3, g_m = 0.9, L_c = 10$ , and $L_{B,\tau} = 1$ . . . . .	46
3.3	A plot of $-\log(P_{out}^u)$ in (3.9) versus training power with $r = 3, \gamma = 1$ and $L_c = 100$ for different values of $g_m$ and $L_{B,\tau}$ . The curves labeled orthogonal refer to the lower bound on the outage probability with channel-agnostic orthogonal RCT in [3], which requires a minimum training duration of $r = 3$ symbols. . . . .	47
4.1	<i>Node A</i> initiated training. . . . .	56
4.2	<i>Node B</i> initiated training. . . . .	64
4.3	Capacity lower bound in (4.10) versus data power for $P_{B,\tau} = 14$ and 6dB. Here, we have used a $3 \times 3$ MIMO system with $L_c = 100$ . The figure demonstrates that the optimal $L_{B,\tau}$ is not always equal to one. . . . .	72
4.4	Capacity lower bound in (4.10) versus training power with $P_{A,d} = 4$ dB for a $3 \times 3$ MIMO system, and $L_c = 100$ . The figure demonstrates that $L_{B,\tau} = 1$ is not always optimal. It also illustrates that quantization of the BF vectors in a fixed-point implementation has a negligible effect on the performance for $B \geq 6$ bits per real dimension. . . . .	73
4.5	Capacity lower bound versus data power for a $3 \times 3$ MIMO system with $L_c = 100$ and $P_{B,\tau} = 2P_{A,d}$ . Using the low SNR optimal $L_{B,\tau}$ is nearly optimal for a wide range of training/data powers. Also, the loss in the data rate due to using the low SNR approximation is negligible. . . . .	75
4.6	Capacity lower bound (exact and approximate) versus data power for a $3 \times 3$ MIMO system with training power $P_{B,\tau} = P_{A,d}$ , and $L_c = 100$ . The figure illustrates the performance gain offered by the proposed training method over the conventional orthogonal training scheme (e.g., [4, 5]), and shows that the approximate lower bound is tight at all data powers. . . . .	78
4.7	Exact and approximate capacity lower bound versus the average power ( $\rho$ ) for the energy efficient resource sharing scheme analyzed in Sec. 4.3.3, with $L_{B,\tau} = 1$ . The figure illustrates the tightness approximate bound and that the sharing of powers derived in Theorem 7 is optimal. . . . .	79

4.8	$C_{A,B,\text{approx}}$ in (4.25) versus the average power ( $\rho$ ) for a $3 \times 3$ MIMO system for the <i>node B</i> initiated training scheme. . . . .	80
4.9	Plot of $S_A$ and $S_B$ in (4.31) and (4.32) corresponding to <i>node A</i> initiated training and <i>node B</i> initiated training scaled by $\rho \mathbb{E}\sigma_1^2$ versus the coherence time ( $L_c$ ) for an $n_A = 8, n_B = 5$ MIMO system with $\rho = 0\text{dB}$ and $10\text{dB}$ . . . . .	81
5.1	The MIMO $n_B \times n_A$ SM system, showing the RCT and the forward-link data transmission. . . . .	90
5.2	MSE versus training power for a $3 \times 4$ MIMO system with $m = 3$ . . . . .	104
5.3	The figure demonstrates the tightness of the proposed approximation in (5.23) for the capacity lower bound in (5.22) for a $3 \times 4$ MIMO system with the data power of $P_{A,d} = P_{B,\tau}, L_{B,\tau} = 3$ symbols. . . . .	108
5.4	Capacity lower bound in (5.22) for a $3 \times 4$ MIMO system versus RCT power $P_{B,\tau}$ , with $P_{A,d} = P_{B,\tau}, L_{B,\tau} = 3$ symbols and $m = 3$ modes. . . . .	109
5.5	Capacity lower bound in (5.22) for a $3 \times 4$ MIMO system versus training duration with $P_{A,d} = P_{B,\tau} = 6\text{dB}$ . . . . .	110
5.6	Capacity lower bound in (5.22) for a $3 \times 4$ multiuser MIMO system, versus training power $P_{B,\tau}$ , with the data power $P_{A,d} = P_{B,\tau}$ , and for the scheduling scheme described in Sec. 5.3. . . . .	111
5.7	Normalized data rate versus number of users $M$ for a $2 \times 2$ multiuser MIMO BF system with $P_{A,d} = P_{B,\tau} = 6\text{dB}$ . Here, the data rate is normalized by $\log_2 \log M$ . . . . .	112

# List of Tables

2.1	Three way training in a TDD-SIMO system . . . . .	24
4.1	Optimum training duration for a $3 \times 3$ MIMO system with $L_c = 100$ and $P_{B,\tau} = 2P_{A,d}$ . . . . .	74

# Chapter 1

## Introduction

Over the past decade or so, increasing demand for data, audio and video services has lead to an enormous increase in the data rate requirement of wireless communication systems. One possible solution to cater to this ever-increasing demand for data rate is to increase the bandwidth and/or the transmit power. Unfortunately, the spectrum available for wireless communication is limited. Moreover, in a network, increasing the transmit power is often not a viable option, as higher power to one user acts as a strong interference to unintended users, resulting in reduced sum data rate. To make matters worse, unlike wired communication systems, wireless systems pose several other challenges such as fading, path loss, etc. Therefore, wireless communication engineers are faced with the challenge of designing systems that efficiently combat fading by providing higher data rate for a given bandwidth and transmit power. Multiple antennas at both the transmitter and the receiver is a key technology to meet these challenges. Physically placing the antennas with an inter-antenna distance of about 10 times the wavelength of the carrier frequency is known to result in fade values that are uncorrelated across antennas. Having independent fade values across antennas in turn yields

higher diversity and multiplexing gains by using methods such as Maximum Ratio Transmission (Combining) (MRT(C)), Spatial Multiplexing (SM), Beamforming (BF), etc [6, 7]. Practical realizability of multiple antenna communication systems require the antenna separation to be within a few centimeters for it to be implementable on the form factor of hand held devices such as mobile phones. This requires the transmitting devices to operate in the range of Giga-Hertz to ensure independent fade values, which is feasible in practice, and is also a part of many modern day wireless standards such as IEEE 802.11, 3GPP-LTE, LTE-Advanced, etc. Owing to the aforementioned benefits offered by multiple antenna communications, it has received tremendous attention in the recent years, both in academia and in industry. However, the professed benefits of multiple antenna systems are realizable only if the fade values are known at the receiver and/or the transmitter. Thus, fast and reliable acquisition of fade values at both ends of the communication system is an important aspect in designing modern wireless communication systems; this is the main focus of this thesis. A well-accepted mathematical model for Multiple-Input Multiple-Output (MIMO) systems is presented next.

## 1.1 Wireless Channel Model

Figure 1.1 shows a model for a point-to-point MIMO wireless communication system with  $n_A$  and  $n_B$  antennas at *node A* and *node B*, respectively. The wireless channel from *node A* to *node B* is modeled as a quasi-static block fading channel, represented by the matrix  $H \in \mathcal{C}^{n_B \times n_A}$ . That is, the channel is assumed to remain constant for a frame of duration  $L_c$ , and evolve in an i.i.d. fashion from frame to frame. We assume a Time Division Duplex (TDD) mode of operation. When the channel is perfectly reciprocal,

by considering the complex conjugate of the transmitted and received signals as the reverse channel input and output, respectively, the reverse-link channel is  $H^H$ . The baseband equivalent of the received signal at *node A* and *node B* corresponding to the input  $\mathbf{x}_A \in \mathcal{C}^{n_A}$  and  $\mathbf{x}_B \in \mathcal{C}^{n_B}$  are given by

$$\mathbf{y}_B = H\mathbf{x}_A + \mathbf{w}_B, \quad (1.1)$$

$$\mathbf{y}_A = H^H\mathbf{x}_B + \mathbf{w}_A, \quad (1.2)$$

where  $\mathbf{x}_A \in \mathcal{C}^{n_A}$  ( $\mathbf{x}_B \in \mathcal{C}^{n_B}$ ),  $\mathbf{y}_B \in \mathcal{C}^{n_B}$  ( $\mathbf{y}_A \in \mathcal{C}^{n_A}$ ) and  $\mathbf{w}_A \in \mathcal{C}^{n_A}$  ( $\mathbf{w}_B \in \mathcal{C}^{n_B}$ ) are the input, output and the additive noise vectors at *node A* (*node B*), *node B* (*node A*) and *node A* (*node B*), respectively. The input vectors are assumed to satisfy an average power constraint, i.e.,  $\mathbb{E}\{\mathbf{x}_A^H \mathbf{x}_A\} = P_A$  and  $\mathbb{E}\{\mathbf{x}_B^H \mathbf{x}_B\} = P_B$ , where  $P_A$  and  $P_B$  are the transmit powers available at *node A* and *node B*, respectively. The entries of the noise vectors  $\mathbf{w}_B \in \mathcal{C}^{n_B}$  and  $\mathbf{w}_A \in \mathcal{C}^{n_A}$  are distributed as  $\mathcal{CN}(0, 1)$ .

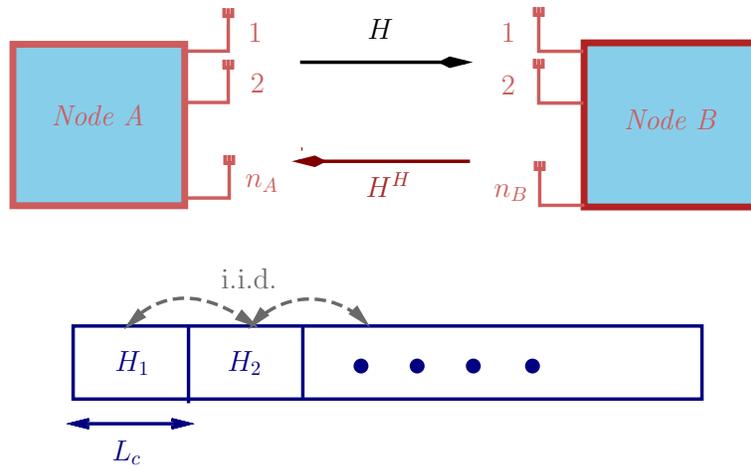


Figure 1.1: System model of a reciprocal MIMO considered in this thesis.

## 1.2 Channel Knowledge and Its Acquisition

It is well understood by now that the capacity and reliability of multiple antenna communication systems are significantly increased by having perfect channel knowledge at both *node B* and *node A* [7]. These gains are primarily due to the fact that multiple antenna communication enables methods such as MRT(C), BF etc, which requires channel knowledge at *node B* and/or *node A*. Therefore, in practical systems, efficient acquisition of channel knowledge at *node B* and *node A* is an important problem. Channel State Information (CSI) at *node B* (CSIB) can be acquired by sending a known training sequence from *node A* to *node B*, from which the latter computes an estimate of the channel. Since data and training signals are transmitted from *node A*, there exists an inherent tradeoff between resources, such as training duration, training power, data power and data duration that are spent for training and data transmission. Higher resources spent in training result in a higher estimation accuracy, which helps in improving the rate achievable during the data transmission phase. However, a higher training overhead leaves less resources for data transmission, which might undo the benefits of the improved channel estimation, and lead to a net reduction in the average data rate. The problem of sharing of resources between data and training in an optimal way was first studied in [8], where, among other things, it is shown that a training duration of  $n_A$  symbols is optimal in terms of a capacity lower bound. Since then, several researchers have studied the tradeoff under various scenarios; these will be elaborated upon in later chapters.

The capacity and reliability is further increased by having CSI at *node A* (CSIA), in addition to CSIB. For example, in a MIMO system with perfect CSIA, it can be shown that

the probability of error falls exponentially with SNR, which is not possible with systems with no CSIA [9]. In this thesis, we address the problem of efficiently acquiring CSIA in reciprocal MIMO systems. The conventional method for acquiring CSIA in reciprocal MIMO systems is to transmit a known training sequence such as the Hadamard matrix in the reverse-link, from which *node A* computes an estimate of the reverse channel  $H^H$ . This incurs a training overhead of at least  $n_B$  symbols, which could be significant, particularly when  $L_c$  is small, i.e., in a fast fading environment. Can we do better than the conventional training? In this thesis, we answer this question in the affirmative, by proposing novel channel-dependent Reverse Channel Training (RCT) schemes. The following example motivates the study of RCT sequence design considered in this thesis by comparing channel-agnostic conventional RCT with channel-dependent RCT that adapts to the current CSIB.

- *Example:* Consider a point to point reciprocal Single-Input Single-Output (SISO) system with perfect CSIB where the forward channel is a complex random variable  $h = |h|e^{j\theta}$  and the reverse channel is  $h^* = |h|e^{-j\theta}$ .<sup>1</sup> Note that knowledge of  $|h|$  would suffice at *node A* for power control or adaptive modulation and coding, since the phase  $\theta$  can be compensated for at *node B* without changing the noise statistics. In the conventional training, *node B* transmits

$$x_{B,\tau} = \sqrt{P_{B,\tau}} \cdot 1, \quad (1.3)$$

where  $P_{B,\tau}$  is the training power. The corresponding received training signal at

---

<sup>1</sup>The reverse channel is  $|h|e^{-j\theta}$  due to reciprocity.

*node A* is given by

$$y_{A,\tau} = \sqrt{P_{B,\tau}}h^* + w_A, \quad (1.4)$$

where  $w_A$  is distributed as  $\mathcal{CN}(0, 1)$ . Using  $y_{A,\tau}$ , *node A* computes an estimate of  $|h|$  as

$$|\hat{h}|_{(\text{conv})} \triangleq \frac{y_{A,\tau}^*}{\sqrt{P_{B,\tau}}} = \left| h + \frac{w_A^*}{\sqrt{P_{B,\tau}}} \right|. \quad (1.5)$$

Note that the conventional training method enables *node A* to estimate both  $|h|$  and  $\theta$ , but the estimation of  $\theta$  is not required for data transmission. Now, we consider a different, channel-dependent reverse training signal for estimating  $|h|$  at *node A*. Suppose we employ  $x_{B,\tau}^{(\text{prop})} = \sqrt{P_{B,\tau}}e^{-j\theta}$  as the training signal, then, the corresponding received signal at *node A* is given by

$$y_{A,\tau} = \sqrt{P_{B,\tau}}e^{-j\theta}h + w_A = \sqrt{P_{B,\tau}}|h| + w_A. \quad (1.6)$$

Note that the proposed training signal also satisfies the training power constraint of  $P_{B,\tau}$  since  $|e^{-j\theta}| = 1$ . *Node B* computes an estimate of  $|h|$  by discarding the imaginary part, as follows

$$|\hat{h}|_{(\text{prop})} \triangleq \frac{|\Re\{y_{A,\tau}\}|}{\sqrt{P_{B,\tau}}} = \left| |h| + \frac{\Re\{w_A\}}{\sqrt{P_{B,\tau}}} \right|. \quad (1.7)$$

In the above, since the desired part of the signal ( $|h|$ ) is real, the imaginary part of the noise does not corrupt the signal, and hence removing it reduces the noise variance by a factor of 2, and results in a reduced Mean Square Error (MSE). Figure 1.2 plots the MSE in the estimation of  $|h|$  as a function of the training

power in dB. The improvement in the MSE by employing the proposed channel-dependent RCT is clear from the graph. This simple example illustrates that the use of channel-dependent RCT can enable *node A* to efficiently estimate only the part of the CSI that is required for data transmission.

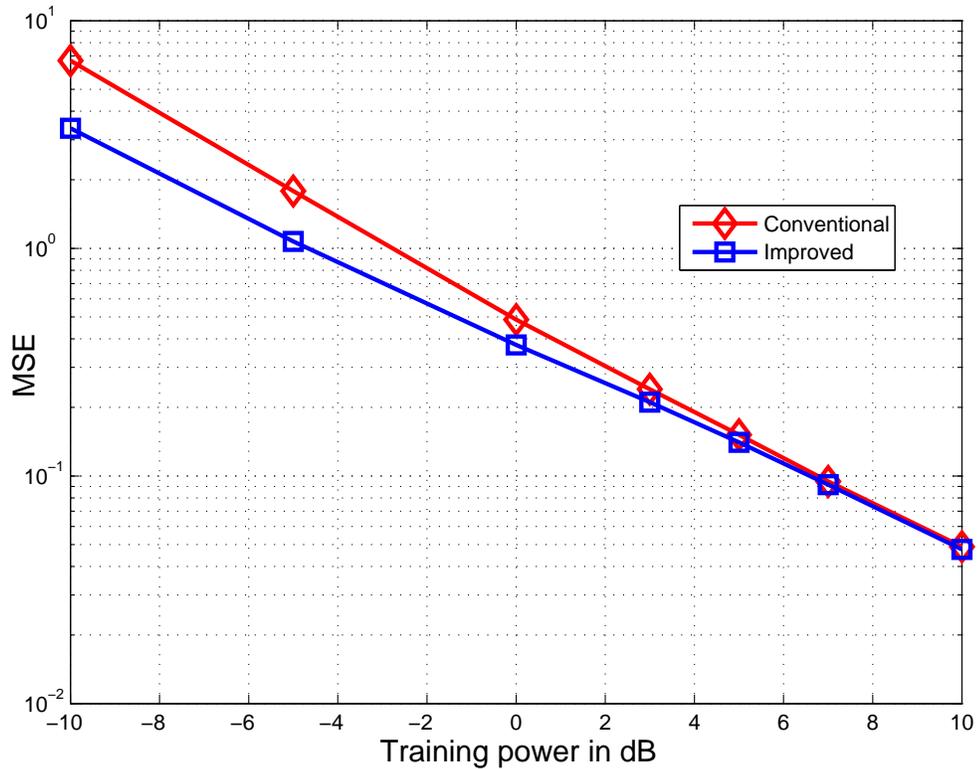


Figure 1.2: MSE versus training power in the reverse-link in a TDD-SISO system. Here, `conventional` refers to the channel-agnostic training scheme, whereas `improved` refers to the proposed channel-dependent training sequence.

### 1.3 Outline of the Thesis

Chapter 2 of this thesis generalizes the idea of channel-dependent RCT presented in the example above to a Single-Input Multiple-Output (SIMO) channel, and investigates

the benefits offered by the scheme in terms of the Diversity Multiplexing gain Tradeoff (DMT). The following models for CSI are studied: (i) perfect CSIB and noisy CSIA, and (ii) noisy CSIB and noisy CSIA. In both cases, a channel-dependent fixed-power training scheme is proposed for acquiring CSIA, along with a forward-link data transmit power control scheme. With perfect CSIB, the proposed scheme is shown to achieve a diversity order that is quadratically increasing with the number of receive antennas. This is in contrast to conventional orthogonal RCT schemes, where the diversity order is known to saturate as the number of antennas at *node B* is increased, for a given  $L_c$ . Moreover, the proposed scheme can achieve a larger DMT compared to the orthogonal training scheme. With noisy CSIB and noisy CSIA, a three-way training scheme is proposed and its DMT performance is analyzed. It is shown that nearly the same diversity order is achievable as in the perfect CSIB case. The outage performance of the proposed scheme is illustrated through Monte Carlo simulations. The contents of this chapter have been published in parts in [10, 11].

Chapter 3 extends the constant power training proposed in chapter 2 to a power controlled RCT scheme that enables *node A* to directly estimate the power control parameter to be used for the forward-link data transmission. We show that our proposed scheme, with an RCT power of  $\bar{P}^\gamma$ ,  $\gamma > 0$ , and a forward data transmission power of  $\bar{P}$ , achieves an infinite diversity order for  $0 \leq g_m < \frac{L_c - L_{B,\tau}}{L_c} \min(\gamma, 1)$  and  $r > 2$ , where  $g_m$  is the multiplexing gain,  $L_c$  is the channel coherence time,  $L_{B,\tau}$  is the RCT duration and  $r$  is the number of receive antennas. We also derive an upper bound on the outage probability and show that it goes to zero asymptotically as  $\exp(-\bar{P}^E)$ , where  $E \triangleq \left(\gamma - \frac{g_m L_c}{L_c - L_{B,\tau}}\right)$ , at high  $\bar{P}$ . Thus, the proposed scheme achieves a significantly

better DMT performance compared to the finite diversity order achieved by channel-agnostic and fixed-power RCT schemes. The contents of this chapter have been published in [12].

In chapter 4, we propose and optimize channel-dependent training schemes for reciprocal MIMO channels with BF at the *node A* and *node B*. First, assuming that CSI is available at *node B*, a channel-dependent RCT signal is proposed that enables efficient estimation of the BF vector at *node A* with a minimum training duration of only one symbol. In contrast, conventional orthogonal training requires a minimum training duration equal to the number of receive antennas. A tight approximation to a lower bound on the capacity of the system is derived, which is used as a performance metric to optimize the parameters of the RCT. Next, assuming that CSI is available at *node A*, a channel-dependent *forward-link training* signal is proposed and its power and duration are optimized with respect to an approximate capacity lower bound. Finally, we demonstrate the significant performance improvement offered by the proposed channel-dependent training schemes over the existing channel-agnostic orthogonal training schemes through simulations. The contents of this chapter have been published in [13].

Chapter 5 extends the RCT scheme in chapter 4 to multiuser MIMO Spatial Multiplexing (SM) systems. In particular, using the channel knowledge at *node B*, a novel, channel-dependent power-controlled RCT sequence is proposed, using which the *node A* estimates the required BF vectors for the forward-link data transmission. Tight approximate expressions for (i) the Mean Square Error (MSE) in the estimate of the BF vectors, and (ii) a Capacity Lower Bound (CLB) for an SM system, are derived, and are

used to optimize the parameters of the training sequence. Moreover, an extension of the channel-dependent training scheme and the data rate analysis to a multiuser scenario with  $M$  user terminals is presented. For the single mode BF system, a closed-form expression for an upper bound on the average sum data rate is derived, which is shown to scale with  $M$  as  $\frac{L_c - L_{B,\tau}}{L_c} \log_2 \log M$ , where  $L_c$  and  $L_{B,\tau}$  are the channel coherence time and training duration, respectively. Using simulation results, the significant performance gain offered by the proposed training sequence over the conventional constant-power orthogonal RCT sequence is demonstrated. Also, it is shown that optimal spatial power allocation during training outperforms its equal power allocation counterpart, while optimal temporal power allocation only offers a marginal improvement in performance. The contents of this chapter have been published in part in [14], and has been submitted as [15].

## 1.4 List of Publications

### Journal Papers

1. B. N. Bharath and C. R. Murthy, "On the DMT of TDD-SIMO systems with channel-dependent reverse channel training," *IEEE Trans. on Commun.*, vol. 60, no. 11, pp. 3332-3341, Nov. 2012
2. B. N. Bharath and C. R. Murthy, "Power Controlled Reverse Channel Training Achieves an Infinite Diversity Order in a TDD-SIMO System with Perfect CSIR," *IEEE Commun. letters*, vol. 16, no. 11, pp. 1800-1803, Nov. 2012.
3. B. N. Bharath and C. R. Murthy, "Channel training signal design for reciprocal

- multiple antenna systems with beamforming," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, p. 1, Aug. 2012.
4. B. N. Bharath and C. R. Murthy, "Reverse Channel Training in a Multi-User TDD-MIMO Spatial Multiplexing System," *IEEE Trans. on Veh. Technol.*, accepted for publication, June 2013.

### Conference Papers

1. B. N. Bharath and C. R. Murthy, "Reverse channel training for reciprocal MIMO systems with spatial multiplexing," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, Taipei, Taiwan, Apr. 2009, pp. 2673-2676.
2. B. N. Bharath and C. R. Murthy, "On the improvement of diversity-multiplexing gain tradeoff in a training based TDD-SIMO system," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, Dallas, TX, USA, Mar. 2010, pp. 3366-3369.
3. R. Prasad, B. N. Bharath, and C. R. Murthy, "Joint Data Detection and Dominant Singular Mode Estimation in Time Varying Reciprocal MIMO Systems," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, Prague, Czech Republic, May 2011, pp. 3240-3243.
4. B. N. Bharath and C. R. Murthy, "Channel estimation at the transmitter in a reciprocal MIMO spatial multiplexing system," in *National Conf. on Commun. (NCC)*, Kharagpur, India, Feb. 2012, pp. 1-5.

# Chapter 2

## On the DMT of TDD-SIMO Systems with Channel-Dependent Reverse Channel Training

### 2.1 Introduction

Reliability and system throughput are two fundamental parameters of interest in any wireless communication system, and the inherent tradeoff between the two at high SNR was elegantly captured by the Diversity Multiplexing gain Tradeoff (DMT) proposed in the seminal work of Zheng and Tse [16]. It is known that a significant improvement in the outage performance can be obtained if the Channel State Information (CSI) at the receiver (CSIR) and the transmitter (CSIT) are perfect [17], [9], while [16] considered perfect CSIR and no CSIT.

In a Time Division Duplex (TDD) system, CSI could be estimated at the transmitter and receiver by sending a known training sequence in the forward and reverse-link directions, respectively. This has two consequences. First, the *estimation error* results in

incorrect data rate or power adaptation at the transmitter, in turn leading to higher outage rate. Second, training incurs a *time overhead*, which could be non-trivial when the training occupies a significant fraction of the channel coherence time, as it affects the pre-log term in the achievable data rate [8]. This chapter therefore focuses on the important problem of analytically comparing the DMT performance of different channel estimation techniques and identifying training signals and data power control schemes that result in a good performance in terms of the achievable DMT. We start with a brief survey of related literature.

The impact of imperfect CSIT on the DMT of a multiple antenna system has been a popular area of research, and it is known that even with imperfect CSIR and CSIT, a significant improvement in DMT can be obtained, compared to the no-CSIT case (see, for example, [18–20]). The effect of imperfect CSIR on the DMT of a MIMO system was first studied in [21]. The DMT analysis of a multiple antenna system with perfect CSIR and when the CSIT is modeled as the CSI plus Gaussian noise whose variance decreases with training SNR was investigated in [22–24]. In a TDD setup, the achievable DMT improvement using power control based on noisy CSIT was shown in [1, 24, 25]. Other works that study the DMT performance with quantized feedback of CSI and/or target data rate control based on noisy CSIT include [18, 19, 23, 26–29]. In [2, 28], the DMT of two-way and multi-round training schemes in a TDD system was derived. In these studies, the channel feedback signal on the reverse-link is chosen to satisfy an average power constraint, rather than an instantaneous power constraint.

Most of the aforementioned studies of the DMT with imperfect CSI typically ignore the training duration overhead. Hence, they are primarily applicable to slowly varying

channels, where the time overhead in training occupies an insignificant fraction of the channel coherence time. An exception is [1], where, taking the training overhead into account, the authors concluded that for nonzero multiplexing gain  $g_m$ , the diversity order saturates as  $r$  increases, where  $r$  is the number of receive antennas. Hence, for fast varying channels, the authors suggest turning off receive antennas in order to achieve higher multiplexing gains. It is important to account for the training duration overhead in deriving the achievable DMT, because, as the SNR goes to infinity, although the estimation error goes to zero, the training duration overhead remains fixed and has a direct impact on the DMT. Also, by modeling the CSIT as the sum of the true CSI and an additive error, most of the past studies implicitly assume that a channel-agnostic orthogonal training signal is employed for channel estimation. When the training signal is channel-dependent, the imperfect CSI can no longer be modeled as the sum of the true CSI and an additive noise. Due to this, the existing results cannot be directly extended to analyze the DMT performance of channel-dependent training schemes.

When the channel is reciprocal and block-fading, e.g., in a TDD system, the receiver could exploit its channel knowledge (acquired through an initial forward-link training phase) in designing its reverse-training sequence, not only to reduce the channel estimation error at the transmitter, but also to reduce the required training duration overhead. Hence, the goals of this chapter are two-fold: (a) to analyze the DMT performance of a *channel-dependent* training scheme for acquiring CSIT and an associated power control mechanism for data transmission; and (b) to contrast the DMT performance of the proposed training and power control schemes with that achieved by conventional channel-agnostic training schemes. Our study focuses on point-to-point Single-Input

Multiple-Output (SIMO) systems. This is of practical importance, since it applies, for example, to the uplink of wireless networks where the base station has multiple antennas, the mobile users have a single antenna, and orthogonal access is used (e.g., OFDM/TDMA) as in WLANs and 4G/LTE systems. The channel-dependent training sequence employed here was first proposed by us in [30] and [10] in a MIMO and SIMO context, respectively, and was independently explored in [31], although not in a DMT context.

In this chapter, for analytical simplicity and clarity of presentation, we start by assuming that perfect CSI is available at the receiver, as in [22–24]. We propose a fixed-power RCT sequence, using which, the CSI can be estimated at the transmitter using a minimum duration of only one symbol, i.e., with a *factor of  $r$  reduction* in training duration compared to orthogonal RCT. For data transmission, we propose a modified truncated channel inversion-type power control scheme based on the noisy CSIT. For this system, we show that a diversity of  $d(g_m) = r \left( s + 1 - \frac{g_m L_c}{L_c - L_{B,\tau}} \right)$  is achievable. Here,  $g_m$  is the multiplexing gain,  $L_c$  is the coherence time,  $L_{B,\tau} \geq 1$  is the reverse training duration, and  $1 \leq s < r$  is a parameter in the data power control scheme. (See Section 2.3.)

Next, we consider the more practical case where noisy CSIR is acquired via a forward-link training sequence, and propose a *three-way* training scheme followed by data transmission. We show that a DMT of  $d(g_m) = r \left( s + 1 - \frac{g_m L_c}{L_c - \beta} \right)$  is achievable, where  $\beta \geq 3$  is the total training overhead from all three training phases, which is again an improvement over conventional orthogonal training schemes. For example, a nonzero diversity order can be achieved with  $\frac{L_c - (r+2)}{L_c} \leq g_m < \frac{L_c - 3}{L_c}$ , which is not possible with orthogonal training schemes without switching off receive antennas and incurring an associated

reduction in diversity order. (See Section 2.4.)

Note that although the perfect CSIR case is a special case of the three-way training scheme with infinite forward-link training power, we briefly present the perfect CSIR case also, as it provides insights into the impact of the reverse-training and data power control mechanisms on the DMT. Moreover, it is useful as an upper bound on the performance with imperfect CSIR. Also, we assume that power control is employed only at the transmitter and focus on fixed-power RCT in the sequel. Using power controlled RCT significantly changes the problem; we analyze this case in chapter 3 (also, see [12]).

An important implication of our work is that it shows that by exploiting the receiver's knowledge of the CSI in designing the *reverse channel training (RCT)* sequence and using our proposed data power control scheme, one can achieve a higher diversity order than conventional RCT for all values of  $g_m$ . Somewhat surprisingly, we also demonstrate that although the DMT analysis corresponds to taking the SNR to infinity, it can nonetheless be used to discriminate between different training schemes both in terms of the estimation error as well as the training overhead. At finite SNR, this translates to an improvement in the outage probability performance and the achievable data rate, as will be illustrated through Monte Carlo simulations in Section 2.6.

We use the following notation. Bold face letters are used for vectors and normal font letters are used for scalars. We write  $f(\bar{P}) \doteq \frac{1}{P^k}$  to mean  $-\lim_{\bar{P} \rightarrow \infty} \frac{\log f(\bar{P})}{\log P} = k$ . Similarly, we define  $f(\bar{P}) \preceq \frac{1}{P^k}$  to mean  $-\lim_{\bar{P} \rightarrow \infty} \frac{\log f(\bar{P})}{\log P} \geq k$ .

## 2.2 System Model

The system model consists of two communicating nodes, *node A* with a single antenna and *node B* with  $r$  antennas, with *node A* attempting to send data to *node B* over a wireless channel. The forward channel from *node A* to *node B*, denoted by  $\mathbf{h} \in \mathcal{C}^{r \times 1}$ , is modeled as a Rayleigh flat fading channel whose entries are *i.i.d.* Circularly Symmetric Complex Gaussian (CSCG) random variables with zero mean and unit variance, i.e.,  $\mathcal{CN}(0, 1)$ . The channel is assumed to be block-fading, i.e., it remains constant for a duration of the coherence time  $L_c$ , and evolve in an *i.i.d.* fashion across coherence times. We assume a TDD system with perfect reciprocity, and hence, taking the complex conjugate of the received signal at *node A*, the reverse-link channel is  $\mathbf{h}^H$ . We let  $\mathbf{h} = \sigma \mathbf{v}$ , where  $\sigma = \|\mathbf{h}\|_2$  is the singular value and  $\mathbf{v} \triangleq \frac{\mathbf{h}}{\|\mathbf{h}\|_2}$  is the singular vector of  $\mathbf{h}$ . Since our goal is to study the achievable DMT performance with channel training, we first explain the two-way training protocol used for acquiring CSI at *node B* and *node A*. Later, in Sec. 2.4, an additional phase of forward-link training is introduced, which is not presented here for simplicity of exposition.

### Phase I (Forward-link training)

Here, the training sequence  $x_{A,\tau} = \sqrt{PL_{A,\tau_1}}$  is transmitted from *node A* to *node B*, where  $L_{A,\tau_1}$  denotes the training duration and  $\bar{P}$  is the training power. Strictly speaking,  $x_{A,\tau} = \sqrt{\bar{P}}$  is transmitted repeatedly  $L_{A,\tau_1}$  times. Mathematically, this is equivalent to using  $x_{A,\tau} = \sqrt{PL_{A,\tau_1}}$  for a duration of one unit.. Throughout this chapter, we use  $\bar{P}$  as the average power constraint during both training and data transmission. The

corresponding received training signal is given by,

$$\mathbf{y}_{B,\tau} = \mathbf{h} \sqrt{\bar{P}L_{A,\tau_1}} + \mathbf{w}_{B,\tau}. \quad (2.1)$$

The entries of  $\mathbf{w}_{B,\tau} \in \mathcal{C}^{r \times 1}$  are assumed to be distributed as *i.i.d.*  $\mathcal{CN}(0, 1)$ . From the received training signal  $\mathbf{y}_{B,\tau}$ , *node B* computes an MMSE estimate of  $\mathbf{h}$ , denoted  $\hat{\mathbf{h}}$ . The error in the estimate, denoted  $\tilde{\mathbf{h}} \triangleq \mathbf{h} - \hat{\mathbf{h}}$ , has *i.i.d.*  $\mathcal{CN}(0, 1/(1 + \bar{P}L_{A,\tau_1}))$  distributed entries.

In a TDD-SIMO system, *node A* only requires knowledge of  $\sigma$  to perform power control, which in turn improves the diversity order compared to the no-CSIT case. Therefore, in phase II, we estimate only  $\sigma$  at *node A*, using a channel-dependent training sequence.

### Phase II (Reverse-link training)

Since *node B* has an estimate (say,  $\hat{\mathbf{v}} \triangleq \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}$ ) of the channel, in this phase, it exploits its CSI to transmit the following training sequence [10,30]:

$$\mathbf{x}_{B,\tau} = \sqrt{\bar{P}L_{B,\tau}} \hat{\mathbf{v}}, \quad (2.2)$$

where  $L_{B,\tau}$  is the reverse training duration. Using the corresponding received signal,  $y_{A,\tau} \triangleq \mathbf{h}^H \mathbf{x}_{B,\tau} + w_{A,\tau}$ , where the AWGN  $w_{A,\tau} \in \mathcal{C}$  is distributed as  $\mathcal{CN}(0, 1)$ , *node A* computes an estimate of the singular value as follows:

$$\hat{\sigma} \triangleq \frac{\Re\{y_{A,\tau}\}}{\sqrt{\bar{P}L_{B,\tau}}} = \sigma \Re\{\mathbf{v}^H \hat{\mathbf{v}}\} + \bar{w}_{A,\tau}, \quad (2.3)$$

where  $\bar{w}_{A,\tau} \triangleq \frac{\Re\{w_{A,\tau}\}}{\sqrt{P}L_{B,\tau}}$ . Note that the estimate  $\hat{\sigma}$  could be negative; this is taken care of by the power control proposed in Sec. 2.3, which uses  $\hat{\sigma}$  only when it is greater than a positive threshold. Since a low or negative  $\hat{\sigma}$  is likely to be inaccurate, the thresholding technique helps to avoid the poor DMT performance due to such estimates. The RCT scheme employed above is different from existing channel-agnostic methods in that the minimum training length in the proposed scheme is only 1 symbol. This represents a factor of  $r$  reduction compared to orthogonal RCT schemes, where the minimum training length increases linearly with  $r$ , and this difference in overhead could be significant when  $L_c$  is small. Also, if  $\hat{\mathbf{v}}$  is error-free, it is the optimal beamforming vector for estimating  $\sigma$  at *node A*.

### Multiplexing Gain and Diversity Order

We recall the definitions of the multiplexing gain,  $g_m$ , and the diversity order  $d$  from [16]:

$$g_m \triangleq \lim_{\bar{P} \rightarrow \infty} \frac{R_{\bar{P}}}{\log \bar{P}}, \quad d \triangleq - \lim_{\bar{P} \rightarrow \infty} \frac{\log P_{out}}{\log \bar{P}}, \quad (2.4)$$

where  $R_{\bar{P}}$  is the target data rate when the average data power constraint is  $\bar{P}$ , and  $P_{out}$  is the corresponding outage probability, i.e., the probability that  $R_{\bar{P}}$  exceeds the channel capacity. In this work, the target data rate  $R_{\bar{P}} = g_m \log \bar{P}$  is fixed and is independent of the CSIT; the extension of our proposed methods to joint rate and power adaptation is relegated to future work. The rate of data transmission  $R_{\bar{P}}$  is increased with  $\bar{P}$  by increasing the cardinality of the signal set, keeping the symbol duration fixed. We ignore the effect of spectral leakage, and assume that the signal bandwidth remains fixed as  $\bar{P}$  goes to infinity. Also, we use outage probability as a proxy for the probability of

error at high SNR with finite-length codes; this is because the probability of error can be made to decrease as fast as the outage probability using finite-length approximately universal codes [32,33].

In the next section, we assume perfect CSI at *node B* and derive the achievable DMT performance of our proposed training and data transmission schemes.

## 2.3 DMT Analysis with Perfect CSIR

When the CSIR is perfect, we have  $\hat{\mathbf{v}} = \mathbf{v}$ , and in this case, it is easy to see that (2.2) is optimal for estimating  $\sigma$  given a power constraint  $\bar{P}$  on the training signal. This is because, in general, the training signal can be expressed as the linear combination  $\mathbf{x}_{B,\tau} = \delta\mathbf{v} + \beta\mathbf{v}_\perp$ , where  $\mathbf{v}_\perp$  is orthogonal to  $\mathbf{v}$  and  $\delta$  and  $\beta$  are some constants. Then, the received training signal at *node A* is  $y_{A,\tau} = \delta\sigma + w_{A,\tau}$ , i.e., the power in  $\mathbf{v}_\perp$  does not help in estimating  $\sigma$ . From (2.3), an *unbiased* estimator of the singular value at *node A* is given by

$$\hat{\sigma} = \sigma + \bar{w}_{A,\tau}. \quad (2.5)$$

Note that since the channel is assumed to be Rayleigh fading,  $\sigma^2$  is chi-square distributed with  $2r$  degrees of freedom. Also, we employ this estimator primarily because we are interested in deriving the achievable DMT performance, and for this purpose, this simple unbiased estimator is sufficient.

### 2.3.1 Power-Controlled Data Transmission from *Node A* to *Node B*

Given the CSIT  $\hat{\sigma}$  in (2.5), *node A* uses a power  $\mathcal{P}(\hat{\sigma})$  in the forward-link *data transmission* phase, to avoid outages while satisfying the average data power constraint  $\bar{P}$ . The

corresponding *data signal* received at *node B* is given by,

$$\mathbf{y}_{B,d} = \sqrt{\mathcal{P}(\hat{\sigma})}\mathbf{h}x_{A,d} + \mathbf{w}_{B,d}, \quad (2.6)$$

where  $x_{A,d} \sim \mathcal{CN}(0, 1)$ , and with appropriate power normalization, the entries of the AWGN  $\mathbf{w}_{B,d} \in \mathcal{C}^{r \times 1}$  are assumed to be *i.i.d.*  $\mathcal{CN}(0, 1)$ . Also,  $\mathcal{P}(\hat{\sigma})$  is chosen independent of  $x_{A,d}$  such that  $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = \bar{P}$ , where the expectation is with respect to  $\hat{\sigma}$  given in (2.5), taken across all coherence blocks. Since  $\mathbb{E}\{|x_{A,d}|^2\} = 1$  within a block, this ensures that the average data power constraint at *node A* is satisfied.

We now present the data power control function  $\mathcal{P}(\hat{\sigma})$  considered in this chapter. Our proposed power control function is motivated as follows. The capacity of a fading channel with *mismatched* CSIT and CSIR is not known in closed form [34]. Since the outage probability computation requires a closed form expression for the capacity, we consider a genie-aided receiver as in [3], where *node B* is assumed to know  $\mathcal{P}(\hat{\sigma})$ . This is schematically illustrated in Fig. 2.1. Then, the achievable data rate conditioned on the knowledge of  $\sqrt{\mathcal{P}(\hat{\sigma})}\mathbf{h}$  is given by [34]

$$C \triangleq \frac{L_c - L_{B,\tau}}{L_c} \log(1 + \sigma^2 \mathcal{P}(\hat{\sigma})). \quad (2.7)$$

An outage occurs when  $R_{\bar{P}}$ , the target data rate, exceeds  $C$ . Its probability is upper bounded by

$$P_{out} \triangleq \Pr\left(\frac{L_c - L_{B,\tau}}{L_c} \log(1 + \sigma^2 \mathcal{P}(\hat{\sigma})) < R_{\bar{P}}\right). \quad (2.8)$$

Note that the exact outage probability is obtained by minimizing the right hand side above over all  $\mathcal{P}(\hat{\sigma})$  satisfying  $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = \bar{P}$ . Hence, using our proposed data power control scheme leads to an upper bound on the outage probability, which is sufficient

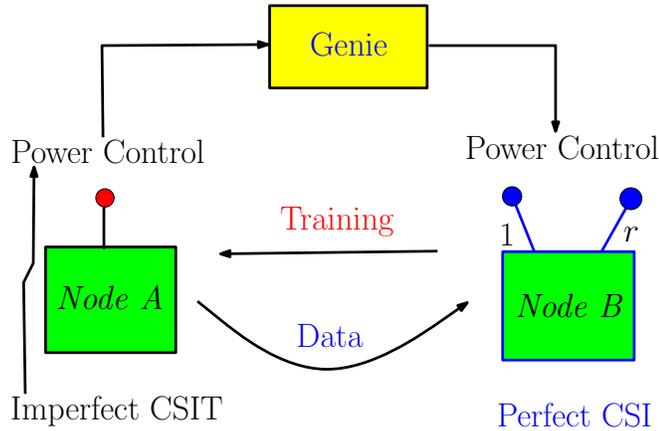


Figure 2.1: System model for reverse channel training with perfect CSIR used in Section 2.3.

for obtaining the achievable DMT performance. If the CSIT is perfect (i.e.,  $\hat{\sigma}^2 = \sigma^2$ ), it is shown in [17] that the power control that minimizes the outage probability is given by

$$\Phi(\sigma^2) \triangleq \frac{\exp\left(\frac{RL_c}{L_c - L_{B,\tau}}\right) - 1}{\sigma^2}. \quad (2.9)$$

Note that since  $R_{\bar{P}} = g_m \log \bar{P}$  and  $\mathbb{E}\left\{\frac{1}{\sigma^2}\right\} = \frac{1}{r-1}$ ,  $\Phi(\sigma^2)$  satisfies  $\mathbb{E}\{\Phi(\sigma^2)\} \leq \bar{P}$  for large enough  $\bar{P}$ , provided  $g_m \leq (L_c - L_{B,\tau})/L_c$ . With inaccurate CSIT, due to the estimation error in  $\hat{\sigma}$ , the natural extension of using a transmission power of  $\Phi(\hat{\sigma}^2)$  could result in allocating insufficient power or more power than required, which could lead to suboptimal performance. Also, inverting the channel for all values of  $\hat{\sigma}$  results in an infinite average power since the Gaussian noise can make the estimate  $\hat{\sigma}$  arbitrarily small with a non-zero probability. One solution is to use a transmit power of  $\Phi(\hat{\sigma}^2)$  when  $\hat{\sigma} > \theta_0$  and a zero power otherwise, where  $\theta_0$  is chosen such that  $\mathbb{E}[\Phi(\hat{\sigma}^2)1_{\hat{\sigma} > \theta_0}] = \bar{P}$ . The drawback of this method is that it results in an outage probability of 1 when  $\hat{\sigma} \leq \theta_0$ , leading to a zero diversity order. To overcome this problem, we choose the threshold  $\theta_0$  such

that  $\theta_0 \rightarrow 0$  as  $\bar{P} \rightarrow \infty$ . Moreover, when  $\hat{\sigma} \leq \theta_0$ , we do not necessarily want to use zero power, since the small value of  $\hat{\sigma}$  could be due to the estimation error. This motivates the following modified power control:

$$\mathcal{P}(\hat{\sigma}) \triangleq \begin{cases} \bar{P}^l & \hat{\sigma} \leq \theta_{\bar{P}}, \\ \kappa_{\bar{P}} \times \Phi(\hat{\sigma}^{2s}) & \hat{\sigma} > \theta_{\bar{P}}, \end{cases} \quad (2.10)$$

where  $s \geq 1$  is a parameter, and we use  $\theta_{\bar{P}} \triangleq \frac{1}{\bar{P}^n}$ ,  $n > 0$ , for mathematical tractability. The parameters  $n$ ,  $\kappa_{\bar{P}}$  and  $l > 0$  are chosen such that  $\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \bar{P}$ . Although similar power control schemes have been employed in the literature with perfect CSIT [17] or orthogonal RCT [1, 2, 24], the form in (2.10) is new. Specifically, the power control scheme in [1, 17, 24] can be obtained from (2.10) by setting  $s = 1$ ,  $\theta_{\bar{P}} = 0$  and  $l = -\infty$ ; while that in [2] can be obtained by setting  $s = r$ ,  $\theta_{\bar{P}} = 0$  and  $l = -\infty$ .

### Power constraint

The description of the power control would be complete if the parameters  $n$ ,  $\kappa_{\bar{P}}$  and  $l$  can be chosen such that  $\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \bar{P}$ , which is the essence of the following Lemma.

**Lemma 1.** *Let  $\theta_{\bar{P}} \triangleq \frac{1}{\sqrt{\bar{P}}}$ . For  $1 \leq s < r$ , there exists a  $\kappa_{\bar{P}} \doteq \frac{1}{\bar{P}^{\frac{\alpha}{2s}-1}}$ , where  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$ , such that  $\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \bar{P}$ , if  $0 \leq l \leq r + 1$ .*

*Proof:* See Appendix A.0.2. ■

Due to Lemma 1, in the rest of this chapter, we consider  $\theta_{\bar{P}} = 1/\sqrt{\bar{P}}$ . Also, in Sec. 2.4, we show that a minor modification of the above data power control scheme can be employed even with imperfect CSIR. The next subsection presents the achievable DMT of the proposed training and power control schemes.

### 2.3.2 Achievable DMT Analysis

**Theorem 1.** Given  $r$  receive antennas and  $L_{B,\tau}$  training symbols being used per coherence interval  $L_c$  to estimate the CSIT in a SIMO system with perfect CSIR and a genie-aided receiver, an achievable diversity order as a function of multiplexing gain  $g_m$  is given by

$$d(g_m) = r \left( \min\{l, s + 1\} - \frac{g_m}{\alpha} \right), \quad (2.11)$$

where  $0 \leq l \leq r + 1$ ,  $1 \leq s < r$ ,  $0 \leq g_m < \alpha$ , and  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$  represents the fractional data transmit duration.

*Proof:* See Appendix A.0.3. ■

**Remark:** From a DMT perspective, it is clear from Theorem 1 that  $s \rightarrow r, l = r + 1$  is superior to  $s = 1, l = 2$ . On the other hand, when  $\hat{\sigma} < 1$ ,  $\Phi(\hat{\sigma}^{2r})$  could be much greater than  $\Phi(\hat{\sigma}^2)$ . Thus, in practical systems with a peak power per transmitted codeword constraint,  $s = 1, l = 2$  could be preferable over  $s \rightarrow r, l = r + 1$ . In the sequel, for convenience, we associate  $l = 2$  with  $s = 1$  and  $l = r + 1$  with  $s \rightarrow r$ , and drop the explicit dependence of the diversity order on  $l$ . Further remarks and discussions on the result obtained here are deferred to Sec. 2.5.

Table 2.1: Three way training in a TDD-SIMO system

Phase	Description	Input-Output Equation
I	Fixed power training (Node A $\rightarrow$ Node B)	$\mathbf{y}_{B,\tau} = \mathbf{h}x_{A,\tau} + \mathbf{w}_{B,\tau}$
II	Fixed power training (Node B $\rightarrow$ Node A)	$y_{A,\tau} = \mathbf{h}^H \mathbf{x}_{B,\tau} + w_{A,\tau}$
III	Power controlled training (Node A $\rightarrow$ Node B)	$\mathbf{y}_{B,\tau_2} = \sqrt{PL_{A,\tau_2}} \mathcal{P}(\hat{\sigma}) \mathbf{h} + \mathbf{w}_{B,\tau_2}$
IV	Power controlled data (Node A $\rightarrow$ Node B)	$\mathbf{y}_{B,d} = \mathbf{h}x_{A,d} + \mathbf{w}_{B,d}$

## 2.4 Three Way Training

In this section, we consider the more practical scenario where training is performed in both directions. We show that with fixed power training, one can achieve nearly the same DMT as derived in Sec. 2.3 for the perfect CSIR case. Unlike in the previous section, the analysis presented here is exact, in the sense that it does not require the assumption of a genie aided receiver, and hence, the DMT derived here is indeed achievable in practice. The transmission protocol now consists of four phases, as shown in Table 2.1. The CSIR and CSIT are obtained by transmitting a fixed power training sequence in both directions, as explained in Sec. 2.2. However, even a small mismatch in the CSI knowledge at *node A* and *node B* can potentially lead to a large mismatch in their estimate of the data transmit power [1]. Thus, it is essential to train *node B* about *node A*'s knowledge of  $\mathcal{P}(\hat{\sigma})$ . This leads to a third phase of training, which is an additional power-controlled forward-link training phase. First, in the following subsection, we explain the power control scheme that is employed here.

### 2.4.1 Power Control Scheme

The power control scheme we propose to employ in this section is as given by (2.10), due to the following. Let  $\hat{\mathbf{h}}$  denote the MMSE estimate of the channel at *node B*, and consider  $\hat{\sigma}$  in (2.3). We have

$$\begin{aligned}\hat{\sigma} &\triangleq \frac{\Re\{y_{A,\tau}\}}{\sqrt{PL_{B,\tau}}} = \Re\{\hat{\mathbf{h}}^H \hat{\mathbf{v}}\} + \Re\{\tilde{\mathbf{h}}^H \hat{\mathbf{v}}\} + \frac{\Re\{w_{A,\tau}\}}{\sqrt{PL_{B,\tau}}} \\ &= \|\hat{\mathbf{h}}\|_2 + \tilde{w}_{eff},\end{aligned}\tag{2.12}$$

where  $\tilde{w}_{eff} \triangleq \Re\{\tilde{\mathbf{h}}^H \hat{\mathbf{v}}\} + \frac{\Re\{w_{A,\tau}\}}{\sqrt{PL_{B,\tau}}}$ . Note that  $\hat{\mathbf{h}}$  and  $\tilde{\mathbf{h}}$  are independent Gaussian random variables.<sup>1</sup> Since  $\hat{\mathbf{v}}$  is uniformly distributed on the unit sphere and is independent of  $\tilde{\mathbf{h}}$ ,  $\Re\{\tilde{\mathbf{h}}^H \hat{\mathbf{v}}\}$  is Gaussian distributed. This implies that the effective noise,  $\tilde{w}_{eff}$ , is Gaussian distributed with  $\mathbb{E}|\tilde{w}_{eff}|^2 \doteq \frac{1}{\bar{P}}$  and independent of  $\hat{\mathbf{h}}$ . Therefore, the estimate of the singular value at *node A* is statistically similar to the estimate given by (2.5) in the perfect CSIR case. Thus, we use a similar power control,  $\mathcal{P}(\hat{\sigma})$  in (2.10), where  $\hat{\sigma}$  is given by (2.12). Also, with a slight abuse of notation,  $\alpha \triangleq \frac{L_c - L_{B,\tau} - L_{A,\tau_1} - L_{A,\tau_2}}{L_c}$ , where  $L_{A,\tau_2}$  is the training duration in the third phase of training (phase III), which is in the forward-link direction.

In this section, without loss of generality, we move the power scaling  $\sqrt{\bar{P}}$  into the data symbol transmitted by *node A*, so that  $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = 1$  (see (2.13) below), where the expectation is taken with respect to the distribution of  $\hat{\sigma}$  in (2.12). Now, in the proof of Lemma 1, using the probability density function (pdf) of  $\|\hat{\mathbf{h}}\|_2$  in place of the pdf of  $\sigma$ , and noting that the effective noise variance  $\doteq 1/\bar{P}$ , we get  $\kappa_{\bar{P}} \doteq \frac{1}{P_{gm}/\alpha}$  and the constraint  $0 \leq l \leq r$  to satisfy  $\mathbb{E}\{\mathcal{P}(\hat{\sigma})\} = 1$  at high SNR. In the next subsection, we explain the third round of training that alleviates the mismatch in the knowledge of the data transmit power.

## 2.4.2 Power-Controlled Forward Link Training Scheme

In phase III of the scheme, *node A* transmits the training sequence:  $x_{A,\tau_2} = \sqrt{PL_{A,\tau_2}}\sqrt{\mathcal{P}(\hat{\sigma})}$ , where  $L_{A,\tau_2}$  is the training duration. The corresponding received training signal at *node*

---

<sup>1</sup> $\hat{\mathbf{h}} \rightarrow \mathbf{h}$  as  $\bar{P} \rightarrow \infty$ . Moreover,  $\|\hat{\mathbf{h}}\|_2$  is a chi distributed random variable.

$B$  is given by,

$$\mathbf{y}_{B,\tau_2} = \sqrt{\bar{P}L_{A,\tau_2}} \sqrt{\mathcal{P}(\hat{\sigma})} \mathbf{h} + \mathbf{w}_{B,\tau_2}, \quad (2.13)$$

where  $\mathbf{w}_{B,\tau_2} \in \mathcal{C}^{r \times 1}$  is the AWGN with  $\mathcal{CN}(0, 1)$  entries. The goal at *node B* is to estimate the composite channel  $\mathbf{p}_c \triangleq \sqrt{\mathcal{P}(\hat{\sigma})} \mathbf{h}$ . Dividing (2.13) by  $\sqrt{\bar{P}L_{A,\tau_2}}$ , we get

$$\tilde{\mathbf{y}}_{B,\tau_2} \triangleq \frac{\mathbf{y}_{B,\tau_2}}{\sqrt{\bar{P}L_{A,\tau_2}}} = \mathbf{p}_c + \frac{\mathbf{w}_{B,\tau_2}}{\sqrt{\bar{P}L_{A,\tau_2}}}. \quad (2.14)$$

From (2.14), *node B* computes an MMSE estimate of  $\mathbf{p}_c$ , denoted by  $\hat{\mathbf{p}}_c$ . Let  $\tilde{\mathbf{p}}_c \triangleq \mathbf{p}_c - \hat{\mathbf{p}}_c$ . Although a closed form expression for  $\hat{\mathbf{p}}_c$  is hard to find, the error  $\tilde{\mathbf{p}}_c$  in the MMSE estimate has the following interesting property, which facilitates the calculation of the outage probability in Sec. 2.4.4. An analogous result has been shown in [35] for the scalar case.

**Lemma 2.**  $\mathbb{E} \|\tilde{\mathbf{p}}_c\|_2^{2z} \preceq \frac{1}{\bar{P}^z}$  for every  $z > 0$ .

*Proof:* See Appendix A.0.4. ■

### 2.4.3 Power-Controlled Data Transmission from *Node A* to *Node B*

In phase IV. of the scheme, using  $\mathcal{P}(\hat{\sigma})$ , *node A* sends the data signal  $x = \sqrt{\bar{P}\mathcal{P}(\hat{\sigma})} x_{A,d}$ , where  $x_{A,d}$  is distributed as  $\mathcal{CN}(0, 1)$  and is independent of  $\mathcal{P}(\hat{\sigma})$ . Note that  $\mathbb{E}|x|^2 = \bar{P}$  by construction, where the expectation is taken with respect to both  $\hat{\sigma}$  and  $x_{A,d}$ . The corresponding signal received at *node B* is

$$\mathbf{y}_{B,d} = \sqrt{\bar{P}\mathcal{P}(\hat{\sigma})} \mathbf{h} x_{A,d} + \mathbf{w}_{B,d} \quad (2.15)$$

$$= \sqrt{\bar{P}} \hat{\mathbf{p}}_c x_{A,d} + \sqrt{\bar{P}} \tilde{\mathbf{p}}_c x_{A,d} + \mathbf{w}_{B,d}. \quad (2.16)$$

Since  $\hat{\mathbf{p}}_c$  is an MMSE estimate, using the worst case noise theorem [8], we have the following lower bound on the mutual information,  $I(x_{A,d}; \mathbf{y}_{B,d} | \hat{\mathbf{p}}_c) \geq C_{AB}$ , where

$$C_{AB} \triangleq \alpha \log \left( 1 + \frac{\bar{P} \|\hat{\mathbf{p}}_c\|_2^2}{\frac{\bar{P}}{r} \mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] + 1} \right), \quad (2.17)$$

and  $\alpha \triangleq \frac{L_c - L_{B,\tau} - L_{A,\tau_1} - L_{A,\tau_2}}{L_c}$  is the fractional data transmit duration after accounting for the time overheads in all three training phases.

#### 2.4.4 Achievable DMT Analysis

**Theorem 2.** *For a SIMO system with  $r$  receive antennas and three phases of training and the data transmission phase as described in Table 2.1, an achievable DMT is given by*

$$d(g_m) = r \left( \min\{l, s\} + 1 - \frac{g_m}{\alpha} \right), \quad (2.18)$$

where  $0 \leq l \leq r$ ,  $1 \leq s < r$ ,  $0 \leq g_m < \alpha$ , and  $\alpha \triangleq \frac{L_c - L_{B,\tau} - L_{A,\tau_1} - L_{A,\tau_2}}{L_c}$ .

*Proof:* See Appendix A.0.5. ■

**Remark:** The above three way training scheme can be generalized to  $k$  training rounds to improve the diversity order, as in [2, 28]. However, this is mathematically cumbersome and out of the scope of our work.

## 2.5 Discussion

Recall that with perfect CSIR and imperfect CSIT, with  $l \geq s + 1$ , and for a genie aided channel, it was shown in Theorem 1 that the following DMT is achievable:

$$d(g_m) = r \left[ s + 1 - \left( \frac{g_m L_c}{L_c - L_{B,\tau}} \right) \right], \quad (2.19)$$

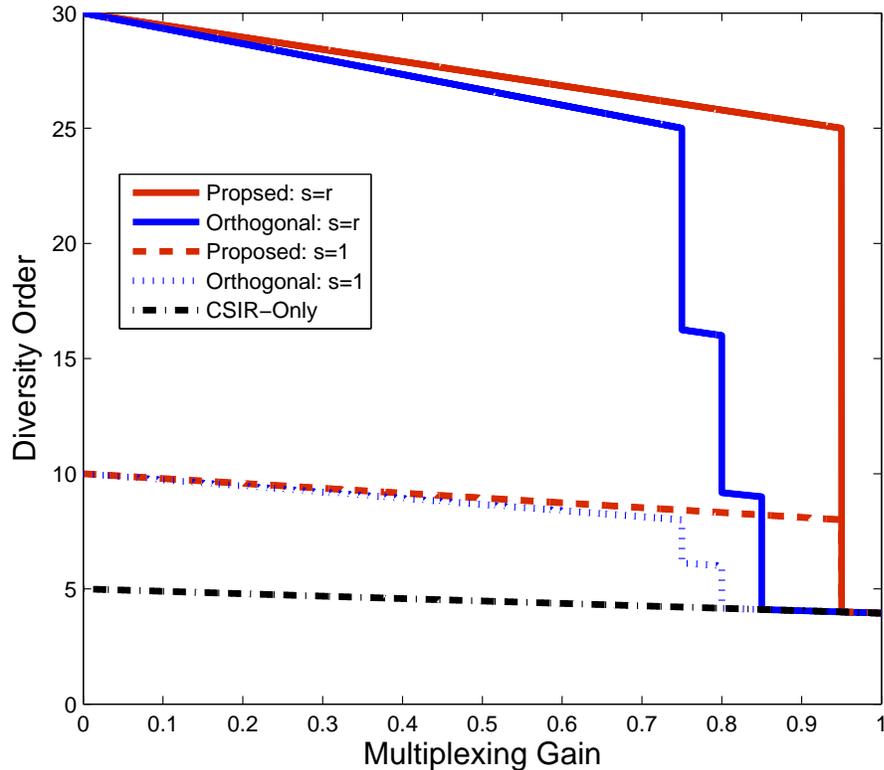


Figure 2.2: The achievable DMT with the training and power control scheme proposed in Sec. 2.3, compared with the performance of the orthogonal RCT and the data power control proposed in [1,2] (and appropriately accounting for the training duration overhead and switching off antennas to achieve higher values of  $g_m$ ). The plot corresponds to a SIMO system with  $r = 5$  antennas, with coherence time  $L_c = 20$  symbols, and reverse training duration of  $L_{B,\tau} = 1$  symbol.

where  $1 \leq s < r$ ,  $0 \leq g_m \leq \frac{L_c - L_{B,\tau}}{L_c}$ . In contrast, for the same genie aided channel, it was shown in [3] that a diversity order of

$$d_s(g_m) = r \left[ 2 - \left( \frac{g_m L_c}{L_c - r L_{B,\tau}} \right) \right], \quad 0 \leq g_m \leq \frac{L_c - r L_{B,\tau}}{L_c} \quad (2.20)$$

is achievable using orthogonal reverse channel training. Note that  $d_s(g_m)$  saturates as  $r$  gets large, as opposed to (2.19), which is monotonically increasing in  $r$ . In order to

achieve a  $g_m > \frac{L_c - rL_{B,\tau}}{L_c}$ , in [1], the authors suggest turning off one receive antenna at a time to reduce the training burden until  $r = 2$ . For example, turning off one antenna,  $g_m \in \left[ \frac{L_c - rL_{B,\tau}}{L_c}, \frac{L_c - (r-1)L_{B,\tau}}{L_c} \right]$  is achievable at a reduced diversity order of  $d_s(g_m) = (r - 1) \left[ 2 - \left( \frac{g_m L_c}{L_c - (r-1)L_{B,\tau}} \right) \right]$ . This is in contrast to our result, which can accommodate a larger multiplexing gain,  $g_m \leq \frac{L_c - L_{B,\tau}}{L_c}$  irrespective of  $r$ , while simultaneously achieving a higher diversity order at each  $g_m$ . We note that for a SIMO channel, a diversity order of  $r(r+1-g_m)$  for  $0 \leq g_m < 1$  was obtained in [2,24], using channel-independent training, and without accounting for the training duration overhead. This corresponds to taking  $L_c \rightarrow \infty$  in (2.19). The performance of the proposed scheme is schematically contrasted with orthogonal RCT in Fig. 2.2 for a SIMO system with  $r = 5$ ,  $L_c = 20$ , and  $L_{B,\tau} = 1$  symbol. The advantage of the proposed scheme at higher values of the multiplexing gain is clear from the plot. The proposed training scheme thus results in a *factor  $r$ -reduction* in the training duration, which, along with the proposed data power control scheme, translates to an increase in the range of achievable multiplexing gains, while simultaneously offering a better diversity order compared to orthogonal RCT schemes.

Comparing Theorems 1 and 2, we see that the DMT performance of a genie aided receiver with perfect CSIR is an upper bound on the performance of the system with imperfect CSIR and CSIT, as expected. Also, the performance of the two systems is similar, except that in the latter case, the factor  $\alpha$  captures the loss in data transmission time due to all three training phases. Similar observations as the above regarding the improvement in DMT can be made for the three way training scheme compared to orthogonal RCT schemes.

## 2.6 Simulation Results

We now briefly present Monte Carlo simulation results to illustrate the outage probability performance of our proposed RCT and forward-link data power control schemes. We consider a Rayleigh fading channel with three receive antennas. We calculate the outage probability by averaging over  $10^8$  *i.i.d.* channel and training noise instantiations. We set the channel coherence time and reverse training duration as  $L_c = 40$  and  $L_{B,\tau} = 1$ , respectively. Figures 2.3 and 2.4 show the outage probability of the proposed fixed-power training scheme and the data power control scheme in (2.10) with  $s = 1$  and  $s = r = 3$ , respectively, as a function of  $\bar{P}$ , with  $g_m = 0$  and  $R_{\bar{P}} = 4$  bits/channel use (1 and 1.5 bits/channel use in case of Fig. 2.4), and with  $g_m = 0.8$ . Although the slopes of the curves do not match with the theoretical diversity order because the latter requires infinite SNR, the improved performance of the proposed schemes is clear from the graphs. Also, in Fig. 2.3, since the proposed scheme uses only  $L_{B,\tau} = 1$  training symbol while the orthogonal RCT scheme uses  $rL_{B,\tau} = 3$  training symbols, the former shows a higher outage than the latter at lower SNRs. Note that, we have not plotted the outage performance of the three-way training scheme in Sec. 2.4. This is because the outage probability is hard to compute, since a closed-form expression for  $\hat{p}_c$  is not available.

## 2.7 Conclusions

This chapter proposed reverse training and data power control schemes for a TDD-SIMO system with perfect/imperfect CSIR and investigated its DMT performance. It was shown that a diversity order of  $d(g_m) = r \left( s + 1 - \frac{g_m}{\alpha} \right)$  is achievable for  $l \geq s + 1$ ,

$1 \leq s < r$  and  $0 \leq g_m < \alpha$ , where  $\alpha$  represents the fractional data transmit duration. In contrast to channel-agnostic orthogonal training schemes, the diversity order was shown to increase monotonically with  $r$  at nonzero multiplexing gain, which is a significant improvement. The DMT analysis was extended to a more practical situation where the training is done in both directions. In this case also, it was shown that the DMT performance can improve quadratically with the number of receive antennas, and nearly the same DMT can be achieved as that with perfect CSIR and a genie-aided receiver. In terms of system design for reciprocal SIMO systems, the key messages from this chapter are that it is important to (a) exploit the CSI at the receiver in designing the RCT and (b) use a modified channel-inversion type power control scheme that transmits data at some non-zero power even when the estimated singular value at the transmitter is poor. For fast varying channels, these ingredients can lead to a significant advantage in DMT performance, which, at finite SNR, can translate to a large improvement in outage probability performance compared to orthogonal training schemes. In the next chapter, we propose a power-controlled RCT scheme and analyze its DMT performance. We show that our proposed power-controlled RCT can achieve an infinite diversity order.

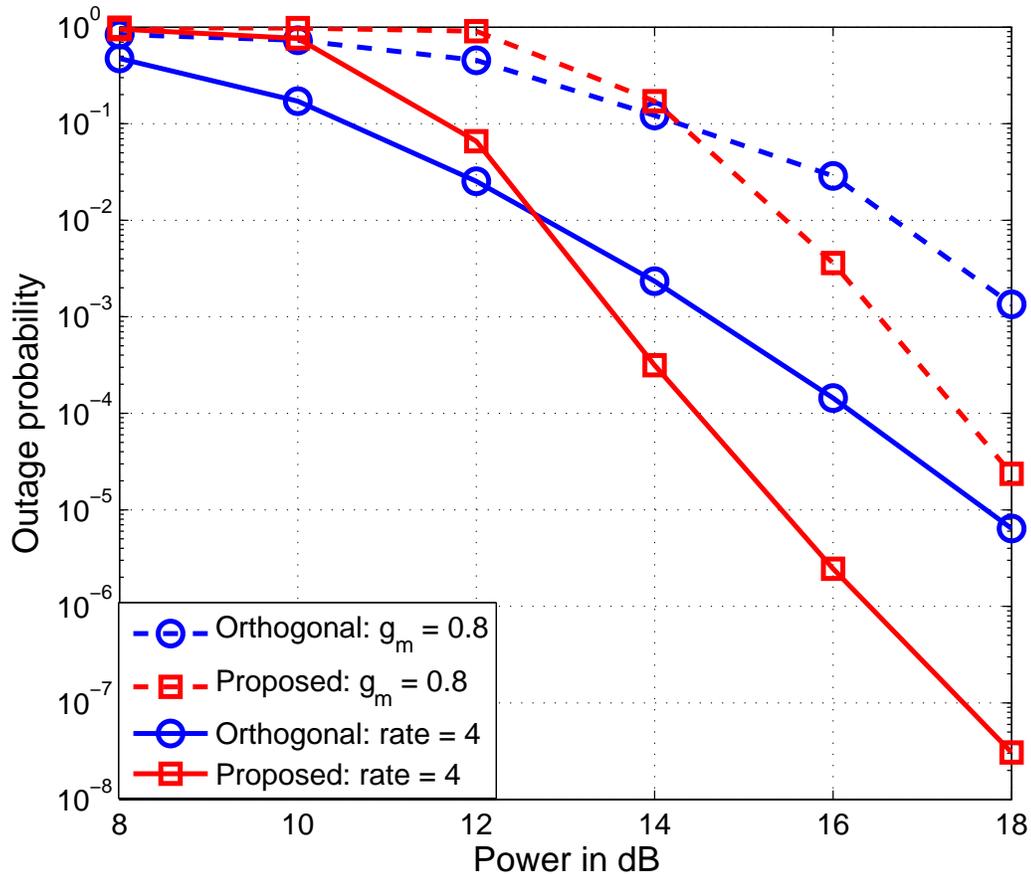


Figure 2.3: Outage probability versus the average data power  $\bar{P}$  for the fixed-power training scheme proposed in Sec. 2.3, with the data power control scheme given by (2.10) with  $s = 1$ . Here,  $r = 3$ ,  $L_c = 40$  and  $L_{B,\tau} = 1$ . With  $g_m = 0.8$ , the target data rate was set as  $R_{\bar{P}} = 4 + g_m \log \bar{P}$  to facilitate the comparison of the curves.

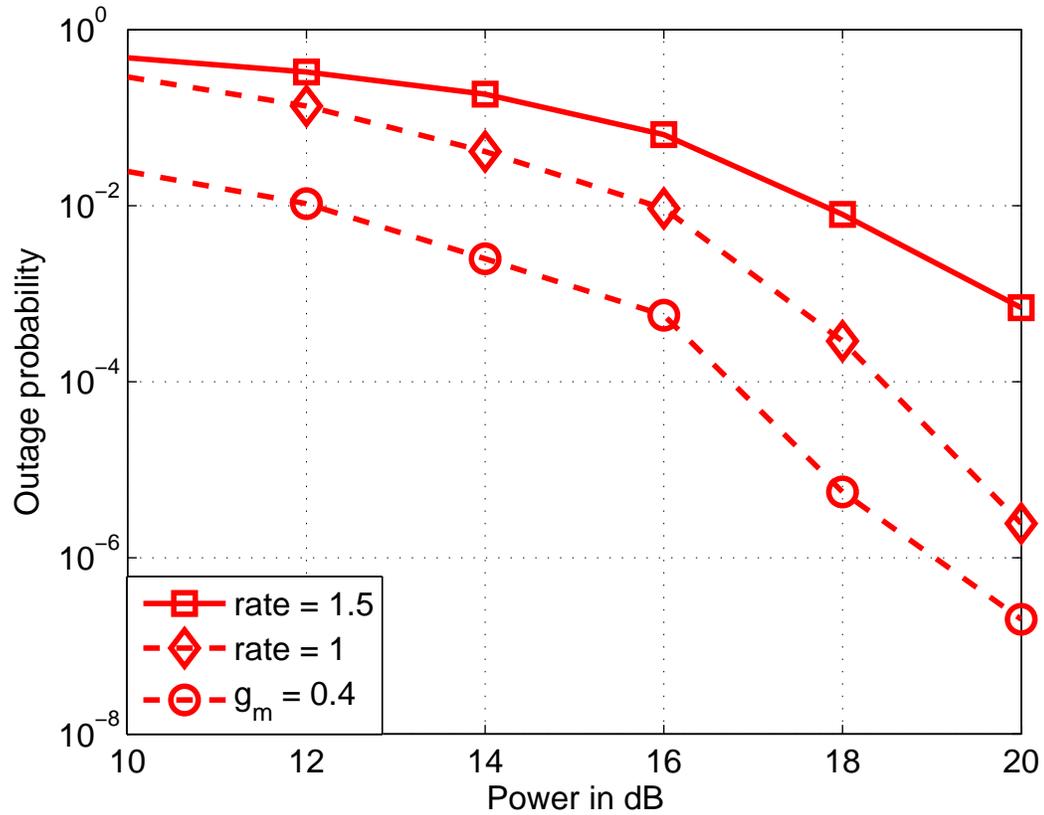


Figure 2.4: Outage probability versus the average data power  $\bar{P}$  for the fixed-power training scheme proposed in Sec. 2.3, with the data power control scheme given by (2.10) with  $s = r$ . Here,  $r = 3$ ,  $L_c = 40$  and  $L_{B,\tau} = 1$ .

# Chapter 3

## Power Controlled Reverse Channel Training Achieves an Infinite Diversity Order in a TDD-SIMO System with Perfect CSIR

### 3.1 Introduction

As mentioned earlier, the diversity multiplexing gain trade-off (DMT), first proposed in [16], elegantly captures the tradeoff between throughput and reliability in a wireless communication system at high SNR. While [16] assumed perfect Channel State Information at the Receiver (CSIR), the DMT analysis has been extended to imperfect CSIR and no CSI at the Transmitter (CSIT) [21], perfect CSIT and CSIR [9], noisy CSIT and perfect CSIR [1–3, 24]. In the studies with noisy CSIT, the error in the CSI is modeled as an independent Gaussian noise matrix whose covariance approaches zero as the training power is increased. This is statistically equivalent to the CSIT estimate in a reciprocal Time Division Duplex (TDD) system acquired using a *channel-agnostic*

*fixed-power orthogonal* Reverse Channel Training (RCT) sequence. In the presence of CSIR acquired through an initial forward-link training phase, channel-agnostic RCT has the disadvantage that it does not exploit the available CSIR as side information in designing the RCT sequence. For example, in a Single-Input Multiple-Output (SIMO) system, using a normalized version of the estimated channel as a beamforming vector to transmit a fixed-power RCT sequence enables the transmitter to directly estimate the norm of the channel, which can then be used for forward-link data power control. A DMT performance advantage results from obviating the need to first estimate the channel vector and then estimate the channel norm and data power control from the channel estimate [11]. However, one can potentially achieve an even better DMT by using *power-controlled* RCT, where the training sequence is designed to satisfy an *average* power constraint, rather than an instantaneous power constraint, and this is the subject of study in this chapter.

In this chapter, we consider a SIMO system consisting of *node A* with a single antenna and *node B* with  $r$  antennas, and with data transmission from *node A* to *node B* using an average power of  $\bar{P}$  per channel use. We focus on the DMT of a reciprocal TDD SIMO system, with perfect CSI at *Node B* (CSIB) as in [24]. Our contributions are:

- We propose a channel-dependent Power Controlled RCT (PCRCT) scheme with average power  $\bar{P}^\gamma$ ,  $\gamma > 0$ , that enables *node A* to directly estimate the power control required for data transmission at the target rate.
- For  $r > 2$ , we show that an *infinite* diversity order is achievable using the proposed PCRCT for  $0 \leq g_m < \frac{L_c - L_{B,\tau}}{L_c} \min(\gamma, 1)$ , where  $g_m$  is the multiplexing gain,  $L_{B,\tau}$  is the RCT duration, and  $L_c$  is the coherence time of the channel.

- We also show that as  $\bar{P}$  gets large, an upper bound on the outage probability goes to zero asymptotically as  $\exp(-\bar{P}^E)$ , where  $E \triangleq \left(\gamma - \frac{g_m L_c}{L_c - L_{B,\tau}}\right)$ .
- We show that, for the proposed PCRCT and data power control, there exists an uncoded data transmission scheme for which the probability of error goes to zero exponentially with  $\bar{P}$ .

An advantage of our proposed method is that it requires a minimum training duration of only one symbol, unlike channel-agnostic training which requires at least  $r$  symbols. This could be significant when the channel coherence time is low, i.e., for fast-varying channels. Our infinite diversity order result is in contrast with most existing results, that obtain only a finite diversity order with imperfect CSIT, even with perfect CSIR. In [28, 36], an infinite diversity order result is shown using *digital* feedback over the reverse-link, when the number of feedback bits is increased with the reverse training power. Unlike [28], our scheme uses a beamformed PCRCT that leads to direct estimation of the data power control at *node A*, due to which, our analytical development and the corresponding DMT result are different.

Our outage exponent result shows that the rate at which the diversity order goes to infinity is faster for smaller values of  $g_m$ . Moreover, the larger the value of  $\gamma > 1$  (i.e., using higher RCT power than the data transmission power), the larger the outage exponent. The RCT scheme proposed in this chapter can thus significantly improve the DMT performance compared to channel-agnostic orthogonal RCT schemes.

*Notation:* We write  $f(\bar{P}) \doteq \frac{1}{\bar{P}^k}$  to mean  $-\lim_{\bar{P} \rightarrow \infty} \frac{\log f(\bar{P})}{\log \bar{P}} = k$ . Similarly, we define  $f(\bar{P}) \preceq \frac{1}{\bar{P}^k}$  to mean  $-\lim_{\bar{P} \rightarrow \infty} \frac{\log f(\bar{P})}{\log \bar{P}} \geq k$ .

## 3.2 System Model

The system model consists of *node A* with a single antenna and *node B* with  $r > 2$  antennas<sup>1</sup> as shown in Fig. 3.1. The data transmission is from *node A* to *node B*. The wireless channel from *node A* to *node B*, denoted by  $\mathbf{h} \in \mathbb{C}^{r \times 1}$ , is modeled as a Rayleigh flat fading channel whose entries are *i.i.d.* Circularly Symmetric Complex Gaussian (CSCG) random variables with zero mean and unit variance, denoted  $\mathcal{CN}(0, 1)$ . The channel is assumed to be block-fading, i.e., it remains constant for a duration of the coherence time  $L_c$ , and evolves in an *i.i.d.* fashion across coherence times. We assume a reciprocal TDD system [1], and hence, without loss of generality, taking the complex conjugate of the received signal at *node A*, the reverse-link channel is  $\mathbf{h}^H$ . We let  $\mathbf{h} = \sigma \mathbf{v}$ , where  $\sigma \triangleq \|\mathbf{h}\|_2$  and  $\mathbf{v} \triangleq \frac{\mathbf{h}}{\|\mathbf{h}\|_2}$ . For analytical tractability, we assume that *node B* has perfect CSI, as in [3], [24]. Extension of the analysis to handle imperfect CSIB is difficult, because the capacity of a fading channel with mismatched CSI at *node A* and *node B* is unknown [34]. Moreover, in a TDD setup, when (say) *node B* transmits a channel-dependent training sequence based on an imperfect estimate to *node A*, the effective noise in the RCT signal at *node A* is no longer Gaussian, due to which, even lower bounding the capacity is not straightforward [37]. In the next section, we present the proposed PCRCT and the data transmission scheme, which will set the stage for the achievable DMT analysis in Sec. 3.4.

---

<sup>1</sup>This condition is required for the analysis to follow.

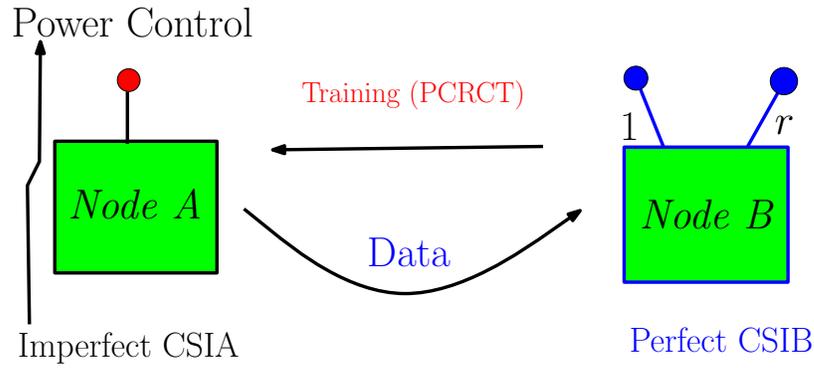


Figure 3.1: System model for PCRCT with perfect CSIR.

### 3.3 PCRCT and Data Transmission Protocol

In this chapter, we propose to employ the following *channel-dependent, power-controlled* training signal from *node B* to *node A*:

$$\mathbf{x}_{B,\tau} = \frac{\sqrt{\bar{P}^\gamma L_{B,\tau}} \sqrt{(r-1)(r-2)}}{\|\mathbf{h}\|_2^2} \mathbf{v}, \quad (3.1)$$

where  $\gamma > 0$  is a parameter that controls the RCT power. In writing the above, we have assumed  $r > 2$ . Further, the training sequence in (3.1) satisfies an *average* power constraint, i.e.,  $\mathbb{E}\|\mathbf{x}_{B,\tau}\|_2^2 = \bar{P}^\gamma L_{B,\tau}$ , since  $\mathbb{E}\frac{1}{\|\mathbf{h}\|_2^4} = \frac{1}{(r-1)(r-2)}$  for a Rayleigh fading channel. The minimum duration  $L_{B,\tau}$  of the above training signal is a single symbol, unlike a channel-agnostic orthogonal RCT scheme, which is of minimum duration equal to  $r$  symbols. Due to the channel reciprocity, the corresponding received signal at *node A* is

$$y_{A,\tau} = \sqrt{\bar{P}^\gamma L_{B,\tau}} \frac{\sqrt{(r-1)(r-2)}}{\|\mathbf{h}\|_2} + w_{A,\tau}, \quad (3.2)$$

where  $w_{A,\tau}$  is the AWGN at *node A*, distributed as  $\mathcal{CN}(0, 1)$ . Note that, in the absence of noise,  $y_{A,\tau}$  is a scaled version of the optimal channel inversion-based power control,

which achieves infinite diversity order when the CSIT is perfect [9]. Motivated by this, even in the presence of noise, using  $y_{A,\tau}$ , *node A* computes  $g_c = c_{\bar{P}} \Re\{y_{A,\tau}\}$  as the amplitude scaling applied to a unit-power data symbol at *node A*, where  $c_{\bar{P}} \triangleq \sqrt{\frac{2\bar{P}}{2(r-2)\bar{P}^\gamma L_{B,\tau} + 1}}$ . Using  $\mathbb{E}\left\{\frac{1}{\|\mathbf{h}\|_2^2}\right\} = \frac{1}{r-1}$  and  $\mathbb{E}\{\Re\{w_{A,\tau}\}^2\} = \frac{1}{2}$ , it can be shown that the average power in  $g_c$  is  $\mathbb{E}\{|g_c|^2\} = \bar{P}$ . Observe that  $c_{\bar{P}} \doteq \bar{P}^{\frac{1-\gamma}{2}}$ ; this will be used in the outage probability analysis below. Note that the training sequence in (3.1) is different from the RCT scheme in [28]. While (3.1) enables *node A* to directly estimate the data power control as a scaled version of the RCT signal, in [28], *node A* first detects an index from the received power level of the RCT signal, and then maps the decoded power index to the forward-link data power. However, it can be shown that, as  $\bar{P} \rightarrow \infty$ , the data power control in [28] and the proposed data power control both approach perfect channel inversion, albeit with different scaling factors. Let  $x_{A,d}$  denote the data symbol at *node A*, distributed as  $\mathcal{CN}(0, 1)$ . Now, *node A* multiplies the symbol  $x_{A,d}$  by  $g_c$  and transmits it to *node B*. At *node B*, pre-multiplying the received data signal by  $\mathbf{v}^H$ , we get

$$\tilde{\mathbf{y}}_{B,d} \triangleq \|\mathbf{h}\|_2 g_c x_{A,d} + \mathbf{v}^H \mathbf{w}_{B,\tau}, \quad (3.3)$$

$$= c_{\bar{P}} \sqrt{\bar{P}^\gamma L_{B,\tau} \sqrt{(r-1)(r-2)}} x_{A,d} + n_{eff}, \quad (3.4)$$

where  $\mathbf{w}_{B,\tau}$  is the AWGN at *node B* with i.i.d.  $\mathcal{CN}(0, 1)$  entries, and

$$n_{eff} \triangleq \|\mathbf{h}\|_2 c_{\bar{P}} \Re\{w_{A,\tau}\} x_{A,d} + \mathbf{v}^H \mathbf{w}_{B,\tau}.$$

Since  $\mathbb{E}\{|x_{A,d}|^2\} = 1$  and  $\mathbb{E}\{|g_c|^2\} = \bar{P}$ , the power constraint  $\bar{P}$  on the data signal is satisfied. Also, the data and the effective noise are uncorrelated given the channel, i.e.,  $\mathbb{E}[n_{eff} x_{A,d}^* | \|\mathbf{h}\|_2] = 0$ . Therefore, using the worst case noise theorem in [8], a lower

bound on the capacity of the system in (3.4) can be obtained by treating  $n_{eff}$  as Gaussian distributed noise. With some algebraic manipulation, it is given by

$$C_L \triangleq \alpha \log \left( 1 + \frac{2c_{\bar{P}}^2(r-1)(r-2)\bar{P}^\gamma L_{B,\tau}}{c_{\bar{P}}^2 \|\mathbf{h}\|_2^2 + 2} \right), \quad (3.5)$$

where  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$ . Using (3.5), an upper bound on the outage probability can be written as

$$P_{out} \leq P_{out}^u \triangleq \Pr \{C_L < R_{\bar{P}}\}, \quad (3.6)$$

where  $R_{\bar{P}}$  denotes the target data rate. We now state and prove the achievable DMT result for the proposed PCRCT and data transmission scheme.

### 3.4 Outage Analysis

We start by recalling the definitions of the multiplexing gain,  $g_m$ , and the diversity order,  $d(g_m)$ , [16]:

$$g_m \triangleq \lim_{\bar{P} \rightarrow \infty} \frac{R_{\bar{P}}}{\log \bar{P}}, \quad d(g_m) \triangleq \lim_{\bar{P} \rightarrow \infty} \frac{-\log P_{out}}{\log \bar{P}}, \quad (3.7)$$

where  $R_{\bar{P}} = g_m \log(\bar{P})$  is the target data rate with average data power  $\bar{P}$ , and  $P_{out}$  is the outage probability, i.e., the probability that  $R_{\bar{P}}$  exceeds the channel capacity. Here, we use the outage probability as a proxy for the probability of error at high SNR with finite-length codes. This is because the latter can be made to decrease as fast as the former using finite-length approximately universal codes [32, 33]. Since our proposed scheme achieves an infinite diversity order, it will be useful to consider the following additional definition: An RCT and data transmission scheme are said to achieve an

outage exponent  $\varsigma$  if the outage probability  $P_{out}$  satisfies

$$\lim_{\bar{P} \rightarrow \infty} \frac{-\log P_{out}}{\bar{P}^\varsigma} = 1. \quad (3.8)$$

The above definition captures the magnitude of the exponent with which the outage probability goes to zero, as  $\bar{P} \rightarrow \infty$ , and thus is useful to distinguish between schemes that achieve an infinite diversity order. We note that a similar concept of an exponential diversity order was defined in [38]. We are now ready to state our main result in the following theorem.

**Theorem 3.** *An upper bound on the outage probability,  $P_{out}^u \triangleq \Pr\{C_L < R_{\bar{P}}\}$ , corresponding to the PCRCT and the data transmission protocol proposed in Sec. 3.3, is given by*

$$P_{out}^u = e^{-S_{\bar{P}}} \sum_{k=0}^{r-1} \frac{(S_{\bar{P}})^{r-k-1}}{(r-k-1)!}, \quad \text{with} \quad (3.9)$$

$$S_{\bar{P}} \triangleq \frac{2(r-1)(r-2)\bar{P}^\gamma L_{B,\tau}}{(\exp(R_{\bar{P}}/\alpha) - 1)} - \frac{2}{c_{\bar{P}}^2} \quad (3.10)$$

for  $0 \leq g_m < \alpha$ . Here,  $r > 2$ , and  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$  as before. Moreover, the proposed scheme achieves the DMT

$$d(g_m) = \infty, \quad 0 \leq g_m < \alpha \min(\gamma, 1), \quad (3.11)$$

and an outage exponent  $\varsigma = \gamma - \frac{g_m}{\alpha}$ ,  $0 \leq g_m < \alpha \min(\gamma, 1)$ .

*Proof:* After some manipulation, (3.6) can be rewritten as

$$P_{out}^u = \Pr \{ \|\mathbf{h}\|_2^2 \geq S_{\bar{P}} \}, \quad (3.12)$$

where  $S_{\bar{P}}$  is defined as in (3.10). Note that, since  $c_{\bar{P}}^2 \doteq \bar{P}^{1-\gamma}$ , we have  $S_{\bar{P}} \doteq \bar{P}^{(\gamma - \frac{g_m}{\alpha})} (1 - \bar{P}^{(\frac{g_m}{\alpha} - 1)}) > 0$  if  $0 \leq g_m \leq \alpha$ . Also, the exponent of  $\bar{P}$  in  $S_{\bar{P}}$  is positive, provided  $g_m < \gamma\alpha$ .

Using the fact that  $\|\mathbf{h}_2\|^2$  is chi-square distributed with  $2r$  degrees of freedom, (3.12) can be written as

$$P_{out}^u = \frac{1}{\Gamma(r)} \int_{S_{\bar{P}}}^{\infty} x^{r-1} e^{-x} dx = e^{-S_{\bar{P}}} \sum_{l=0}^{r-1} \frac{(S_{\bar{P}})^{r-l-1}}{(r-l-1)!}, \quad (3.13)$$

which results in (3.9). Since the terms inside the summation in (3.9) are polynomial in  $S_{\bar{P}}$ , due to the  $e^{-S_{\bar{P}}}$  term, we get the diversity order as  $d(g_m) = \infty$ , for  $0 \leq g_m < \alpha \min(\gamma, 1)$ . Recall that the condition  $r > 2$  is required for satisfying the RCT power constraint of  $\bar{P}^\gamma$ . Using (3.9) in the definition in (3.8), it immediately follows that  $\varsigma = \gamma - \frac{g_m}{\alpha}$ . ■

**Remarks:** The above result shows that an infinite diversity order can be obtained using the proposed channel-dependent PCRCT and data power control scheme. In contrast, channel-agnostic orthogonal RCT with a channel inversion based power control only achieves a finite diversity order [1, 3, 24]. Also, for a given  $g_m$ , the above result shows that the outage exponent increases with  $\gamma$ , i.e., using a higher power for the RCT signal is beneficial. In addition to the outage exponent, it is interesting to investigate the probability of error exponent with finite length codes and ascertain whether it achieves an infinite diversity order. We address this in Theorem 4 below.

### 3.5 An Achievable Coding Scheme

In this section, we consider the probability of error exponent of the above PCRCT and power control, and show that it achieves an infinite diversity order. The proof involves proposing a transmission scheme that achieves a rate of  $\frac{g_m}{\alpha} \log \bar{P}$ , upper bounding its pairwise probability of error, and showing that the union bound on the probability of error achieves an infinite diversity order.

**Theorem 4.** For the proposed PCRCT and power control, there exists a data transmission scheme that achieves an infinite diversity order for  $0 \leq g_m < \alpha \min(\gamma, 1)$ , for all block lengths  $L_d \geq 1$ . In particular, the probability of error  $P_e \preceq \exp(-\bar{P}^E)$ , where

$$E = \begin{cases} \frac{1}{2} \left( \gamma - \frac{g_m}{\alpha} \right), & 0 \leq \frac{g_m}{\alpha} < \min(2 - \gamma, \gamma, 1) \\ 1 - \frac{g_m}{\alpha}, & \min(2 - \gamma, \gamma, 1) \leq \frac{g_m}{\alpha} < \min(\gamma, 1). \end{cases} \quad (3.14)$$

*Proof:* See Appendix B.0.6. ■

The above result shows that the outage analysis captures the fact that an infinite diversity order can be achieved using finite block length code, and hence, the proposed scheme is DMT optimal for  $\gamma \geq 1$ . However, it also indicates that there is a mismatch in the exponential order with which the probability of error of the proposed uncoded transmission scheme and the outage probability go to zero. Further, we see that, for a given  $\frac{g_m}{\alpha}$ , increasing the PCRCT power exponent  $\gamma$  beyond  $2 - \frac{g_m}{\alpha}$  does not further improve the probability of error exponent. On the other hand, when  $\frac{g_m}{\alpha} < \gamma < 2 - \frac{g_m}{\alpha}$ , the probability of error exponent linearly improves with  $\gamma$ .

### 3.6 Numerical Results

We now briefly present numerical results to illustrate the outage probability performance of our proposed channel-dependent PCRCT and forward-link data power control schemes. We consider a Rayleigh fading channel with three receive antennas. We plot the upper bound on the outage probability in (3.9) versus data power for different values of  $\gamma$  in Fig. 3.2. The figure corresponds to  $r = 3$ ,  $g_m = 0.9$ ,  $L_c = 10$  and  $L_{B,\tau} = 1$ , which results in  $\alpha = 0.9$ . Thus, when  $\gamma = 1$ , the outage exponent becomes zero, which

results in a constant outage probability (see (3.9)) as illustrated in the figure. Also, when  $\gamma$  is increased, a lower outage probability can be achieved, as expected.

Figure 3.3 shows a plot of  $-\log(P_{out}^u)$  from (3.9) versus  $\bar{P}$  for a  $3 \times 1$  system with  $L_c = 100$ , for different values of  $g_m$  and  $L_{B,\tau}$ . As expected, the slope of the curves corresponding to  $g_m = 0.3$  is higher than that of the curves corresponding to  $g_m = 0.6$ . Also, the performance with  $L_{B,\tau} = 3$  is marginally lower than that with  $L_{B,\tau} = 1$ , indicating that  $L_{B,\tau} = 1$  is the optimal choice, i.e., it is optimal to train for a single symbol duration in the reverse channel. The curves obtained from orthogonal RCT saturate as the training power is increased, corresponding to the finite diversity order achieved with orthogonal training [3]. As the training power is increased, the proposed RCT scheme significantly outperforms orthogonal RCT. The exact SNR at which the curves cross would be lower than the value seen in the graph, because the upper bound derived here is being compared with the lower bound on the outage probability with orthogonal RCT derived in [3]. Thus, the proposed power-controlled training and data transmission scheme offers a significantly better DMT performance compared to the orthogonal RCT scheme.

### 3.7 Conclusions

In this chapter, we proposed a channel-dependent PCRCT and a data power control scheme in a TDD-SIMO system with perfect CSIB, and analyzed its DMT performance. We showed that the proposed scheme achieves an infinite diversity order for  $0 \leq g_m < \alpha \min(\gamma, 1)$ . Also, at high SNR, the derived upper bound on the outage probability goes to zero approximately as  $\exp(-\bar{P}^{(\gamma - \frac{g_m}{\alpha})})$ , where  $\gamma > 0$  is the exponent of the RCT power.

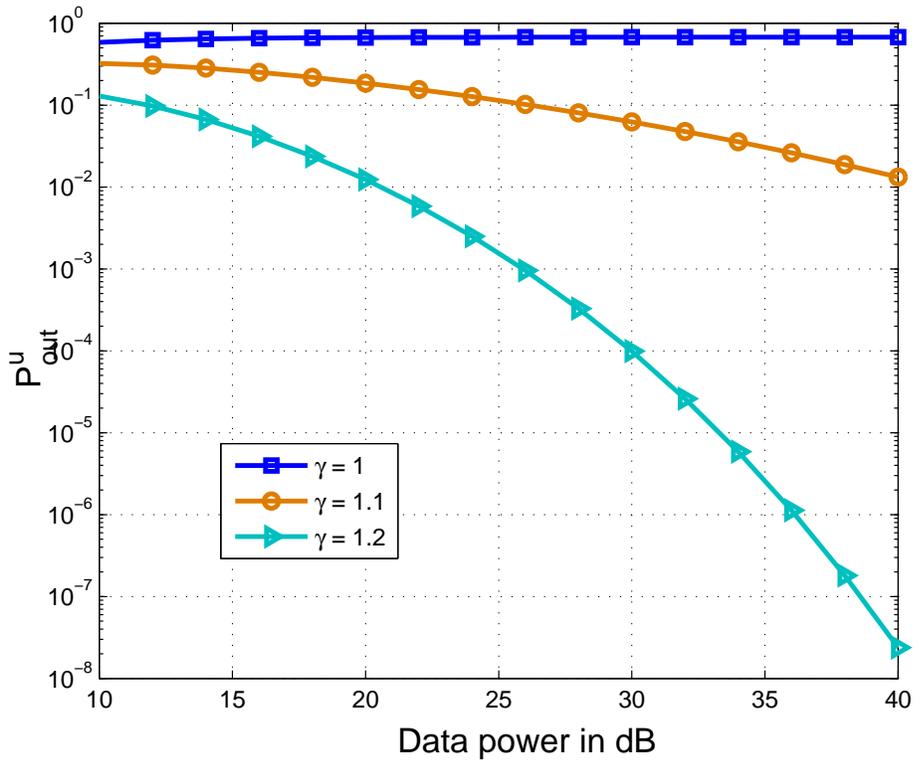


Figure 3.2: Upper bound on the outage probability versus data power for different values of  $\gamma$ , with  $r = 3$ ,  $g_m = 0.9$ ,  $L_c = 10$ , and  $L_{B,\tau} = 1$ .

We also showed that there exists an uncoded data transmission scheme for which the probability of error exhibits an infinite diversity order for  $0 \leq g_m < \alpha \min(\gamma, 1)$ . The proposed scheme can thus achieve a significantly better DMT performance compared to fixed-power, channel-agnostic orthogonal RCT schemes and the RCT scheme of the previous chapter. The next chapter presents a fixed-power RCT sequence design for general MIMO-BF systems with a lower bound on the capacity as the performance metric.

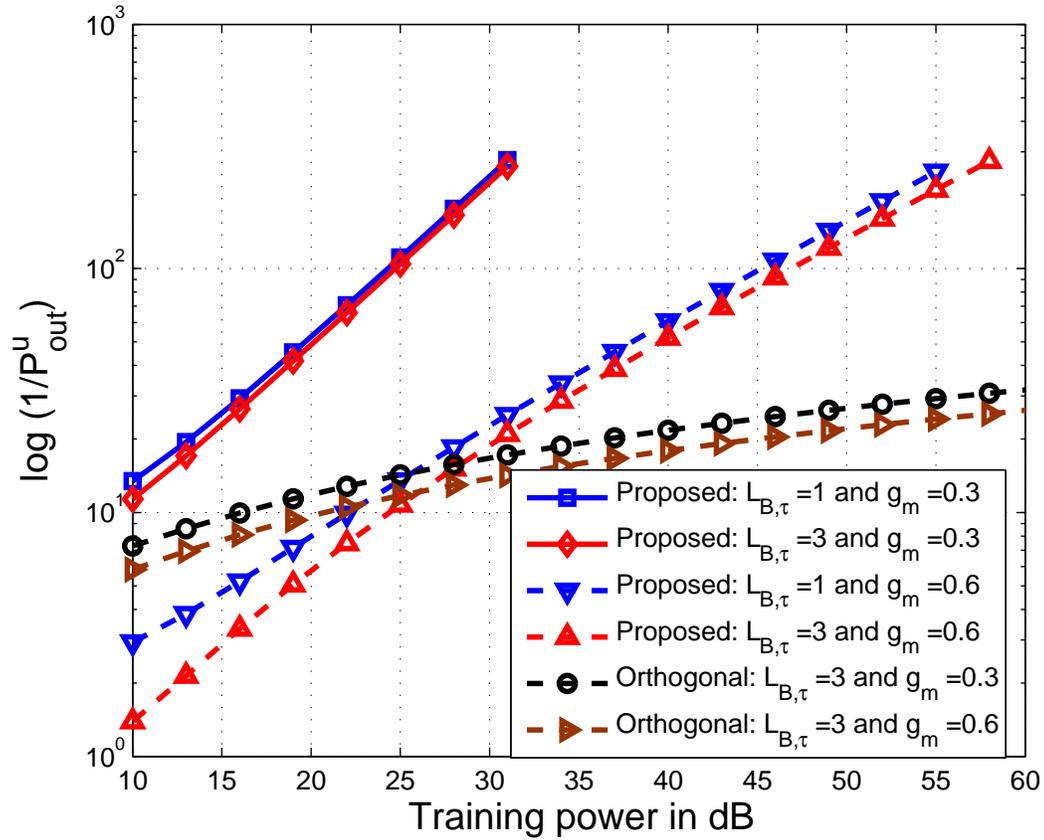


Figure 3.3: A plot of  $-\log(P_{out}^u)$  in (3.9) versus training power with  $r = 3$ ,  $\gamma = 1$  and  $L_c = 100$  for different values of  $g_m$  and  $L_{B,\tau}$ . The curves labeled orthogonal refer to the lower bound on the outage probability with channel-agnostic orthogonal RCT in [3], which requires a minimum training duration of  $r = 3$  symbols.

# Chapter 4

## Channel Training Signal Design for Reciprocal Multiple Antenna Systems with Beamforming

### 4.1 Introduction

In mobile and vehicular communications, fast and reliable acquisition of Channel State Information (CSI) at the Transmitter (CSIT) and Receiver (CSIR) is key to realizing the professed diversity benefits of multiple antenna transmission techniques such as maximal ratio transmission Beamforming (BF) [6]. Typically, CSIR is acquired by sending a known training signal from the transmitter to the receiver [8]. CSIT, on the other hand, can be obtained using feedback of *quantized* CSI from the receiver to the transmitter when the former has CSI [39–41]; training in the reverse-link, also called *Reverse Channel Training (RCT)* [5]; or a combination of the two [42]. The quantized feedback approach assumes that the reverse-link is a fixed-rate channel, independent of the forward-link. This chapter focuses on the RCT approach, which relies on *channel reciprocity*, i.e., that

the forward and reverse channels are the same.<sup>1</sup> The reciprocal channel is a good model for the reverse-link, for example, in Time Division Duplex (TDD) systems.

In this chapter, we consider two communicating nodes, *node A* and *node B*, with  $n_A$  and  $n_B$  antennas, respectively; and with data being transmitted from *node A* to *node B* using BF over the dominant mode of the channel. In conventional channel training, one employs a known, orthogonal training signal in the forward and reverse directions successively to convey the CSI to *node A* (CSIA) and *node B* (CSIB), respectively, prior to data transmission. This requires a minimum of  $n_A$  (and  $n_B$ ) symbols for acquiring CSIB (and CSIA). The time overhead for channel estimation could be significant, especially in mobile communications when the channel is fast-varying, since longer training reduces the time available for data transmission, leading to a reduction in the average data rate.

One option that has not been explored much in the literature for TDD BF-based systems is to exploit the CSI obtained from the initial training in one direction, to design the training signal in the opposite direction and thereby improve the efficiency and accuracy of estimation at the other node. At first glance, it would appear that using a CSI-dependent training signal is not feasible: training-based channel estimation requires that the training signal be *known* at both ends. In the case of BF-based data transmission, the transmitter needs to estimate only the transmit BF vector, rather than the entire channel matrix. Due to this, it turns out that using a simple channel-dependent BF-based RCT scheme, it is possible to efficiently estimate the transmit BF vector with a significantly lower time overhead in training. In [30], we had proposed the idea of using CSIB as a side information in designing the RCT signal; this is explored in depth

---

<sup>1</sup>Channel reciprocity requires well-calibrated transmit and receive RF chains at both ends, which is assumed here.

in this chapter.

We start with a survey of related literature. It is shown in [8] that for a MIMO Spatial Multiplexing (SM) system with no CSI at the transmitter and with a total energy constraint on the data and training signals, orthogonal training of duration equal to  $n_A$  symbols is optimal with respect to a lower bound on the forward-link capacity. The idea of using fixed, orthogonal signals for acquiring CSIA is considered in [4, 5, 43, 44]; while RCT sequences that exploit CSI have appeared in [31, 45–49]. In [44], the authors propose a channel-dependent training sequence in the form of analog feedback of the received pilot symbols, for a multiuser MIMO system that helps users acquire the relevant CSI using minimal training overhead. In [45] and [46], the reverse-link training signal is a scaled version of the received training signal in the forward-link. Although it is shown that this outperforms orthogonal training in the reverse-link, it has the disadvantage that the transmitter estimates the entire channel, which is not required for certain types of data transmission such as BF. The authors in [47–49] propose methods for directly estimating the dominant singular vectors blindly without estimating the entire TDD MIMO channel matrix. However, they require several rounds of communication, reasonably high data SNR, and the channel to be relatively slowly varying, in order to converge and track the dominant mode of the channel. In [31], Zhou et al. consider a Multiple-Input Single-Output (MISO) two-way communication system, and design the forward and reverse training and data powers subject to a total constraint on the available resources to approximately minimize the SER. The diversity multiplexing gain trade-off analysis with imperfect CSIA and CSIB has also been considered, e.g., in [24]. However, none of the aforementioned works design the RCT signal specifically

to enable the transmitter to directly estimate the BF vector to be used for data transmission, nor do they consider the optimization of the training duration from an average data rate perspective, which is the focus of this chapter. We consider a Multiple-Input Multiple-Output (MIMO) communication system with the achievable data rate as our performance metric, and address the following issues:

*CSIA acquisition given perfect CSIB:* Assuming perfect CSI is available at *node B*, we propose a *channel-dependent* RCT signal that enables *node A* to directly estimate the BF vector with a minimum training duration of only one symbol, an  $n_B$  factor reduction in the training overhead compared to orthogonal RCT (e.g., [4, 5]).

For our proposed RCT scheme, we derive a tight approximate lower bound on the capacity with the estimated CSIA, and use it to obtain a closed-form, near-optimal solution for the reverse training duration. Through numerical simulations, we illustrate the tightness of the approximate capacity lower bound, and also show that the proposed RCT scheme significantly outperforms the existing orthogonal RCT methods.

We next consider the optimal sharing of resources (both the power and duration) for training and data transmission when the two nodes have a joint power constraint. The analysis yields interesting insights into the tradeoff between RCT and the forward data rate; similar studies have been considered in the past literature (e.g., [31]) also. We show that when the total energy available for communication is constrained, regardless of the specific value of the available energy, it is optimal in terms of the approximate capacity lower bound to employ a single RCT symbol, and that the fractional power that should be spent on data transmission is  $\frac{\sqrt{2(L_c-1)}}{\sqrt{2(L_c-1)+1}}$ , where  $L_c$  is the channel coherence time. (See Sec. 4.3)

*CSIB acquisition given perfect CSIA:* We propose an analogous scheme for acquiring CSIB when *node A* has perfect CSI. We again derive an approximate capacity lower bound with our proposed training scheme, and maximize it with respect to the training and data powers and durations. We show that about half the total available power should be spent on the forward training, and that the optimal training duration is 1 symbol, at low data SNR. (See Sec. 4.4)

We also consider the question of whether to initiate the training at *node A* or at *node B*. With the proposed scheme, *node A* (similarly, *node B*) initiated training incurs a minimum total overhead of  $n_A + 1$  (similarly,  $n_B + 1$ ) training symbols. We show that *node A* initiated training (i.e., perfect CSIB) along with the proposed channel-dependent RCT scheme outperforms *node B* initiated training (i.e., perfect CSIA) if  $L_c > \max(n_A, n_B) + 2$  and  $L_c \geq 2n_A - n_B + 1$ , which is typically the case in practice. (See Sec. 4.5)

We corroborate our theoretical results through Monte Carlo simulations in Sec. 4.6, and offer concluding remarks in Sec. 4.7. Proofs of the theorems and some detailed derivations are relegated to the Appendix. Finally, note that, although we derive our results for a single-user system, they can be directly applied in a multiuser setting with user selection and orthogonal access schemes, since the results are derived on a per channel instantiation basis.

*Notation:* In addition to the notation we have employed thus far, we write  $f(x) = \mathcal{O}(g(x))$  to mean that  $\limsup_{x \rightarrow \infty} \left| \frac{f(x)}{g(x)} \right| < \infty$ . All logarithms in the sequel are base- $e$ .

## 4.2 System Model

The system model considered in this chapter consists of two nodes, *node A* and *node B*, with  $n_A$  and  $n_B$  antennas, respectively. The wireless channel from *node A* to *node B*, denoted by  $H \in \mathcal{C}^{n_B \times n_A}$ , is modeled as a quasi-static flat fading channel with coherence time  $L_c$ . That is, the channel is assumed to remain constant for a frame of duration  $L_c$ , and evolve in an i.i.d. fashion from frame to frame. Let  $H = U\Sigma V^H$  be the Singular Value Decomposition (SVD) of  $H$ . The diagonal entries of  $\Sigma \in \mathcal{R}^{n_B \times n_A}$ , denoted  $\sigma_1, \dots, \sigma_n$ , are the singular values of  $H$ ; with  $n \triangleq \text{rank}(H)$ . The columns of unitary matrices  $U \in \mathcal{C}^{n_B \times n_B}$  and  $V \in \mathcal{C}^{n_A \times n_A}$  are the eigenvectors of  $HH^H$  and  $H^H H$ , respectively. Assuming a TDD system with perfect reciprocity, and considering the complex conjugate of the transmitted and received signals as the channel input and output, respectively, the channel from *node B* to *node A* is  $H^H$ . The equations corresponding to training and data transmission in either direction are given by

$$\text{Forward-link training: } \mathbf{y}_{B,\tau} = H\mathbf{x}_{A,\tau} + \mathbf{w}_{B,\tau}, \quad (4.1)$$

$$\text{Reverse-link training: } \mathbf{y}_{A,\tau} = H^H\mathbf{x}_{B,\tau} + \mathbf{w}_{A,\tau}, \quad (4.2)$$

$$\text{Forward-link data: } \mathbf{y}_{B,d} = H\mathbf{x}_{A,d} + \mathbf{w}_{B,d}, \quad (4.3)$$

where  $\mathbf{x}_{i,\tau}$  (and  $\mathbf{x}_{i,d}$ )  $\in \mathcal{C}^{n_i}$  and  $\mathbf{y}_{j,\tau}$  (and  $\mathbf{y}_{j,d}$ )  $\in \mathcal{C}^{n_j}$  are the transmitted and received training (and data) signals at nodes  $i$  and  $j$ , respectively. The signal  $\mathbf{x}_{A,d}$  consists of a  $\mathcal{CN}(0, 1)$  distributed data symbol multiplied by a BF vector. The entries of the noise vectors  $\mathbf{w}_{B,\tau}$ ,  $\mathbf{w}_{A,\tau}$  and  $\mathbf{w}_{B,d}$  are assumed  $\mathcal{CN}(0, 1)$  distributed. The training signals are

assumed to be transmitted at a fixed-power. Power controlled training, where the training power is varied with the channel state, significantly changes the problem, and is explored in [12].

In this chapter, we consider a MIMO-TDD BF system with perfect CSI at one of the nodes. Such an assumption has been made in past literature in both reciprocal MIMO systems (e.g., [3, 24]) and quantized-feedback based MIMO systems (e.g., [39, 40]). We refer to the scheme where *node B* (similarly, *node A*) has perfect CSI as *node A* (similarly, *node B*) *initiated training*, corresponding to an initial high power training signal sent from the respective node. The perfect CSIB (CSIA) assumption makes the analysis tractable, and isolates the effect of reverse (forward) channel training on the achievable data rate performance. Another reason for considering the perfect CSI assumption is that the capacity of a fading channel with *mismatched* CSIT and CSIR is not known in closed form [1, 34]. Moreover, in a TDD setup, when (say) *node B* uses an imperfect channel estimate to transmit a channel-dependent training sequence to *node A*, the effective noise in the reverse training signal at *node A* is no longer Gaussian, due to which, even lower bounding the capacity is not straightforward [37]. Therefore, extending this work to allow for imperfect CSI at both nodes requires one to consider other performance metrics such as the outage probability; the interested reader is referred to an analysis from a Diversity-Multiplexing Gain Tradeoff (DMT) perspective described in Chapter 2 [11]. In the next section, we present the proposed RCT signal and analyze its average data rate performance.

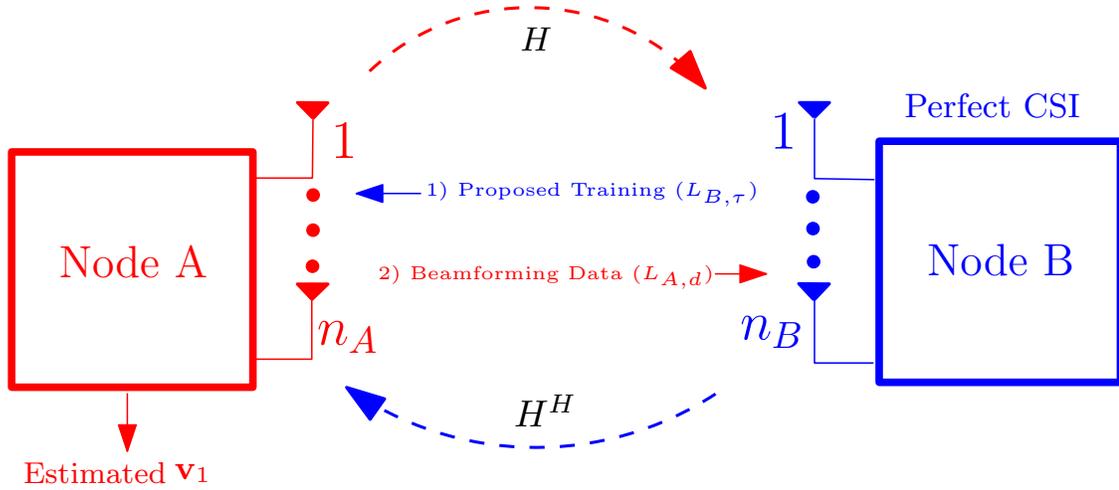
### 4.3 RCT Signal Design With Perfect CSIB

In this section, we consider a MIMO-TDD BF system with *node A* initiated training, i.e., with perfect CSIB. For BF, it suffices for *node A* to acquire knowledge of  $\mathbf{v}_1$ , the dominant right singular vector of  $H$ . *Node A* then uses its estimate of  $\mathbf{v}_1$  to send beamformed data to *node B*. The transmission protocol is shown in Fig. 4.1. It consists of an RCT phase of duration  $L_{B,\tau}$  symbols, followed by a forward-link BF-based data transmission phase of  $L_{A,d} = L_c - L_{B,\tau}$  symbols.

Now, since *node B* has perfect CSI, unlike conventional channel-agnostic training, we propose the following training signal that exploits the CSI at *node B*:

$$\mathbf{x}_{B,\tau} = \sqrt{P_{B,\tau} L_{B,\tau}} \mathbf{u}_1, \quad (4.4)$$

where  $P_{B,\tau}$  and  $L_{B,\tau}$  are training power and training duration at *node B*, which is known at both nodes, while the left dominant singular vector of the channel,  $\mathbf{u}_1$ , is adapted at *node B* based on the CSIB. Note that, strictly speaking,  $\mathbf{x}_{B,\tau} = \sqrt{P_{B,\tau}} \mathbf{u}_1$  is transmitted repeatedly  $L_{B,\tau}$  times; this is mathematically equivalent to using  $\mathbf{x}_{B,\tau} = \sqrt{P_{B,\tau} L_{B,\tau}} \mathbf{u}_1$  for a duration of one symbol. Also, in practical fixed-point implementations, the quantization process would introduce errors in the BF vectors employed for transmission. Incorporating this effect in our present model is difficult, as it makes the analysis mathematically intractable. However, in Sec. 4.6, we show through simulations that the proposed schemes are robust to quantization errors. Now, the training signal in (4.4) satisfies an instantaneous power constraint of  $P_{B,\tau} L_{B,\tau}$ , since  $\|\mathbf{x}_{B,\tau}\|_2^2 = P_{B,\tau} L_{B,\tau}$ . The

Figure 4.1: *Node A* initiated training.

corresponding received training signal at *node A* is given by

$$\mathbf{y}_{A,\tau} = \sqrt{P_{B,\tau} L_{B,\tau} \sigma_1} \mathbf{v}_1 + \mathbf{w}_{A,\tau}. \quad (4.5)$$

Using (4.5), *node A* estimates  $\mathbf{v}_1$  as

$$\hat{\mathbf{v}}_1 = \frac{\mathbf{y}_{A,\tau}}{\|\mathbf{y}_{A,\tau}\|_2}. \quad (4.6)$$

Note that, in the noiseless case, (4.6) results in  $\hat{\mathbf{v}}_1 = \mathbf{v}_1$ . Moreover, conditioned on  $\sigma_1$ , it asymptotically achieves a Constrained Cramér-Rao Lower Bound (CCRLB), a result that is stated and proved as Theorem 14 in Appendix C.0.7.

Using  $\hat{\mathbf{v}}_1$  in (4.6) for BF data transmission, from (4.3), the received data signal at *node B* is

$$\mathbf{y}_{B,d} = \sqrt{P_{A,d}} H \hat{\mathbf{v}}_1 x_{A,d} + \mathbf{w}_{B,d}, \quad (4.7)$$

where the data symbol  $x_{A,d} \in \mathcal{C}$  is  $\mathcal{CN}(0, 1)$  distributed and hence satisfies  $\mathbb{E}|x_{A,d}|^2 = 1$ ,

which ensures that the average power constraint of  $P_{A,d}$  is satisfied. Now, since *node B* has perfect CSI, pre-multiplying (4.7) by  $\mathbf{u}_1^H$ , we get

$$\begin{aligned}\tilde{y}_{B,d} &\triangleq \mathbf{u}_1^H \mathbf{y}_{B,d} \\ &= \sqrt{P_{A,d}}\sigma_1 x_{A,d} - \sqrt{P_{A,d}}\sigma_1 \mathbf{v}_1^H \tilde{\mathbf{v}}_1 x_{A,d} + \mathbf{u}_1^H \mathbf{w}_{B,d}, \\ &= \sqrt{P_{A,d}}\sigma_1 (1 - \mathbf{v}_1^H \mathbb{E}\{\tilde{\mathbf{v}}_1|H\})x_{A,d} + \tilde{w}_{\text{eff}},\end{aligned}\tag{4.8}$$

where  $\tilde{\mathbf{v}}_1 \triangleq \mathbf{v}_1 - \hat{\mathbf{v}}_1$ , and for notational convenience, we define

$$\tilde{w}_{\text{eff}} \triangleq \sqrt{P_{A,d}}\sigma_1 \mathbf{v}_1^H (\mathbb{E}\{\tilde{\mathbf{v}}_1|H\} - \tilde{\mathbf{v}}_1)x_{A,d} + \mathbf{u}_1^H \mathbf{w}_{B,d}.\tag{4.9}$$

Note that the first term in (4.8) is a deterministic function of  $H$ , and that the effective noise is uncorrelated with the data given the channel, i.e.,  $\mathbb{E}\{\tilde{w}_{\text{eff}}x_{A,d}^*|H\} = 0$ . Thus, for the system in (4.8), an application of the worst case noise theorem [8] results in the following capacity lower bound:

$$C_{BA,L} \triangleq \alpha \mathbb{E}_H \log \left( 1 + P_{A,d} \frac{\sigma_1^2 |a_1|^2}{1 + P_{A,d} \sigma_1^2 \mathbb{E}|a_2|^2} \right),\tag{4.10}$$

where  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$ ,  $a_1 \triangleq 1 - \mathbf{v}_1^H \mathbb{E}\{\tilde{\mathbf{v}}_1|H\}$ ,  $a_2 \triangleq \mathbf{v}_1^H (\tilde{\mathbf{v}}_1 - \mathbb{E}\{\tilde{\mathbf{v}}_1|H\})$ . Note that the expectation inside the bracket is with respect to the noise in the received training symbols, and  $\mathbb{E}_H$  denotes the expectation with respect to the channel statistics. The above expression uses the fact that  $\mathbb{E}\{\mathbf{v}_1|H\} = \mathbf{v}_1$ . Note that, (4.10) is valid regardless of the method of estimating the BF vector at *node A*. In particular, it is also valid for conventional orthogonal training. Now, we would like to solve the following optimization

problem:

$$\max_{L_{B,\tau}: 1 \leq L_{B,\tau} \leq L_c} C_{BA,L}. \quad (4.11)$$

The above problem can be numerically solved using an off-line search algorithm. However, in order to find a tractable and closed form solution, in the following, we derive an approximate expression for the capacity lower bound  $C_{BA,L}$  in (4.10), and use it to optimize the training power and duration.

### 4.3.1 Approximate Capacity Lower Bound

We have the following tight approximation for  $C_{BA,L}$  in (4.10). The accuracy of the approximation is captured in Lemma 3; and will be illustrated through simulations in Sec. 4.6. The approximate capacity lower bound is as follows:

$$C_{BA,L} \approx C_{B,A,\text{approx}} \triangleq \alpha \mathbb{E} \log \left[ 1 + \frac{P_{A,d} \sigma_1^2}{1 + \frac{P_{A,d}}{2P_{B,\tau} L_{B,\tau}}} \right], \quad (4.12)$$

where the expectation is taken with respect to the distribution of  $\sigma_1$ . The derivation of (4.12) is provided in Appendix C.0.8. The right hand side above succinctly captures the effect of training duration: as  $L_{B,\tau}$  is increased,  $\alpha$  captures the data rate loss due to the time overhead in training, while the expectation term captures the data rate improvement due to the increased accuracy of BF vector estimation.

The following Lemma asserts that the expression in (4.12) becomes exact in the limit of large data and training powers, when their ratio is kept fixed. Its proof is straightforward, and is omitted. Further discussion on the tightness of (4.12) is relegated to Sec. 4.6.

**Lemma 3.** *If  $\lim_{P_{A,d}, P_{B,\tau} \rightarrow \infty} \frac{P_{A,d}}{P_{B,\tau}} = \mu_{d,\tau} < \infty$ , then*

$$\lim_{P_{A,d}, P_{B,\tau} \rightarrow \infty} \left| C_{BA,L} - \alpha \mathbb{E} \log \left[ 1 + \frac{P_{A,d} \sigma_1^2}{1 + \frac{P_{A,d}}{2P_{B,\tau} L_{B,\tau}}} \right] \right| = 0. \quad (4.13)$$

### 4.3.2 Optimal Training Duration

In this subsection, given  $P_{A,d}$  and  $P_{B,\tau}$ , we wish to solve the following optimization problem:

$$\max_{1 \leq L_{B,\tau} \leq L_c} C_{B,A,\text{approx}}. \quad (4.14)$$

It will be shown through simulation results in Sec. 4.6 that the solution to (4.14) matches with that obtained by numerically optimizing (4.10) for a wide range of parameters, due to the tightness of the lower bound. Now, the above problem can be solved using a simple line search in the finite interval  $[1, L_c - 1]$ . However, one can get a closed form solution for the optimum training duration using the following expression:

$$f(L_{B,\tau}) \triangleq \left( \frac{L_c - L_{B,\tau}}{L_c} \right) \frac{P_{A,d} \mathbb{E} \sigma_1^2}{1 + \frac{P_{A,d}}{2P_{B,\tau} L_{B,\tau}}}. \quad (4.15)$$

The above is an approximation of  $C_{B,A,\text{approx}}$  at low data SNR, i.e., when  $P_{A,d} \ll 1$ . In Sec. 4.6, we show through simulations that the optimum  $L_{B,\tau}$  obtained by optimizing  $f(L_{B,\tau})$  closely matches with that obtained by numerically optimizing (4.12) (as well as (4.10)), at low data SNR. Hence, we solve the following optimization problem:

$$\max_{1 \leq L_{B,\tau} \leq L_c} f(L_{B,\tau}). \quad (4.16)$$

Note that  $f(L_{B,\tau})$  is a concave function of  $L_{B,\tau}$ . Therefore, it is easy to show that the optimal solution is given by

$$L_{B,\tau}^* = \max \left\{ 1, -\beta + \sqrt{\beta^2 + \beta L_c} \right\}, \quad (4.17)$$

where  $\beta \triangleq \frac{P_{A,d}}{2P_{B,\tau}}$ . If  $L_{B,\tau}^*$  is not an integer, we choose  $\lceil L_{B,\tau}^* \rceil$  as the solution if  $f(\lceil L_{B,\tau}^* \rceil) > f(\lfloor L_{B,\tau}^* \rfloor)$ ; otherwise,  $\lfloor L_{B,\tau}^* \rfloor$  is the solution. Note that, since  $\beta > 0$ ,  $-\beta + \sqrt{\beta^2 + \beta L_c} < L_c$ .

Now, since  $f(L_{B,\tau})$  is unimodal, comparing  $f(1)$  with  $f(2)$ , it can be shown that  $L_{B,\tau}^* > 1$  if and only if  $\beta > \frac{2}{L_c - 3}$ , i.e., for  $P_{B,\tau} < \frac{P_{A,d}(L_c - 3)}{4}$ . In other words, using a single RCT symbol is optimal in terms of the approximate capacity lower bound only when the RCT power is at least one fourth of the total power in all the forward-link data symbols sent within the channel coherence time put together. At lower RCT powers, using more than one RCT symbol is preferable.

Next, when  $\beta < 1$  is some fixed, small value, and  $L_c$  is large compared to  $\beta$ , it can be seen that  $L_{B,\tau}^* \approx \sqrt{\beta L_c}$ , i.e., the time that should be allocated for training increases as the square-root of the channel coherence time. More precisely, the optimal training duration scales as an *integer* close to  $\sqrt{\beta L_c}$  that maximizes  $f(L_{B,\tau})$ . Numerical results and further insights into the solution are provided in Sec. 4.6.

### 4.3.3 Optimal Sharing of Resources

In this subsection, we consider the RCT design when the overall energy efficiency of the communication system is taken into account, including the training and data phases. We study the optimal energy and time allocation between *data transmission* at *node A*

and *reverse training* at *node B*, for a given *total energy budget* of  $\rho L_c$ , where  $\rho$  is the average power. Although nodes do not share a power source, the study presented here gives insights on the impact of RCT on the forward data transmission, by providing a mechanism for tying the energy spent in RCT into the energy cost associated with the forward-link communication. For example, it will turn out that given a constraint on the total available energy for communication, it is optimal to train for the minimum possible duration and allot a fraction  $1/(\sqrt{2(L_c - 1)} + 1)$  of the total available energy for training.

The energy efficient design is pertinent when one considers a scenario where *node B* also has data to send to *node A*. In this case, the energy spent by *node B* in sending training to *node A* is also energy lost from transmitting its own data. Using a total energy constraint is an indirect way of ensuring fairness in the utilization of energy and improving the efficiency of the two-way data communication. Moreover, the result presented here facilitates the comparison of *node A* initiated training versus *node B* initiated training on a level-playing field (see Sec. 4.5). A similar analysis, albeit in a different context of the MISO channel, with the SER as the performance metric, was considered in [31]. Here, we consider the data rate as our performance measure in a MIMO-BF system, and solve the following problem:

$$\max_{P_{B,\tau}, P_{A,d} \geq 0, L_{B,\tau}, L_{A,d} \geq 1} \alpha \mathbb{E} \log \left[ 1 + \frac{P_{A,d} \sigma_1^2}{1 + \frac{P_{A,d}}{2P_{B,\tau} L_{B,\tau}}} \right], \quad (4.18)$$

subject to the total energy constraint  $P_{B,\tau} L_{B,\tau} + P_{A,d} L_{A,d} = \rho L_c$ , and with the data transmit duration  $L_{A,d}$  satisfying  $L_{B,\tau} + L_{A,d} = L_c$ .

Intuitively, one would expect that the larger the  $\rho$ , the smaller the optimal  $L_{B,\tau}$  and the

larger the optimal fractional energy used for the data signal. Surprisingly, it turns out that both the optimal training duration ( $L_{B,\tau}^*$ ) and the optimal fraction of the total energy used for data depend only on  $L_c$ , and are independent of  $\rho$ , as shown below. Now, since  $P_{B,\tau}$  and  $P_{A,d}$  enter into the capacity lower bound only inside the  $\log(\cdot)$  function, without loss of optimality, we first optimize  $P_{A,d}$  and  $P_{B,\tau}$  for a given data duration, substitute the result into the objective function, and then optimize  $L_{B,\tau}$  and  $L_{A,d}$ . The solution is given in Theorems 5 and 6.

**Theorem 5.** *For a given data duration  $1 \leq L_{A,d} \leq L_c$  and with  $L_{B,\tau} = L_c - L_{A,d}$ , the optimal values of  $P_{B,\tau}$  and  $P_{A,d}$  that solve (4.18) are given by*

$$P_{A,d}^* = \frac{1}{L_{A,d}} \alpha^* \rho L_c, \quad \text{and} \quad P_{B,\tau}^* = \frac{1}{L_c - L_{A,d}} (1 - \alpha^*) \rho L_c, \quad (4.19)$$

where  $\alpha^* \triangleq \frac{\sqrt{2L_{A,d}}}{\sqrt{2L_{A,d}+1}}$ . The corresponding capacity lower bound is

$$C_{B,A,\text{approx}} = \frac{L_{A,d}}{L_c} \mathbb{E} \log \left( 1 + \frac{2\rho L_c \sigma_1^2}{(\sqrt{2L_{A,d}+1})^2} \right). \quad (4.20)$$

*Proof:* See Appendix C.0.9. ■

Now, it remains to find the optimal value of  $L_{A,d}$  that maximizes  $C_{B,A,\text{approx}}$  in (4.20), which is the essence of the following Theorem.

**Theorem 6.** *The optimal data duration is given by  $L_{A,d}^* = L_c - 1$ , i.e.,  $L_{B,\tau}^* = 1$ . The corresponding optimal capacity lower bound expression is*

$$C_{B,A,\text{approx}}^* = \frac{L_c - 1}{L_c} \mathbb{E} \log (1 + \rho_{\text{eff}} \sigma_1^2), \quad (4.21)$$

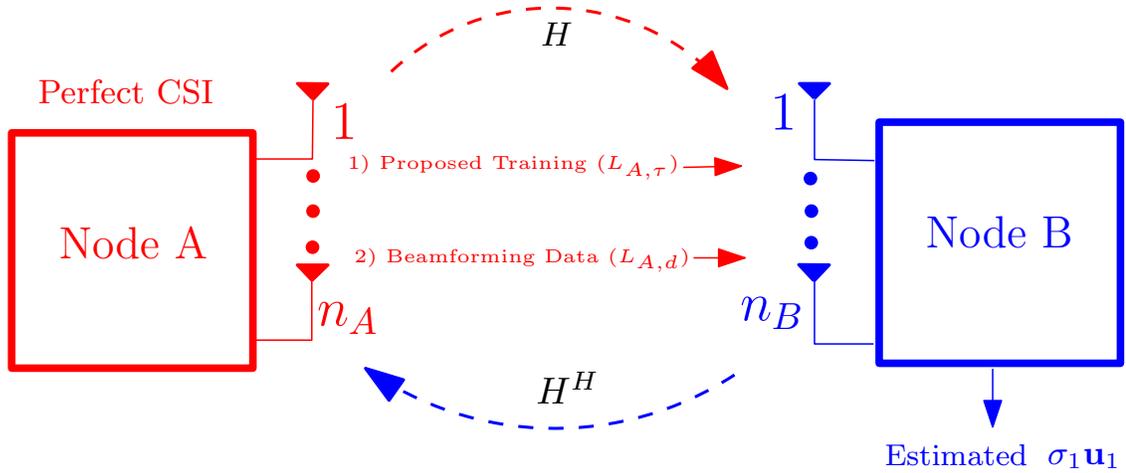
where  $\rho_{\text{eff}} \triangleq \frac{2\rho L_c}{(\sqrt{2(L_c-1)+1})^2}$ .

*Proof:* See Appendix C.0.10. ■

*Remark:* Theorems 5 and 6 suggest that given a constraint on the total energy available for communication, it is optimal to allocate the minimum possible time duration of one symbol for training. It is interesting to note that, as  $L_c$  gets large, the optimal data and training energy per symbol in (4.19) approach  $\rho$  and  $\rho\sqrt{L_c/2}$ , respectively. In the previous subsection, we had shown that the optimal reverse training duration scales with the coherence time as  $\sqrt{\beta L_c}$ . Here, we see that, for the energy efficient design, the optimal training duration remains fixed at one symbol, but the optimal training energy scales as  $\sqrt{L_c}$ . This is a consequence of the total energy constraint imposed under this design.

## 4.4 Forward-Link Training With Perfect CSIA

In this section, assuming perfect CSI at *node A* (i.e., *node B* initiated training), we consider *node A* transmitting both training and data to *node B*. In particular, *node A* transmits data along the dominant mode of the channel [6]. This requires the knowledge of  $\sigma_1 \mathbf{u}_1$  at *node B* for data detection, which is conveyed via a training scheme similar to the one proposed in Sec. 4.3.1. The transmission protocol is shown in Fig. 4.2. It consists of a forward-link training phase of duration  $L_{A,\tau}$  symbols followed by BF data transmission of duration  $L_{A,d} = L_c - L_{A,\tau}$  symbols. In this scenario, since resources such as the duration and power for training and data transmission are all allocated at *node A*, an optimal sharing of the same is necessary. Towards this, we derive a tractable, approximate lower bound on the capacity of the system with the proposed channel-dependent

Figure 4.2: *Node B* initiated training.

training scheme. We then find the optimal sharing of resources that maximizes the approximate capacity lower bound. Note that the model here is different from that in Sec. 4.3, where we had assumed perfect CSI at *node B* and studied the implications of our proposed training from *node B* to *node A* on the data transmission from *node A* to *node B*.

### Estimation at *Node B*

Since *node A* has perfect CSI, it uses BF along the vector  $\mathbf{v}_1$  to transmit the training signal to *node B*. From the received training signal

$$\mathbf{y}_{B,\tau} = \sqrt{P_{A,\tau} L_{A,\tau}} \sigma_1 \mathbf{u}_1 + \mathbf{w}_{B,\tau},$$

*node B* computes an MMSE estimate of  $\mathbf{b} = \sigma_1 \mathbf{u}_1$ , denoted by  $\hat{\mathbf{b}}_{\text{mmse}}$ . We use the MMSE estimate because it facilitates the use of the worst case noise theorem in [8] to obtain a

capacity lower bound. Let

$$\mathbf{b} = \hat{\mathbf{b}}_{\text{mmse}} + \tilde{\mathbf{b}}_{\text{mmse}},$$

where  $\tilde{\mathbf{b}}_{\text{mmse}}$  is the estimation error.

### Data Transmission from Node A to Node B

Since *node A* has perfect CSI, it transmits data using BF along the dominant mode of the channel as

$$\mathbf{x}_{A,d} \triangleq \sqrt{P_{A,d}} \mathbf{v}_1 x_{A,d}.$$

The corresponding received data signal is given by

$$\mathbf{y}_{B,d} = \sqrt{P_{A,d}} \sigma_1 \mathbf{u}_1 x_{A,d} + \mathbf{w}_{B,d}, \quad (4.22)$$

$$= \sqrt{P_{A,d}} \hat{\mathbf{b}}_{\text{mmse}} x_{A,d} + \mathbf{n}_{\text{eff}}. \quad (4.23)$$

where  $\mathbf{n}_{\text{eff}} \triangleq \sqrt{P_{A,d}} \tilde{\mathbf{b}}_{\text{mmse}} x_{A,d} + \mathbf{w}_{B,d}$ . Note that  $\mathbb{E} \|\mathbf{x}_{A,d}\|_2^2 = P_{A,d}$  since  $\mathbb{E} |x_{A,d}|^2 = 1$ .

By the orthogonality property of MMSE estimation,  $\mathbb{E}\{x_{A,d}^* \mathbf{n}_{\text{eff}} | \mathbf{y}_{B,\tau}\} = 0$ , and hence, premultiplying (4.23) by  $\frac{\hat{\mathbf{b}}_{\text{mmse}}^H}{\|\hat{\mathbf{b}}_{\text{mmse}}\|_2}$  and using the worst case noise theorem [8], we get the following lower bound on the capacity:

$$C_{AB,L} \triangleq \frac{L_c - L_{A,\tau}}{L_c} \mathbb{E} \log(1 + \text{SNR}_{\text{eff}}), \quad (4.24)$$

where

$$\text{SNR}_{\text{eff}} \triangleq \frac{P_{A,d} \|\hat{\mathbf{b}}_{\text{mmse}}\|_2^2}{\frac{P_{A,d}}{n_B} \mathbb{E} \left\{ \|\tilde{\mathbf{b}}_{\text{mmse}}\|_2^2 \middle| \mathbf{y}_{B,\tau} \right\} + 1},$$

and the expectation in (4.24) is taken with respect to the distribution of  $\mathbf{y}_{B,\tau}$ . Now, the goal is to find the optimal training power and duration to maximize the capacity lower

bound in (4.24), subject to the transmit power constraint at *node A*. However, the distribution of the vector  $\mathbf{b} \triangleq \sigma_1 \mathbf{u}_1$  is complicated in the general MIMO case, due to which, a closed form expression for  $\hat{\mathbf{b}}_{\text{mmse}}$  is hard to find. This, in turn, makes it difficult to directly optimize the capacity lower bound. Therefore, we consider a similar approximation to the capacity lower bound as in [31] to obtain the following performance metric:

$$C_{AB,L} \approx C_{A,B,\text{approx}} \triangleq \frac{L_c - L_{A,\tau}}{L_c} [\mathbb{E}\sigma_1^2]^2 \text{SNR}_L, \quad (4.25)$$

where

$$\text{SNR}_L \triangleq \frac{P_{A,\tau} L_{A,\tau} P_{A,d}}{(P_{A,d} + P_{A,\tau} L_{A,\tau}) \mathbb{E}\sigma_1^2 + n_B}. \quad (4.26)$$

The derivation of the above is detailed in Appendix C.0.11. The approximation facilitates the optimization of the training and data powers in closed form in the following subsection.

### Optimal Training Power and Duration Allocation

Let  $\rho$  be the average power constraint at *node A*. Here, we solve the following optimization problem:

$$\max_{P_{A,d}, P_{A,\tau}, L_{A,\tau}, L_{A,d}} \frac{L_c - L_{A,\tau}}{L_c} \text{SNR}_L, \quad (4.27)$$

subject to  $P_{A,d} L_{A,d} + P_{A,\tau} L_{A,\tau} = \rho L_c$  and  $L_{A,d} + L_{A,\tau} = L_c$ .

As in the previous section, without loss of global optimality, we first maximize  $\text{SNR}_L$  with respect to  $P_{A,d}$  and  $P_{A,\tau}$  since it enters into the expression in (4.27) only through  $\text{SNR}_L$ . Theorems 7 and 8 present the solution for the optimal training power and training duration, respectively.

**Theorem 7.** For a given training duration, the optimal training and data powers that solve the problem in (4.27) are  $P_{A,d}L_{A,d} = \alpha^*\rho L_c$  and  $P_{A,\tau}L_{A,\tau} = (1 - \alpha^*)\rho L_c$ , where

$$\alpha^* = \begin{cases} \theta - \sqrt{\theta(\theta - 1)} & \text{if } L_{A,d} > 1, \\ 0.5 & \text{if } L_{A,d} = 1, \end{cases} \quad (4.28)$$

and

$$\theta \triangleq \frac{L_{A,d}}{L_{A,d} - 1} + \frac{n_B L_{A,d}}{\mathbb{E}\sigma_1^2 \rho L_c (L_{A,d} - 1)}. \quad (4.29)$$

Also, the corresponding expression for optimal  $C_{A,B,\text{approx}}$ , denoted  $C_{A,B,\text{approx}}^*$ , becomes

$$C_{A,B,\text{approx}}^* \triangleq \begin{cases} \frac{\rho \mathbb{E}\sigma_1^2 L_{A,d}}{(L_{A,d} - 1)} \left( \sqrt{\theta} - \sqrt{\theta - 1} \right)^2 & \text{if } L_{A,d} > 1, \\ \frac{\mathbb{E}(\sigma_1^2)^2}{4L_c} \frac{(\rho L_c)^2}{n_B + \mathbb{E}\sigma_1^2 \rho L_c} & \text{if } L_{A,d} = 1. \end{cases} \quad (4.30)$$

*Proof:* See Appendix C.0.12. ■

We now solve for the optimal  $L_{A,d}$  that maximizes the objective function in (4.30) in the following Theorem.

**Theorem 8.** The data duration that maximizes  $C_{A,B,\text{approx}}^*$  in (4.30) subject to  $1 < L_{A,d} \leq L_c - 1$  is  $L_{A,d} = L_c - 1$ .

*Proof:* See Appendix C.0.13. ■

*Remark:* At low SNR, i.e., as  $\rho \rightarrow 0$ ,  $\alpha \rightarrow \frac{1}{2}$  and  $C_{A,B,\text{approx}}^*$  is quadratic in  $\rho$ . Thus, at low SNR, it is optimal to use just one training symbol, but expend half the total available power on training. It is interesting to note that this result coincides with that obtained at low SNR with SM of data and no CSIT (see [8], Corollary 1).

## 4.5 *Node A* Initiated Training Versus *Node B* Initiated Training

The goal of this section is to compare the schemes in Secs. 4.3 and 4.4 in terms of the approximate capacity lower bound, at low data SNR. We wish to determine whether it is better to have perfect CSI at *node A*, that would allow it to perform *ideal transmit beamforming*, but with *node B* using an *estimated* beamforming vector to receive and decode the data signal; or to have perfect CSI at *node B*, that would allow it to perform *ideal receive beamforming*, but with *node A* transmitting data using an *estimated* beamforming vector.

At first glance, it appears that having perfect CSI at *node A* is beneficial as it allows *node A* to exploit the CSI to accurately transmit the beamformed data and optimize the capacity lower bound. However, acquiring the perfect CSI using orthogonal training inherently involves a minimum time overhead that depends on the number of antennas at the nodes. If *node B* has a significantly larger number of antennas than *node A*, the time overhead of initiating training at *node B* could offset the potential gains one could obtain by exploiting CSI at *node A*, and vice versa. Moreover, an error in the transmit beamforming vector has a different effect compared to an error in the receive beamforming vector on the achievable data rate. Loosely speaking, this is because, in the former case, the error in the transmit beamforming vector does not change the noise statistics at the receiver, while in the latter case, the error in the receive beamforming vector multiplies both the data and noise components of the received signal. The analysis presented below determines which of the two options offers a better performance.

### 4.5.1 Node A Initiated Training

Consider the *node A* initiated training and data transmission scheme proposed in Sec. 4.3, and as depicted in Fig. 4.1. Let the average per-symbol power constraint be  $\rho$  units. Since acquisition of perfect CSIB inherently requires  $n_A$  initial training symbols from *node A*, the time remaining for RCT and data transmission is  $L_c - n_A$  symbols, and the total available power is  $\rho(L_c - n_A)$ . We use  $\rho_{\text{eff}}\mathbb{E}\sigma_1^2$  in (4.21), scaled by the fraction of the time spent for data transmission, as the performance metric. This corresponds to a low SNR approximation of  $C_{B,A,\text{approx}}^*$  in (4.21). The performance metric is given by

$$S_A \triangleq \frac{2(L_c - n_A - 1)}{(\sqrt{2(L_c - n_A - 1)} + 1)^2} \rho \mathbb{E}\sigma_1^2, \quad (4.31)$$

where we have used  $L_{A,d} = L_c - n_A - 1$ .

### 4.5.2 Node B Initiated Training

Consider the training and data transmission scheme proposed in Sec. 4.4, and as depicted in Fig. 4.2. In this case, using (4.30), the approximate capacity lower bound  $C_{A,B,\text{approx}}^*$  scaled by the fraction of the time spent for data transmission is

$$S_B \triangleq \rho \mathbb{E}\sigma_1^2 \frac{L_{A,d}}{(L_{A,d} - 1)} \left( \sqrt{\theta} - \sqrt{\theta - 1} \right)^2, \quad (4.32)$$

where  $L_{A,d} = L_c - n_B - 1$  and  $\theta$  is as defined in (4.29), with  $L_c$  replaced by  $L_c - n_B$  to factor the initial *node B* to *node A* training overhead of  $n_B$  symbols into (4.32).

Now, we compare the two schemes using (4.31) and (4.32). *Node A* initiated training outperforms *node B* initiated training if  $S_A > S_B$  and vice versa. Assume that  $L_c > \max(n_A, n_B) + 2$ , which ensures that the  $L_c - n_B - 2$  term in the denominator of (4.32) is

positive. Since  $\theta > \frac{L_{A,d}}{L_{A,d}-1}, (\sqrt{\theta} - \sqrt{\theta-1})^2 < \frac{\sqrt{L_{A,d}-1}}{\sqrt{L_{A,d}+1}}$ . Substituting this in (4.32), it can be shown that *node A* initiated training outperforms *node B* initiated training if

$$\frac{2(L_c - n_A - 1)(\sqrt{L_c - n_B - 1} + 1)^2}{(\sqrt{2(L_c - n_A - 1)} + 1)^2(L_c - n_B - 1)} \geq 1. \quad (4.33)$$

The above implies that *node A* initiated training is better if  $L_c \geq 2n_A - n_B + 1$ .

As  $\rho \rightarrow 0$ , it can be shown that  $S_B$  in (4.32) is proportional to  $\rho^2$ , while that  $S_A$  in (4.31) is proportional to  $\rho$ . Hence, *node A* initiated training outperforms *node B* initiated training if  $L_c > \max(n_A, n_B) + 2$ . In practice, the channel coherence time will be much larger than  $n_A$  and  $n_B$ , and hence, the conditions  $L_c > \max(n_A, n_B) + 2$  and  $L_c \geq 2n_A - n_B + 1$  will typically be satisfied. Hence, in cases of practical interest, *node A* initiated training outperforms *node B* initiated training.

## 4.6 Simulation Results and Discussion

In this section, we present numerical results to validate the theoretical expressions and illustrate the performance benefits offered by the proposed training scheme. The simulation set up consists of a Rayleigh flat fading channel whose coefficients are assumed to be i.i.d. and drawn from  $\mathcal{CN}(0, 1)$ , and an AWGN noise that is also modeled as i.i.d.  $\mathcal{CN}(0, 1)$ . Throughout this section, except for Sec. 4.6.3, we assume a  $3 \times 3$  TDD-MIMO system with  $L_c = 100$ . The parameter values are chosen to illustrate the various trade-offs involved in RCT; in general, the performance advantage of the proposed scheme over orthogonal training is larger for smaller  $L_c$ .

### 4.6.1 Node A Initiated Training

In this subsection, we use the exact lower bound on the capacity in (4.10) to evaluate the performance of various schemes.

To illustrate that the minimum training duration of one symbol is not always optimal, we plot the exact capacity lower bound versus data and training powers in Figs. 4.3 and 4.4, respectively. It can be seen from the two figures that  $L_{B,\tau} = 1$  performs poorly compared to the low SNR optimal solution derived in (4.17) whenever the RCT power is small compared to the data power. This is because, when  $\beta \triangleq P_{A,d}/2P_{B,\tau}$  is large, the term in the denominator of (4.12) has a dominant effect on the data rate compared to the  $(L_c - L_{B,\tau})/L_c$  term. Hence, in order to improve the data rate,  $L_{B,\tau}$  has to be increased, and the optimal  $L_{B,\tau}$  is greater than 1. Next, to capture the effect of errors introduced due to quantization of the beamforming vectors in a fixed-point arithmetic implementation, in Fig. 4.4, we plot the capacity lower bound in (4.10) when the BF vectors at *node A* and *node B* are quantized. We consider a uniform scalar quantizer, with  $B = 4$  and 6 bits of quantization per real dimension of the BF vector. We see that although there is a degradation in the data rate when  $B = 4$  bits, this performance loss becomes insignificant when  $B = 6$  bits. Further, the degradation in the data rate increases with the training power. This is expected, because at high training power, the residual error due to quantization dominates, and the error due to noise in the training phase becomes insignificant.

The training length can be optimized in three ways: (i) numerically optimizing the exact lower bound in (4.10), (ii) numerically optimizing the approximate capacity lower bound in (4.12), and (iii) the low SNR optimal  $L_{B,\tau}$  in (4.17). We compare these three

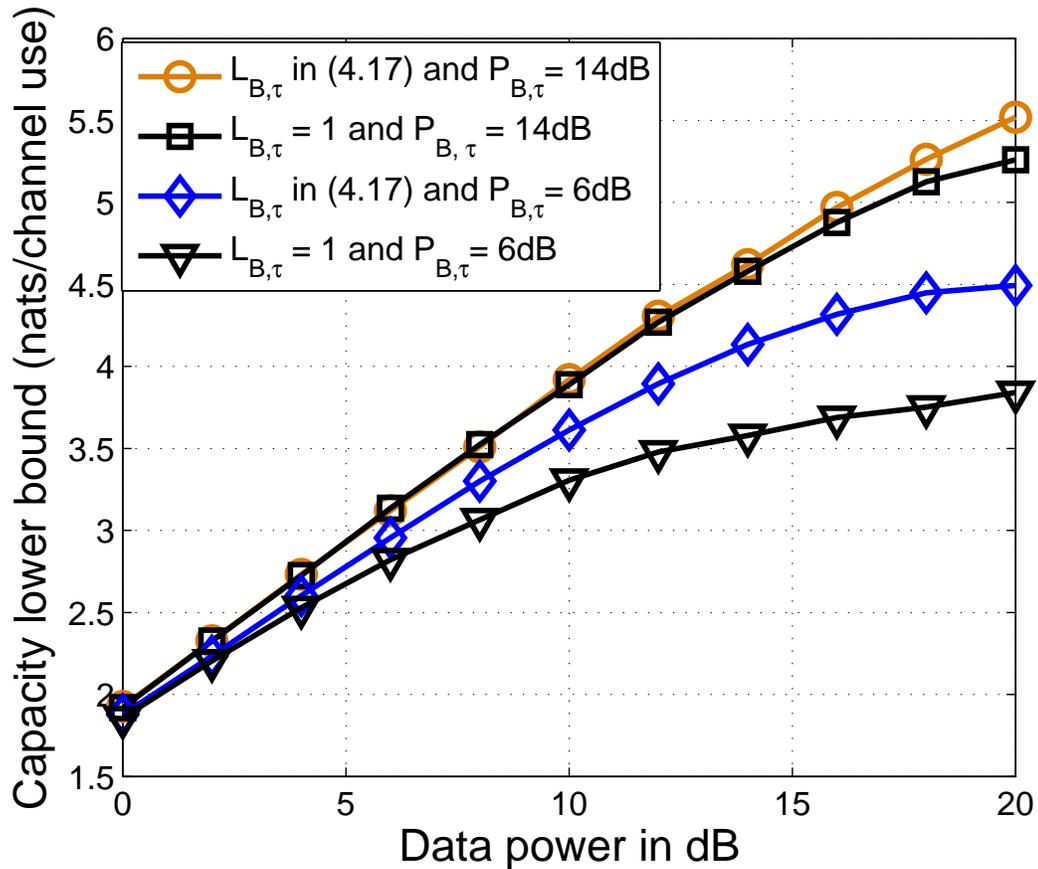


Figure 4.3: Capacity lower bound in (4.10) versus data power for  $P_{B,\tau} = 14$  and 6dB. Here, we have used a  $3 \times 3$  MIMO system with  $L_c = 100$ . The figure demonstrates that the optimal  $L_{B,\tau}$  is not always equal to one.

solutions for various values of  $P_{A,d}$  with  $P_{B,\tau} = 2P_{A,d}$  in Table 4.1. At low  $P_{A,d}$  all three cases coincide, and for  $P_{A,d} \geq 4$  dB, the low SNR optimal solution deviates from the optimal  $L_{B,\tau}$  since the low SNR assumption is no longer valid. However, it is interesting to note that for all  $P_{A,d}$  considered, the training duration obtained by optimizing the approximate capacity lower bound in (4.12) coincides with the optimal solution. This further justifies the use of the approximate lower bound derived in this chapter. Moreover, as can be seen from Fig. 4.5, the data rate obtained from the low SNR optimal  $L_{B,\tau}$

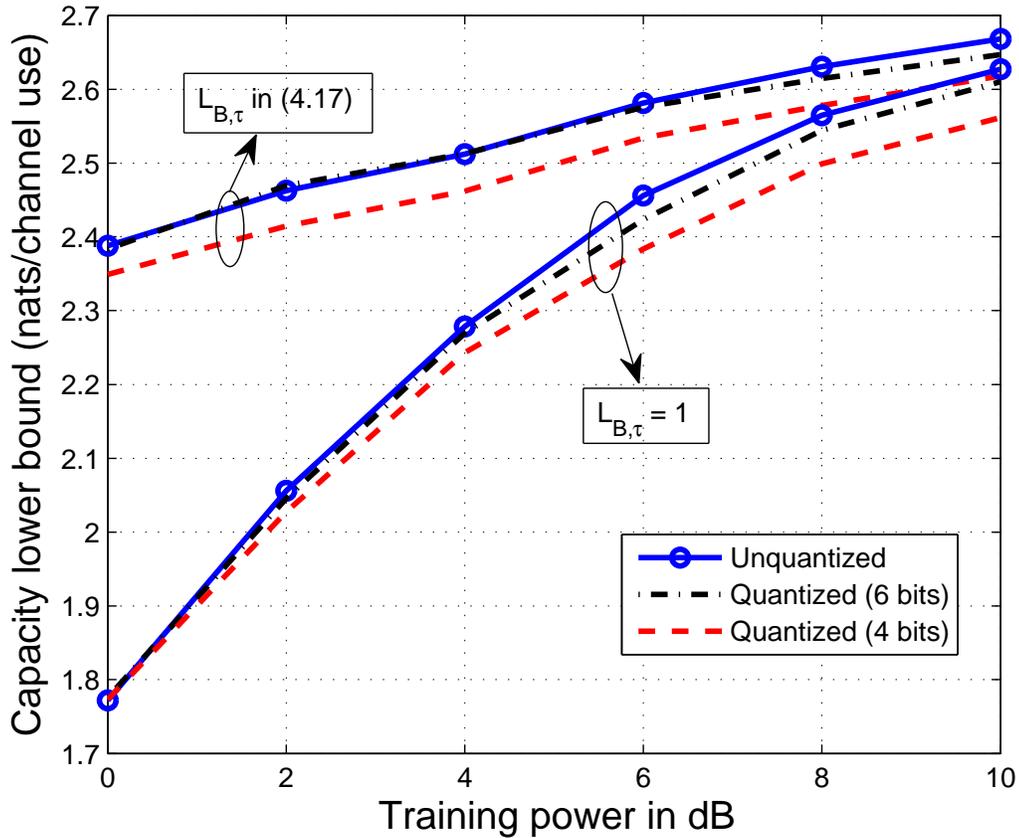


Figure 4.4: Capacity lower bound in (4.10) versus training power with  $P_{A,d} = 4\text{dB}$  for a  $3 \times 3$  MIMO system, and  $L_c = 100$ . The figure demonstrates that  $L_{B,\tau} = 1$  is not always optimal. It also illustrates that quantization of the BF vectors in a fixed-point implementation has a negligible effect on the performance for  $B \geq 6$  bits per real dimension.

in (4.17) approximates that obtained by numerically optimizing the exact lower bound in (4.10) very well, despite the differences in the value of  $L_{B,\tau}$ .

The conventional orthogonal training scheme (e.g., [4, 5]) is compared with the proposed training scheme in Fig. 4.6, where the capacity lower bound versus data power (equal to the training power) is plotted. For fair comparison, we have used the same

Table 4.1: Optimum training duration for a  $3 \times 3$  MIMO system with  $L_c = 100$  and  $P_{B,\tau} = 2P_{A,d}$ .

Data power $P_{A,d}$	2dB	4dB	6dB	8dB	10dB	12dB
$L_{B,\tau}$ using (4.10)	5	3	3	3	3	2
$L_{B,\tau}$ using (4.12)	5	3	3	3	3	2
$L_{B,\tau}$ in (4.17)	5	5	5	5	5	5

training duration values of 3 and 9 symbols for both the proposed and the orthogonal training schemes. The proposed scheme outperforms the orthogonal scheme by over 7dB at all SNRs. This gain is because the proposed scheme directly estimates the dominant mode of the channel, resulting in a significantly improved estimation accuracy compared to the orthogonal scheme. Also, note that the approximate lower bound derived in (4.12) is tight at all data powers.

For a fixed training duration of one symbol, Fig. 4.7 shows a plot of the capacity lower bound (exact and approximate) versus the average power per symbol ( $\rho$ ) for the energy efficient scheme for sharing of resources between *node A* and *node B* described in Sec. 4.3.3. This figure also illustrates the performance loss from setting a suboptimal value of  $\alpha$ , the fraction of the total power used for data transmission. For the sake of comparison, we use the following four different values of  $\alpha$ : (i) the optimal value of  $\alpha^* = \frac{\sqrt{2L_{A,d}}}{\sqrt{2L_{A,d}+1}} = 0.9336$  for  $L_{A,d} = L_c - 1 = 99$ , (ii)  $\alpha = \frac{L_c-1}{L_c} = 0.99$  (i.e.,  $P_{B,\tau} = P_{A,d}$ ), (iii)  $\alpha = 0.5$  (i.e.,  $P_{A,d}L_{A,d} = P_{B,\tau}L_{B,\tau}$ ), and (iv)  $\alpha = 0.1$ . It is clear from the figure that the optimal  $\alpha = 0.9336$  from Theorem 5 results in a significant improvement in the data

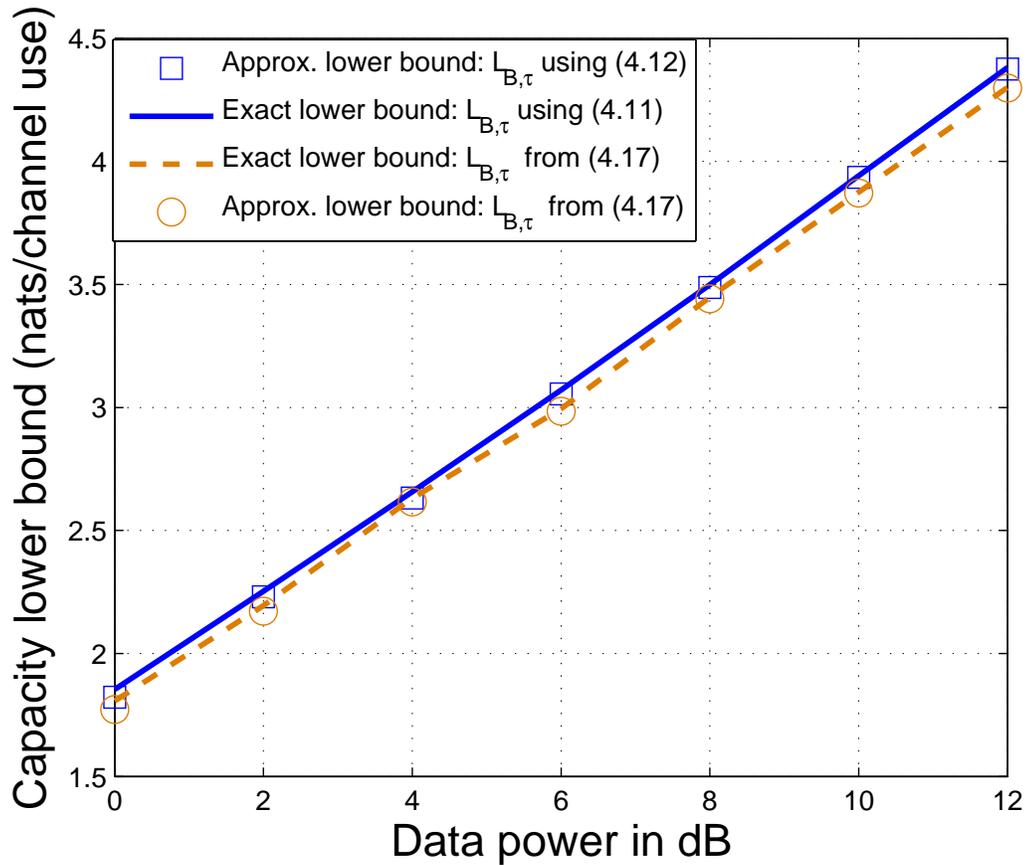


Figure 4.5: Capacity lower bound versus data power for a  $3 \times 3$  MIMO system with  $L_c = 100$  and  $P_{B,\tau} = 2P_{A,d}$ . Using the low SNR optimal  $L_{B,\tau}$  is nearly optimal for a wide range of training/data powers. Also, the loss in the data rate due to using the low SNR approximation is negligible.

rate compared to other values of  $\alpha$ . Note that the approximate lower bound is tight for almost all cases except for the case of  $\alpha = 0.99$  where the gap is approximately 0.3 nats at 0dB. This is because the ratio  $P_{A,d}/P_{B,\tau} = 1$ , and due to this, the higher order terms ignored in the approximation makes the proposed lower bound loose.

### 4.6.2 Node B Initiated Training

Figure 4.8 shows a plot of  $C_{A,B,\text{approx}}^*$  in (4.25) versus  $\rho$ , the total training power for *node B* initiated training described in Sec. 4.4. We compare the performance for the following different values of  $\alpha$ : (i) the optimal  $\alpha^* = 0.93$  from (4.28), (ii)  $\alpha = 0.99$  (i.e.,  $P_{A,\tau} = P_{A,d}$ ), (iii)  $\alpha = 0.5$  (i.e.,  $P_{A,d}L_{A,d} = P_{A,\tau}L_{A,\tau}$ ), and (iv)  $\alpha = 0.2$  (very less power for data transmission). The superior performance of the optimal  $\alpha$  compared to the other values of  $\alpha$  is clear from the graph.

### 4.6.3 Node A Initiated Training Vs. Node B Initiated Training

Figure 4.9 plots  $S_A$  and  $S_B$  in (4.31) and (4.32), corresponding to *node A* initiated training and *node B* initiated training, scaled by  $\rho\mathbb{E}\sigma_1^2$ , versus the coherence time for an average power of  $\rho = 0\text{dB}$  and  $10\text{dB}$ . Due to the scaling, the performance of *node A* initiated training at  $\rho = 0\text{dB}$  and  $10\text{dB}$  are nearly the same, making the curves easy to visualize on a single plot. Here, we have used an  $n_A = 8$  and  $n_B = 5$  MIMO system. From Sec. 4.5, *node A* initiated training is better than *node B* initiated training if  $L_c \geq 2n_A - n_B + 1$ , i.e., when  $L_c \geq 12$ . Also, at low SNR, *node A* initiated training outperforms *node B* initiated training for  $L_c > \max(n_A, n_B) + 2$ , i.e., when  $L_c > 10$ . In the figure, it can be seen that at both  $\rho = 0\text{dB}$  and  $10\text{dB}$ , *node A* initiated training outperforms *node B* initiated training for  $L_c \geq 12$ . At  $\rho = 0\text{dB}$ , *node A* initiated training outperforms *node B* initiated training for  $L_c = 11$  also. Thus, the figure corroborates our theoretical predictions on the conditions for *node A* initiated training to outperform *node B* initiated training.

## 4.7 Conclusions

In this chapter, we considered the design and analysis of a channel-dependent training signal for a TDD-MIMO system with two nodes, *node A* and *node B*, and with data transmission over the dominant mode of the channel from *node A* to *node B*. We assumed that one of the nodes has perfect channel knowledge, and proposed a scheme for estimating the dominant singular vector at the other node. We derived a lower bound on the forward-link capacity and used it to optimize the training power and duration. When *node B* has perfect CSI, with independent power constraints at *node A* and *node B*, it is optimal to spend the minimum possible time for training (i.e.,  $L_{B,\tau} = 1$ ), provided the training power is of the order of one quarter of the total data transmit power. Otherwise, the optimal  $L_{B,\tau} > 1$ . From an energy efficiency perspective, regardless of the available energy,  $L_{B,\tau} = 1$  is optimal, and the fractional power spent on data transmission is  $\frac{\sqrt{2(L_c-1)}}{\sqrt{2(L_c-1)+1}}$ , where  $L_c$  is the channel coherence time. Also, when *node A* has perfect CSI, at low SNR,  $L_{A,\tau} = 1$  is optimal. We showed that if  $L_c \geq 2n_A - n_B + 1$ , which is typically the case in practical systems, initiating training at *node A* is better than initiating training at *node B* in terms of an approximate capacity lower bound. Monte Carlo simulations validated the theoretical expressions and illustrated the performance benefits offered by the proposed channel-dependent training scheme compared to conventional channel-agnostic orthogonal RCT. In the following chapter, we extend the fixed-power RCT scheme presented in this chapter to a PCRCT scheme for TDD-MIMO multiuser spatial multiplexing systems.

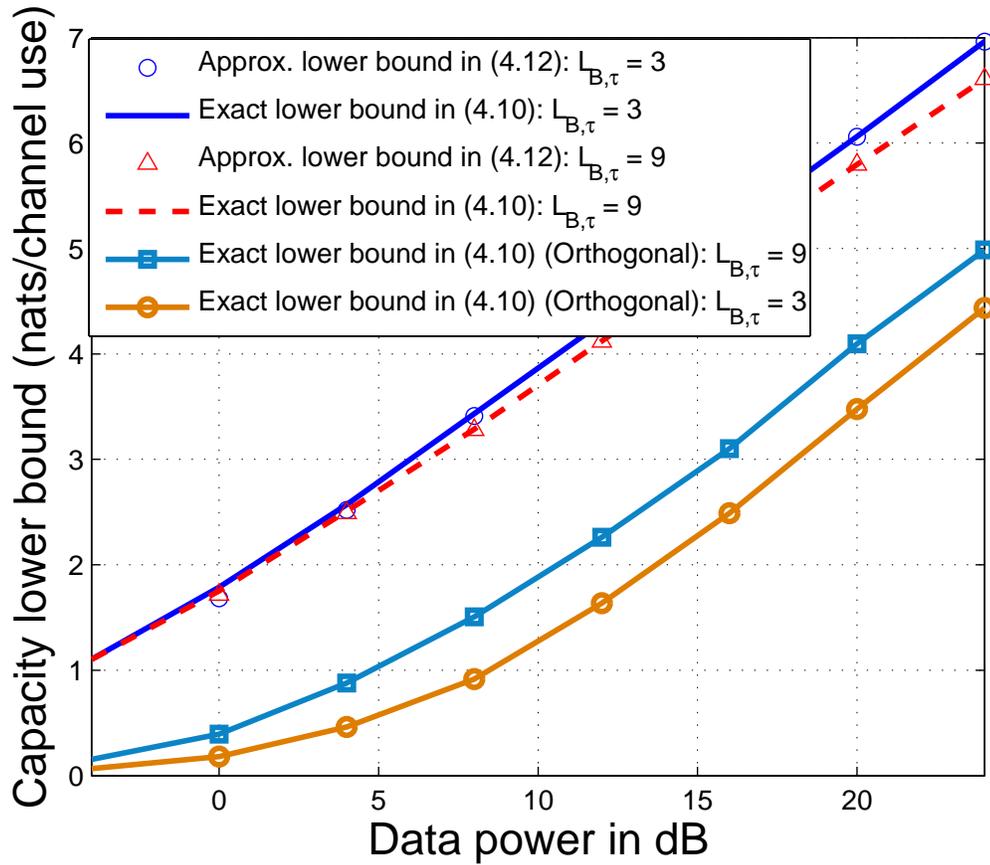


Figure 4.6: Capacity lower bound (exact and approximate) versus data power for a  $3 \times 3$  MIMO system with training power  $P_{B,\tau} = P_{A,d}$  and  $L_c = 100$ . The figure illustrates the performance gain offered by the proposed training method over the conventional orthogonal training scheme (e.g., [4, 5]), and shows that the approximate lower bound is tight at all data powers.

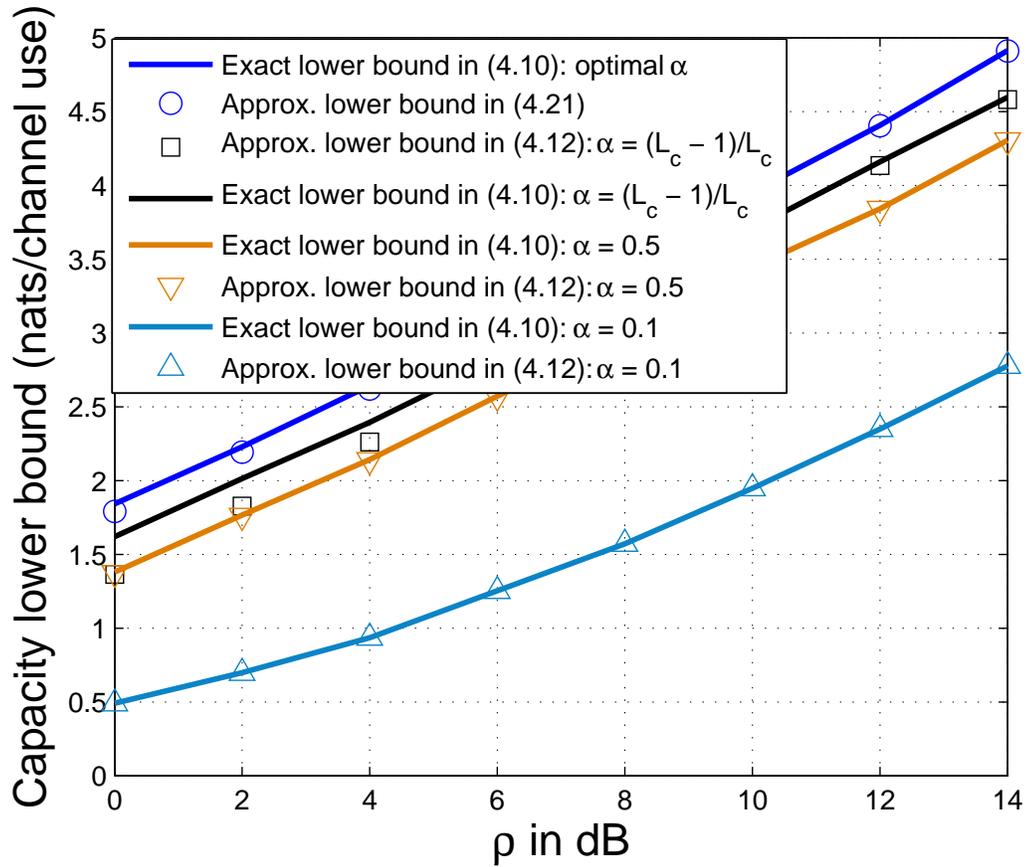


Figure 4.7: Exact and approximate capacity lower bound versus the average power ( $\rho$ ) for the energy efficient resource sharing scheme analyzed in Sec. 4.3.3, with  $L_{B,\tau} = 1$ . The figure illustrates the tightness approximate bound and that the sharing of powers derived in Theorem 7 is optimal.

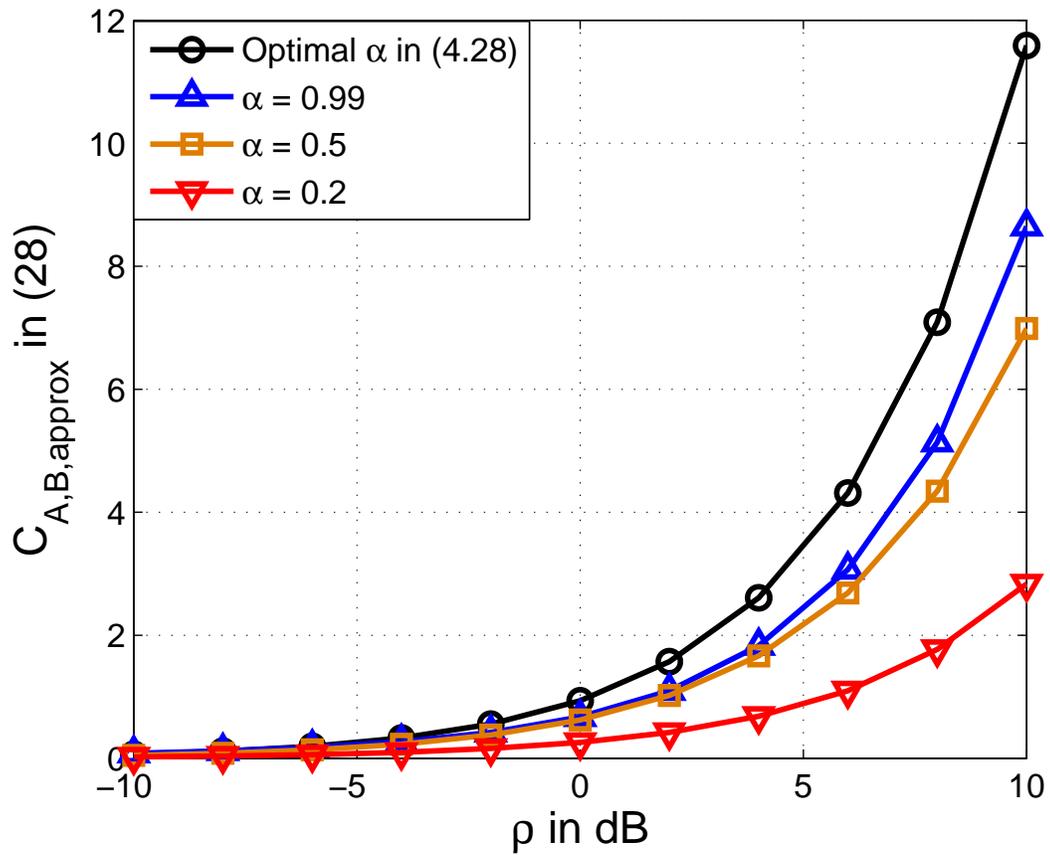


Figure 4.8:  $C_{A,B,approx}$  in (4.25) versus the average power ( $\rho$ ) for a  $3 \times 3$  MIMO system for the *node B* initiated training scheme.

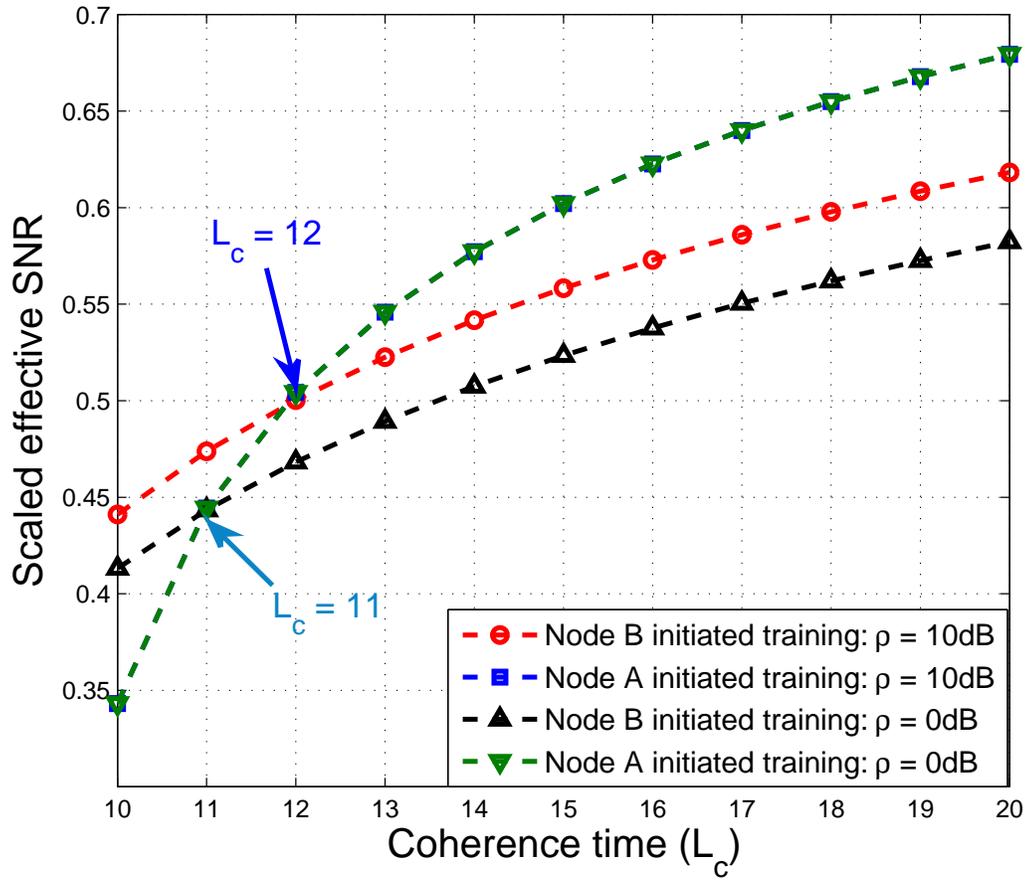


Figure 4.9: Plot of  $S_A$  and  $S_B$  in (4.31) and (4.32) corresponding to *node A* initiated training and *node B* initiated training scaled by  $\rho\mathbb{E}\sigma_1^2$  versus the coherence time ( $L_c$ ) for an  $n_A = 8, n_B = 5$  MIMO system with  $\rho = 0\text{dB}$  and  $10\text{dB}$ .

## Chapter 5

# Reverse Channel Training in a Multi-User TDD-MIMO Spatial Multiplexing System

In the recent years, the use of multiple antennas has emerged as one of the promising technologies for wide band multiuser communication (e.g., IEEE 802.16a, IEEE 802.20 and 4G protocol candidates) as it offers a significant improvement compared to single antenna systems, in terms of reliability and throughput. However, these benefits are realizable only when both the transmitter (Base Station (**BS**)) and receivers (users) have accurate and up-to-date Channel State Information (CSI). Thus, one of the important problems in designing multiple antenna systems is the fast and accurate acquisition of CSI both at the transmitter (**BS**) and receivers (users). CSI can be obtained at the users simultaneously by sending a known training sequence in the forward-link from the transmitter (BS). In Time Division Duplex (TDD) systems, exploiting the reciprocity of the channel, CSI at the Transmitter (CSIT), i.e., at the **BS**, can be acquired by sending a known training signal in the reverse-link, also known as Reverse Channel Training

(RCT). However, as the number of transmit antennas and/or the number of users become large, the overhead due to training can become prohibitive, especially in vehicular or mobile communications, where the channel is relatively fast varying, since the training duration is proportional to the product of number of antennas at each user terminal and the number of users. When CSI is available at the Receiver (CSIR), one can potentially exploit it to design the RCT sequence and selectively feedback only the required part of the CSI to the **BS**. This could result in faster and/or more accurate acquisition of the CSI at the **BS**, leading to an improvement in the effective data rate and/or a reduction in the power required for data transmission; this is the focus of this chapter.

The main body of the existing literature on CSIT acquisition in single and multiuser TDD systems focuses on orthogonal RCT [3–5,44,50], where an orthogonal training sequence such as the scaled identity matrix is employed. The method employed in [45] and [46] to acquire CSIT is to feedback a scaled version of the received forward-link training signal, from which the transmitter estimates the entire channel matrix. Although this outperforms orthogonal RCT, it has the disadvantage that the transmitter estimates the entire channel matrix, which is not required for certain types of data transmission schemes such as Beamforming (BF) or Spatial Multiplexing (SM) along the dominant modes of the channel. Data-aided blind estimation of the dominant BF vector is proposed in [47,48,51]. Reference [42] proposes a two stage protocol consisting of conventional RCT followed by quantized CSI transmission in the reverse-link. A channel-dependent RCT scheme was independently explored in [31], in the context

of a Single-Input Multiple-Output (SIMO) channel, and the sharing of energy spent between training and data was optimized with respect to an approximate expression for the forward-link data SNR.

In this chapter, we consider an SM system with equal power allocation across the  $m$  dominant modes of the channel during data transmission with perfect CSIR and noisy CSIT obtained via RCT. Equal power allocation across modes is known to be nearly optimal for all but low data SNR [52]. The perfect CSIR assumption helps simplify the analysis and isolates the effect of estimation errors in the RCT on the performance. This assumption is common in studies that focus on the achievable data rate or outage probability [4, 24, 39, 50, 53, 54]. Also, we note that the perfect CSIR is required for the analysis to be tractable; however, our proposed RCT scheme is applicable even when the CSI at the receiver is not perfect. Now, SM-based data transmission over  $m < n_A$  dominant modes of the channel only requires knowledge of the  $m$  dominant right singular vectors of the forward-link channel at the transmitter, and not the entire channel matrix, where  $n_A$  is the number of antennas at the **BS**/transmitter. Motivated by this, in this chapter, we explore a novel, channel-dependent, power-controlled RCT scheme that enables the **BS** to estimate *only* the part of the channel that is required for data transmission. The structure of the RCT scheme we consider here is different from the RCT schemes considered in previous chapters, and it allows for both spatial and temporal allocation of the training power. Due to this, the design considerations and the corresponding performance analysis are completely different. The following are our main contributions:

- *Proposed RCT*: We propose an RCT that allows the transmitter to directly estimate

the dominant eigenmodes of the channel required for data transmission. Further, the proposed RCT allows one to allot the power both spatially (across modes) and temporally (across time), while satisfying an average training power constraint. Both the spatial power allocation matrix  $D$  and the temporal power control parameter  $\phi_c > 0$  are optimized using the following two performance metrics: (i) the Mean Square Error (MSE) in the estimated precoding matrix at *node A*, and (ii) a Capacity Lower Bound (CLB) on the downlink data transmission.

- *Optimal RCT with approximate MSE as a metric:* With the approximate MSE in the estimated singular vectors as the performance metric, we obtain an analytical solution for the *optimal* spatial power allocation matrix  $D$  and the optimal temporal power control  $\phi_c$ , as a function of the channel singular values. We show, using simulations, that the optimal  $D$  achieves a lower MSE compared to using equal power allocation across the eigenmodes. On the other hand, temporal power allocation across channel instantiations results in only a marginal benefit in the MSE compared to constant power training over time. For example, in a  $3 \times 4$  MIMO system, the proposed training scheme offers an improvement of over 15dB in the training power required to achieve the same MSE, compared to the orthogonal RCT.
- *Optimal RCT with approximate CLB as a metric:* Here, we analytically optimize spatio-temporal power allocation of the RCT scheme to maximize the approximate CLB. In the  $3 \times 4$  example mentioned above, using the optimal  $D$  outperforms using equal spatial power allocation by approximately 1 bit/channel use at around 16dB of training and data power, while temporal power allocation offers

only a marginal improvement in the data rate over equal temporal power allocation. We also illustrate that using a larger number of modes is not always optimal when the overhead due to training is taken into account.

- *Multiuser case:* We extend the proposed channel-dependent RCT to a multiuser downlink scenario with  $M$  user terminals. We use the approximate CLB derived in this chapter as the metric for user scheduling. In the case of BF with  $m = 1$  mode, we derive an upper bound on the sum data rate with CSIT obtained using the proposed channel-dependent RCT, and show that it scales with  $M$  as  $\frac{L_c - L_{B,\tau}}{L_c} \log_2 \log M$ , where  $L_c$  and  $L_{B,\tau}$  are the coherence time of the channel and the training duration, respectively. Through simulations, we illustrate the benefits of the proposed scheme over conventional orthogonal RCT.

We note that our study of power-controlled, channel-dependent RCT is fundamentally different from past work on reverse-link training for MIMO SM systems, which is based on constant-power, channel-agnostic orthogonal RCT. Our proposed RCT scheme can lead to significant performance improvements over orthogonal RCT both in terms of the achievable data rate as well as the MSE in BF vector estimation, especially when the channel is fast varying. Moreover, from a system designer's point of view, it is useful to know that spatial allocation of the available training power is much more beneficial than temporal allocation.

The rest of this chapter is organized as follows. The system model is described in Sec. 5.1. The proposed training scheme and channel estimation procedure are explained in Sec. 5.2. The training sequence is optimized in Sec. 5.2.3. The performance of the RCT in the multiuser scenario is presented in Sec. 5.3. Simulation results are provided

in Sec. 5.4, and Sec. 5.5 concludes the chapter.

We use the following notation. We use  $\mathbb{E}_{|H}[\cdot]$  to denote the expected value of  $[\cdot]$  conditioned on  $H$ .  $I_{m \times n}$ , with  $n \leq m$ , represents the first  $n$  columns of the  $m \times m$  identity matrix. We use  $\mathbf{x} = \mathcal{O}(\mathbf{y})$ ,  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  to mean that the entries of  $\mathbf{x}$  are less than the corresponding entries of  $c\mathbf{y}$  for some  $0 < c < \infty$ .

## 5.1 System Model

The system model consists of a single cell multiuser system with a base station denoted **BS** and  $M$  active user terminals, denoted  $\mathbf{UT}_1, \dots, \mathbf{UT}_M$ . The **BS** has  $n_A$  antennas and each **UT** has  $n_B$  antennas. Denote the MIMO channel from the **BS** to  $\mathbf{UT}_k$  by  $H_k \in \mathcal{C}^{n_B \times n_A}$ . Let  $H_k = U^{(k)} \Sigma^{(k)} (V^{(k)})^H$  be the SVD of  $H_k$ , where the diagonal entries of  $\Sigma^{(k)} \in \mathbb{R}^{n_B \times n_A}$ , denoted  $\sigma_{1,k}, \dots, \sigma_{n,k}$ , are the singular values of  $H_k$ , with  $n \triangleq \text{rank}(H_k)$ , which equals  $\min(n_A, n_B)$  almost surely. Moreover,  $U^{(k)} \in \mathbb{C}^{n_B \times n_B}$  and  $V^{(k)} \in \mathbb{C}^{n_A \times n_A}$  are unitary matrices whose columns are the eigenvectors of  $H_k H_k^H$  and  $H_k^H H_k$ , respectively. The channel is assumed to remain constant for a frame of duration equal to the channel coherence time  $L_c$ , and evolve in an i.i.d. fashion from frame to frame. We assume a TDD mode of operation with perfect reciprocity [47, 55–57], and thus, without loss of generality, the reverse-link channel of the  $k^{\text{th}}$  user is  $H_k^H$  (see [47]). The transmission protocol consists of the following three phases.

- **Phase I:** This phase consists of a downlink training followed by user scheduling. The downlink training is performed by sending a known pilot sequence from the **BS** to all the **UTs**. Using this, each user terminal computes an estimate of their respective channels. Here, we assume that the resulting estimate is error-free, as

in, for example, [53,54]. This facilitates the derivation of a capacity lower bound, and its analytically tractable tight approximation in closed-form, with respect to which the RCT can be optimized. We consider a scheduling scheme where each user computes a metric (see Sec. 5.3) as a function of its CSI. The user with the highest metric is scheduled for data transmission by the **BS**. The selection of the user with the highest metric can be efficiently implemented using decentralized algorithms such as splitting [58,59] and timer-based schemes [60], which incurs very low overhead<sup>1</sup> in terms of power and delay. We assume that one of these schemes is used to pick the best user, and we ignore the overhead involved in user selection. In this setting, we focus on the problem of RCT sequence design to convey the CSI of the selected user to the **BS**.

- **Phase II:** In this phase, the scheduled user terminal, say  $\mathbf{UT}_k$ , transmits a training sequence  $X_{B,\tau}^{(k)}$  in the uplink direction. The baseband equivalent of the received training signal at the **BS**, denoted  $Y_{A,\tau}$ , is given by

$$\textbf{Reverse-link training: } Y_{A,\tau} = H_k^H X_{B,\tau}^{(k)} + W_{A,\tau}. \quad (5.1)$$

The entries of the noise  $W_{A,\tau}$  are assumed to be i.i.d. complex circularly symmetric standard Gaussian distributed, denoted  $\mathcal{CN}(0, 1)$ . From  $Y_{A,\tau}$ , the **BS** computes an estimate of  $V_m^{(k)}$ , the first  $m$  columns of the matrix  $V^{(k)}$ , which is subsequently used for data transmission over the dominant modes of the channel in the downlink, as explained next. Denote the estimate of  $V_m^{(k)}$  by  $\hat{V}_m^{(k)}$ .

---

<sup>1</sup>In fact, the time overhead in best user selection is bounded irrespective of the number of users [60].

- **Phase III:** This phase consists of data transmission from the **BS** to the scheduled user. For the data transmission scheme, we assume SM of data with equal power allocation. Here, the **BS** sends  $m \geq 1$  i.i.d. data streams,  $\mathbf{x}_{A,d}^{(k)} \in \mathbb{C}^m$ , multiplied by the estimate of the  $m$  dominant right singular vectors of  $H_k$  obtained in **Phase II** [7, 39]. The corresponding received signal at **UT** $_k$ , denoted  $\mathbf{y}_{B,d}^{(k)} \in \mathbb{C}^{n_B \times 1}$ , is given by

$$\text{Forward-link data: } \mathbf{y}_{B,d}^{(k)} = \sqrt{\frac{P_{A,d}}{m}} H_k \hat{V}_m^{(k)} \mathbf{x}_{A,d}^{(k)} + \mathbf{w}_{B,d}^{(k)}. \quad (5.2)$$

In the above,  $\mathbb{E} [\mathbf{x}_{A,d}^{(k)} (\mathbf{x}_{A,d}^{(k)})^H] = I_m$  and  $(\hat{V}_m^{(k)})^H (\hat{V}_m^{(k)}) = I_m$  ensure that the data signal satisfies an average power constraint of  $P_{A,d}$ . The entries of the noise vector  $\mathbf{w}_{B,d}^{(k)} \in \mathbb{C}^{n_B}$  are assumed to be i.i.d.  $\mathcal{CN}(0, 1)$ . Note that, when  $m = 1$ , this corresponds to pure BF based data transmission using the estimated dominant right singular vector of the channel. Thus, the scheme considered in this chapter encompasses BF as a special case.

For the above transmission and scheduling scheme, we consider the problem of designing  $X_{B,\tau}$ , with the following two performance metrics: (i) MSE in  $\hat{V}_m^{(k)}$ , and (ii) an achievable downlink data rate. In the following section, we present our proposed channel-dependent training sequence for the scheduled user in the reverse-link, along with a method for estimating the dominant singular vectors at the **BS**.

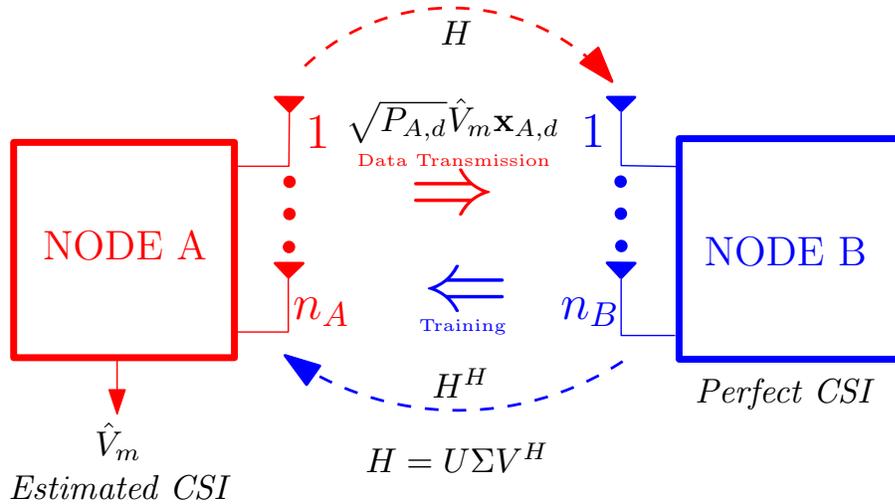


Figure 5.1: The MIMO  $n_B \times n_A$  SM system, showing the RCT and the forward-link data transmission.

## 5.2 Design and Optimization of RCT for the Scheduled User

In this section, we assume that the best user has been scheduled for data transmission using the procedure in Phase I of the protocol described in the previous section, and focus on the design of the RCT sequence to efficiently convey the CSI of the selected user to the **BS** (also see Fig. 5.1). Since the index of the scheduled user does not directly enter the expressions, for the ease of presentation, in this section, we drop the user index.<sup>2</sup> For SM with equal power allocation, the **BS** requires the knowledge of  $V_m \triangleq [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$ , the first  $m$  columns of the right singular matrix  $V$  of the channel  $H$ . We propose the following *channel-dependent, power-controlled RCT sequence* that enables the

<sup>2</sup>For example, instead of writing  $H_k, X_{B,\tau}^{(k)}, U^{(k)}$ , we write  $H, X_{B,\tau}, U$ .

**BS** to directly estimate  $V_m$ :

$$X_{B,\tau} = \sqrt{P_{B,\tau}L_{B,\tau}}\sqrt{\phi_c}UD, \quad (5.3)$$

where  $P_{B,\tau}$  and  $L_{B,\tau}$  are the average training power and training duration, respectively. The temporal power control parameter  $\phi_c \in \mathbb{R}^+$ , the unitary matrix  $U \in \mathbb{C}^{n_B \times n_B}$ , and the spatial power allocation matrix  $D \in \mathbb{R}^{n_B \times m}$  are adapted at the **UT** based on the CSIR. For mathematical tractability,  $D \in \mathbb{R}^{n_B \times m}$  is restricted to be a diagonal matrix with non-negative entries  $d_i$ ,  $i = 1, 2, \dots, m$ , satisfying  $\|D\|_F^2 \leq 1$ . This, along with  $\mathbb{E}\phi_c \leq 1$ , ensures that the training sequence satisfies the average power constraint  $\mathbb{E}\|X_{B,\tau}\|_F^2 \leq P_{B,\tau}L_{B,\tau}$ . The RCT scheme in (5.3) has the following desirable features: (i) for a given channel realization,  $D$  allows the **UT** to selectively allot greater or lesser power for training the different channel eigenmodes (spatial power allocation), (ii) the power control parameter  $\phi_c$  enables the **UT** to perform temporal power allocation.

Note that, in the case of BF, i.e., when  $m = 1$ , the transmitter requires the knowledge of only  $\mathbf{v}_1$ . Using the proposed RCT scheme,  $\mathbf{v}_1$  can be conveyed to *node A* using only one training symbol. In contrast, orthogonal RCT requires at least  $n_B$  training symbols. Hence, the proposed scheme offers savings in terms of the minimum required training duration.

Now, from (5.1), the received training signal at the **BS**, normalized by  $\sqrt{P_{B,\tau}L_{B,\tau}}$  is

$$\bar{Y}_{A,\tau} \triangleq \frac{Y_{A,\tau}}{\sqrt{P_{B,\tau}L_{B,\tau}}} = V\Sigma^H D\sqrt{\phi_c} + \frac{W_{A,\tau}}{\sqrt{P_{B,\tau}L_{B,\tau}}}. \quad (5.4)$$

Denote the  $k^{\text{th}}$  columns of  $\bar{Y}_{A,\tau}$  and  $\frac{W_{A,\tau}}{\sqrt{P_{B,\tau}L_{B,\tau}}}$  by  $\bar{\mathbf{y}}_{k,A,\tau}$  and  $\mathbf{w}_{k,A,\tau}$ , respectively. Note that, in the noiseless case, one can obtain the  $k^{\text{th}}$  column of  $V_m$  by simply normalizing

$\bar{\mathbf{y}}_{k,A,\tau}$ . Motivated by this, even in the presence of noise, using  $\bar{Y}_{A,\tau}$ , the **BS** estimates  $V_m$  as  $\hat{V}_m \triangleq [\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_m]$ , where

$$\hat{\mathbf{v}}_k = \frac{\bar{\mathbf{y}}_{k,A,\tau}}{\|\bar{\mathbf{y}}_{k,A,\tau}\|_2}, \quad 1 \leq k \leq m. \quad (5.5)$$

The choice of the matrix  $D$  and the parameter  $\phi_c$  determine the allocation of the power to different modes and realizations of the channel, which can be used to control the estimation accuracy of  $\hat{V}_m$ . The design of the RCT sequence involves jointly optimizing the matrix  $D$  and  $\phi_c$ . In particular, we wish to solve the following two optimization problems:

- Minimizing the MSE in the estimate of  $V_m$ :

$$\min_{\phi_c > 0, D \in \mathbb{R}^{n_B \times m}} \mathbb{E} \left\| V_m - \hat{V}_m \right\|_F^2 = \min_{\phi_c, D} \mathbb{E} \sum_{k=1}^m \|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2, \quad (5.6)$$

such that  $D$  is diagonal and non-negative,  $\|D\|_F^2 \leq 1$  and  $\mathbb{E}\phi_c \leq 1$ .

- Maximizing the data rate:<sup>3</sup>

$$\max_{L_{B,\tau} > 0, \phi_c > 0, D \in \mathbb{R}^{n_B \times m}} \text{data rate} \quad (5.7)$$

such that  $D$  is diagonal and non-negative,  $\|D\|_F^2 \leq 1$ ,  $\mathbb{E}\phi_c \leq 1$  and  $1 \leq L_{B,\tau} \leq L_c$ .

Note that, in conventional orthogonal RCT, one uses the training sequence

$$X_{B,\tau}^{(\text{conv})} = \sqrt{\frac{P_{B,\tau} L_{B,\tau}}{n_B}} Q, \quad (5.8)$$

where  $Q$  is any  $n_B \times L_{B,\tau}$  matrix with orthonormal rows. Further, the transmitter uses

---

<sup>3</sup>We derive an explicit form for the data rate in Sec. 5.2.3.

the received training signal to estimate the channel matrix  $\hat{H}$  using either least-squares or MMSE estimation,<sup>4</sup> and then employs the  $m$  dominant right singular vectors of the estimated channel as the BF vectors for SM data transmission.

In the following subsection, we present the solution to (5.6). For analytical tractability, we first derive a tight approximate expression for  $\mathbb{E} \left\| V_m - \hat{V}_m \right\|_F^2$ , and then use it to optimize the training sequence.

### 5.2.1 MSE Optimal $D$ and $\phi_c$

First, we state the following theorem, which presents an approximate expression for the MSE,  $\mathbb{E} \left\| V_m - \hat{V}_m \right\|_F^2$ .

**Theorem 9.** *Let the columns of  $\hat{V}_m$  be given by (5.5), and suppose  $\phi_c > 0$ . Then, there exists a  $\hat{V}_{m,approx} \triangleq V_m + E \in \mathbb{C}^{n_A \times m}$  such that*

$$\hat{V}_m = \hat{V}_{m,approx} + \mathcal{O} \left( \frac{1}{P_{B,\tau} L_{B,\tau}} \right), \quad (5.9)$$

with the columns of  $E \triangleq [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m] \in \mathbb{C}^{n_A \times m}$  defined as follows:

$$\mathbf{e}_k \triangleq \frac{-\Re\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\}}{\sigma_k d_k \sqrt{\phi_c}} \mathbf{v}_k + \frac{1}{\sigma_k d_k \sqrt{\phi_c}} \mathbf{w}_{k,A,\tau}. \quad (5.10)$$

Further,

$$\left| \mathbb{E} \left\| V_m - \hat{V}_m \right\|_F^2 - \mathbb{E} \left\| V_m - \hat{V}_{m,approx} \right\|_F^2 \right| = \mathcal{O} \left( \frac{1}{(P_{B,\tau} L_{B,\tau})^2} \right), \quad (5.11)$$

where the expectation in the left hand side is with respect to the channel singular values. Also,

---

<sup>4</sup>Note that the estimate of the BF vectors obtained using the least-squares estimate and the MMSE estimate are the same, as both the estimates are constant multiples of each other.

an approximate expression for the MSE is given by

$$\mathbb{E} \left\| V_m - \hat{V}_{m,\text{approx}} \right\|_F^2 = \left( \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \right) \mathbb{E} \sum_{k=1}^m \frac{1}{\sigma_k^2 d_k^2 \phi_c}. \quad (5.12)$$

*Proof:* See Appendix D.0.14. ■

*Remark 1:* The approximate expression for the MSE in (5.12) differs from  $\mathbb{E} \left\| V_m - \hat{V}_m \right\|_F^2$  only in the second order terms of the training power and duration, due to the  $(P_{B,\tau} L_{B,\tau})^2$  term in the denominator of (5.11). In Sec. 5.4, we illustrate the tightness of the approximate MSE expression at practical training powers.

*Remark 2:* From the above, when  $d_k = 1/\sqrt{m}$  and  $\phi_c = 1$ , i.e., with equal spatio-temporal power allocation, we have

$$\mathbb{E} \left\| V_m - \hat{V}_{m,\text{approx}} \right\|_F^2 = \frac{m(2n_A - 1)}{2P_{B,\tau} L_{B,\tau}} \mathbb{E} \left( \sum_{i=1}^m \sigma_i^{-2} \right). \quad (5.13)$$

Using (5.12), our optimization problem becomes

$$\min_{\phi_c > 0, d_i \geq 0} \left( \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \right) \mathbb{E} \sum_{k=1}^m \frac{1}{\sigma_k^2 d_k^2 \phi_c} \text{ such that } \sum_{i=1}^m d_i^2 \leq 1, \mathbb{E} \phi_c \leq 1. \quad (5.14)$$

The solution is given by the following Lemma.

**Lemma 4.** *The optimal  $D$  and  $\phi_c$  that solve (5.14) are given by*

$$d_k = \sqrt{\frac{\sigma_k^{-1}}{\sum_{i=1}^m \sigma_i^{-1}}} \text{ and } \phi_c = \frac{\sum_{i=1}^m \sigma_i^{-1}}{\mathbb{E} \sum_{i=1}^m \sigma_i^{-1}}. \quad (5.15)$$

*The approximate MSE with  $\phi_c = 1$  and optimal  $D$  is*

$$\mathbb{E} \left\| V_m - \hat{V}_{m,\text{approx}} \right\|_F^2 = \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \mathbb{E} \left( \sum_{i=1}^m \sigma_i^{-1} \right)^2. \quad (5.16)$$

The approximate MSE with the jointly optimal  $D$  and  $\phi_c$  is

$$\mathbb{E} \left\| V_m - \hat{V}_{m,approx} \right\|_F^2 = \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \left( \mathbb{E} \sum_{i=1}^m \sigma_i^{-1} \right)^2. \quad (5.17)$$

*Proof:* See Appendix D.0.15. ■

Comparing the MSE with constant power training ( $\phi_c = 1$ ) in (5.16) and the MSE with power-controlled training in (5.17), it is clear that the power-controlled training outperforms the constant power training, since  $(\mathbb{E} \sum_{i=1}^m \sigma_i^{-1})^2 \leq \mathbb{E} (\sum_{i=1}^m \sigma_i^{-1})^2$  by Jensen's inequality. Thus, spatio-temporal power allocation during training improves the accuracy of the estimate. Note that, the above solution is valid if  $\mathbb{E} \sum_{i=1}^m \sigma_i^{-1} < \infty$ , which is true, for example, when the channel is Rayleigh fading, and  $n_A \neq n_B$ , since [61]

$$\mathbb{E} \sigma_i^{-1} \leq \sqrt{\mathbb{E} \sigma_i^{-2}} < \sqrt{\mathbb{E} \sum_{i=1}^n \sigma_i^{-2}} = \sqrt{\mathbb{E} \text{Trace}\{(H^H H)^{-1}\}} = \sqrt{\frac{n_B}{|n_A - n_B|}} < \infty. \quad (5.18)$$

*Remark 3:* For a given channel instantiation, the MSE in estimating the dominant BF vector is small, compared to the modes with smaller gain. Hence, using the inverse of the channel singular values to fix the power allocation, as given by (5.15), is intuitively satisfying. Similarly, since across time, it is reasonable to allot power proportional to  $\sum_{i=1}^m \sigma_i^{-1}$ , which, roughly speaking, measures the “goodness” of the channel.

In the above, we optimized the power allocated across space and time with the MSE as the performance metric. In the following, we optimize the RCT with the data rate as the performance metric.

## 5.2.2 Data Transmission and Capacity Lower Bound

As mentioned earlier, in Phase III, the *data*  $\mathbf{x}_{A,d} \in \mathbb{C}^{m \times 1}$  is pre-multiplied by  $\hat{V}_m \in \mathbb{C}^{n_A \times m}$  obtained from (5.5), and transmitted to the UT. Since the UT has perfect CSI, it pre-multiplies the received data signal by  $U_m^H$ , where  $U_m \in \mathbb{C}^{n_B \times m}$  is the first  $m$  columns of the matrix  $U$ . From (5.2), the corresponding received data signal is:

$$\mathbf{y}_{B,d} = \sqrt{\frac{P_{A,d}}{m}} \Sigma_m V^H \hat{V}_m \mathbf{x}_{A,d} + U_m^H \mathbf{w}_{B,d}, \quad (5.19)$$

where  $\Sigma_m \in \mathbb{R}^{m \times n_A}$  is the first  $m$  rows of the matrix  $\Sigma \in \mathbb{R}^{n_B \times n_A}$ . Note that the distribution of the entries of  $\tilde{\mathbf{w}}_{B,d} \triangleq U_m^H \mathbf{w}_{B,d}$  is the same as the entries of  $\mathbf{w}_{B,d}$ , since  $U_m^H$  has orthonormal columns. We rewrite (5.19) as

$$\mathbf{y}_{B,d} = \sqrt{\frac{P_{A,d}}{m}} G \mathbf{x}_{A,d} + \tilde{\mathbf{w}}_{\text{eff}}, \quad (5.20)$$

where  $G \triangleq \Sigma_{m,m} - \Sigma_m V^H \mathbb{E}_{|H} \{V_e\}$ ,  $V_e \triangleq V_m - \hat{V}_m$ , and

$$\tilde{\mathbf{w}}_{\text{eff}} \triangleq \sqrt{\frac{P_{A,d}}{m}} \Sigma_m V^H \mathbb{E}_{|H} \{V_e\} \mathbf{x}_{A,d} - \sqrt{\frac{P_{A,d}}{m}} \Sigma_m V^H V_e \mathbf{x}_{A,d} + \tilde{\mathbf{w}}_{B,d}. \quad (5.21)$$

In the above, we have used  $\mathbb{E}_{|H} \{X\}$  to mean the expected value of  $X$  conditioned on  $H$ . Here,  $\Sigma_{m,m} \in \mathbb{R}^{m \times m}$  is the  $m \times m$  principal submatrix of  $\Sigma$ , and (5.20) is obtained by adding and subtracting the first term in (5.21). Note that  $G$  in the first term in (5.20) is a deterministic function of  $H$ , and it is easy to see that the effective noise term  $\tilde{\mathbf{w}}_{\text{eff}}$  is uncorrelated with the data given the channel, i.e.,  $\mathbb{E}_{|H} \{\tilde{\mathbf{w}}_{\text{eff}} \mathbf{x}_{A,d}^H\} = 0$ . Hence, for the system in (5.20), the worst case noise theorem [8] is applicable, which results in the

following lower bound  $C_{\text{LB}}$  on the channel capacity:

$$C_{\text{LB}} \triangleq \alpha \mathbb{E}_H \log_2 \left| I_m + P_{A,d} \frac{GG^H}{\mathbb{E}_{|H}\{\|\tilde{\mathbf{w}}_{\text{eff}}\|_F^2\}} \right|, \quad (5.22)$$

where  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$ . Here,  $\mathbb{E}_H\{\cdot\}$  outside the log function denotes the expectation with respect to  $H$ , and  $\mathbb{E}_{|H}\{\cdot\}$  in the denominator term of (5.22) and in the definition of  $G$  denotes the expectation with respect to the distribution of the additive noise in the training and data phases given  $H$ .

Note that the derivation of the capacity lower bound in (5.22) is independent of the specific training scheme used to estimate the dominant BF vectors. Thus, it is valid for the orthogonal RCT scheme also.

Now, it turns out that directly optimizing the training sequence to maximize the above CLB is analytically intractable due to  $\mathbb{E}_{|H}\{V_e\}$  term in (5.22), which is hard to analyze. Hence, we derive an approximate expression for the CLB in the following theorem, and use it to optimize the training sequence. The solution obtained by optimizing the approximate CLB becomes accurate asymptotically as the data and training powers become large.

**Theorem 10.** *Let*

$$C_{\text{LB},a} \triangleq \alpha \mathbb{E} \log_2 \left| I_m + \frac{P_{A,d}}{m} \frac{\Sigma_{m,m} \Sigma_{m,m}^H}{1 + \sigma_{\text{eff}}^2} \right| \quad (5.23)$$

with

$$\sigma_{\text{eff}}^2 \triangleq \frac{P_{A,d}}{P_{B,\tau} L_{B,\tau} m^2} \sum_{i=1}^m \frac{\beta_i}{d_i^2 \phi_c}, \quad \text{and } \beta_i \triangleq \frac{1}{2} + \frac{\sum_{j=1, j \neq i}^m \sigma_j^2}{\sigma_i^2}. \quad (5.24)$$

Then,  $C_{LB}$  is an approximation to the capacity lower bound given by (5.22), in the sense

$$|C_{LB} - C_{LB,a}| \rightarrow 0 \text{ as } P_{A,d}, P_{B,\tau} \rightarrow \infty \text{ such that } \frac{P_{A,d}}{P_{B,\tau}} \leq \mu, \mu > 0.$$

*Proof:* See Appendix D.0.16.

Note that the effect of channel estimation errors on the data rate is captured via  $\sigma_{\text{eff}}^2$  in the denominator of the approximate expression in (5.23). When the training power is large, the loss in the data rate is small, as expected. The effect of the spatio-temporal power allocation parameters of the training sequence on the data rate is also captured through the  $\sigma_{\text{eff}}^2$  term.

In the next subsection, we optimize the training sequence to maximize (5.23).

### 5.2.3 Capacity Lower Bound Optimal $D$ , $\phi_c$ and $L_{B,\tau}$

We want to find the maximum of  $C_{LB,a}$  given by (5.23) subject to  $\|D\|_F^2 \leq 1$  and  $\mathbb{E}\phi_c \leq 1$ , which is equivalent to first optimizing over  $D$  for a given  $\phi_c$ , and then finding the optimal functional  $\phi_c$ , as follows:

$$\begin{aligned} \mathbf{P1} : \quad & \max_{\phi_c} \max_D C_{LB,a} & (5.25) \\ & \text{such that } \|D\|_F^2 \leq 1, \text{ and } \mathbb{E}\phi_c \leq 1. \end{aligned}$$

The solution to the above problem is given in the following theorem.

**Theorem 11.** For a given  $L_{B,\tau}$ , the optimal  $(d_1, d_2, \dots, d_m)$  that solves **P1** in (5.25) is given by

$$d_i = \sqrt{\frac{\sqrt{\beta_i}}{\sum_{j=1}^m \sqrt{\beta_j}}}, \quad 1 \leq i \leq m, \quad (5.26)$$

where  $\beta_i$  is as defined in (5.24). Also, the optimal temporal power allocation  $\phi_c^*$  satisfies

$$\lambda = \mathcal{H}(\phi_c^*) \triangleq \left( \frac{1}{\tau + \phi_c^*} \right) \sum_{k=1}^m \frac{P_{A,d} \sigma_k^2 \tau}{(P_{A,d} \sigma_k^2 + m) \phi_c^* + m \tau}, \quad (5.27)$$

where  $\tau \triangleq \frac{P_{A,d}}{P_{B,\tau} L_{B,\tau} m^2} \left( \sum_{k=1}^m \sqrt{\beta_k} \right)^2$ , and  $\lambda$  is a Lagrange multiplier, chosen such that  $\mathbb{E} \phi_c^* = 1$ , where the expectation is over the distribution of  $(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$ .

*Proof:* See Appendix D.0.17. ■

Note that, the optimal power control is available only implicitly, as given by (5.27). The value of  $\lambda$  has to be numerically computed. A procedure for finding  $\lambda$  is described in Sec. 5.4.

In the special case of pure BF, i.e., when  $m = 1$ , we have the following simple closed-form solution for the power control policy, and the resulting capacity lower bound.

**Corollary 1.** For  $m = 1$ , the capacity lower bound with optimal power control policy is given by

$$C_{LB,BF} = \alpha \mathbb{E} \log_2 \left( 1 + P_{A,d} \frac{\sigma_1^2}{1 + \sigma_{eff,BF}^2} \right), \quad (5.28)$$

where  $\sigma_{eff,BF}^2 \triangleq \frac{P_{A,d}}{2P_{B,\tau} L_{B,\tau} \phi_c^*}$  with

$$\phi_c^* = \left( \frac{-\tau_2 + \sqrt{\tau_2^2 - 4\tau_1\tau_3}}{2\tau_1} \right)^+.$$

In the above,  $\tau_1 \triangleq 1 + P_{A,d} \sigma_1^2$ ,  $\tau_2 \triangleq 2\tau + \tau P_{A,d} \sigma_1^2$  and  $\tau_3 \triangleq \tau^2 - \frac{P_{A,d} \sigma_1^2 \tau}{\lambda_{bf}}$ . Here,  $\lambda_{bf}$  is chosen to satisfy the average power constraint, and  $\tau$  is as defined earlier, with  $m = 1$ .

*Remark 4:* The only parameter of the training sequence that remains to be optimized is the training duration  $L_{B,\tau}$ . The optimal  $L_{B,\tau}$  can be obtained using a simple off-line search over  $\{1, 2, \dots, L_c\}$ . We relegate the details to Sec. 5.4.

In the above, we found the optimal spatial and temporal power allocation of RCT sequence that maximizes an approximate CLB. In the following section, we consider the design of the RCT in a multiuser setting and its effect on the data rate.

### 5.3 Multiuser Scenario

In this section, we use the previously derived approximate expression for the CLB as a metric to schedule a single user for data transmission, and analyze the data rate of the system. If the  $k^{\text{th}}$  user is scheduled for data transmission, from Theorem 10, with  $\phi_c = 1$ , the approximate data rate achieved by it is given by

$$R_k \triangleq \alpha \log_2 \left| I_m + \frac{P_{A,d}}{m} \frac{\Sigma_{k,m,m} \Sigma_{k,m,m}^H}{1 + \sigma_{k,\text{eff}}^2} \right|, \quad (5.29)$$

where

$$\sigma_{k,\text{eff}}^2 \triangleq \frac{P_{A,d}}{P_{B,\tau} L_{B,\tau} m^2} \left( \sum_{i=1}^m \sqrt{\beta_{i,k}} \right)^2, \quad \beta_{i,k} \triangleq \frac{1}{2} + \frac{\sum_{l=1, l \neq i}^m (\sigma_{l,k})^2}{(\sigma_{i,k})^2}, \quad (5.30)$$

and  $\Sigma_{k,m,m}$  is the  $m \times m$  principal submatrix of the singular value matrix of  $H_k$ . For simplicity, we assume that the data power, RCT power and RCT duration are the same for all users. Further, we assume that the channels are i.i.d. across users. The user is selected for data transmission using max-rate scheduling based on  $R_k$ . Since  $R_k$  can be computed locally at each receiver (user), the user with the highest metric can be efficiently selected using splitting or a timer based scheme, as explained in Sec. 5.1.<sup>5</sup>

---

<sup>5</sup>Note that using the minimum MSE in the estimate of the BF vectors as the criterion for user selection coincides with max-rate scheduling, when  $m = 1$ . Both schemes select the user whose channel matrix has the largest dominant singular value.

With max-rate scheduling, the average sum data rate is given by

$$R_{\text{avg}} \triangleq \mathbb{E}_{H_1, \dots, H_M} [\max\{R_1, R_2, \dots, R_M\}], \quad (5.31)$$

Unfortunately, a closed-form expression for (5.31) is hard to find when  $m > 1$ , as it involves finding the distribution of complicated terms involving  $\beta_{i,k}$ . However, in the case of pure BF, i.e., when  $m = 1$ , a simplified expression involving a single integral can be found, as shown in the following theorem.

**Theorem 12.** *When  $m = 1$ ,  $R_{\text{avg}}$  is given by*

$$R_{\text{avg}} = \int_0^\infty (1 - (\Pr\{\sigma_{1,1}^2 \leq \omega\})^M) d\omega, \quad (5.32)$$

where

$$\omega \triangleq \frac{1}{P_{A,d}} \left(2^{\frac{x}{\alpha}} - 1\right) \left(1 + \frac{P_{A,d}}{2P_{B,\tau}L_{B,\tau}}\right),$$

and

$$\Pr\{\sigma_{1,1}^2 \leq \omega\} \triangleq \frac{e^{-\omega}}{\prod_{k=1}^{n_B} (n_B - k)!(n_A - k)!} \sum_{j=1}^{n_B} \sum_{p=n_A-n_B}^{(n_A+n_B)j-2j^2} c_{j,p} p! \left[ \sum_{s=0}^p \frac{(-1)^j \omega^{p-s}}{(p-s)!} \right]. \quad (5.33)$$

Here, the constants  $c_{j,p}$  are the coefficients of  $e^{-jy}y^p$  term in the pdf of  $\sigma_{1,1}^2$ , the largest eigenvalue of  $H_1^H H_1$ , which can be found using Table I in [62].

*Proof:* See Appendix D.0.18. ■

Now, although the above result provides a simple, easy-to-compute integral using which one can analyze the performance of the proposed RCT scheme, it is hard to obtain insight into the system behavior as the number of users, RCT power, data power, etc. are varied. In the following theorem, we derive an upper bound on  $R_{\text{avg}}$ , and show

that it scales with  $M$  as  $\alpha \log_2 \log M$  with the number of users  $M$ . We illustrate the tightness of the bound in Sec. 5.4.

**Theorem 13.** *When  $m = 1$ , an upper bound on  $R_{avg}$  is given by*

$$R_{avg} \leq R_{avg}^u \triangleq \alpha \log_2 \left( 1 + \frac{P_{A,d}}{1 + \sigma_{BF,eff}^2} \inf_{s \in (0,1)} \left[ \frac{\log M}{s} - \frac{n_A n_B \log(1-s)}{s} \right] \right), \quad (5.34)$$

where  $\sigma_{BF,eff} \triangleq \frac{P_{A,d}}{2P_{B,\tau}L_{B,\tau}}$ . Further,

$$\lim_{M \rightarrow \infty} \frac{R_{avg}^u}{\alpha \log_2 \log M} = 1.$$

*Proof:* See Appendix D.0.19.

It is interesting to note that a similar data rate scaling of  $\log_2 \log M$  was also observed in the case of a multiuser MISO down link channel with perfect CSIR and perfect CSIT [63, 64].

To summarize, we used the previously derived approximate expression for the CLB as a metric to schedule a single user for data transmission, and showed that the upper bound on the data rate of the system scales with  $M$  as  $\alpha \log_2 \log M$ . The proposed RCT scheme directly leads to an improvement in performance compared to orthogonal RCT, due to the dependence of the data rate on the factor  $\alpha$ , which captures the training duration overhead.

## 5.4 Simulation Results

In this section, we illustrate the performance improvement offered by the proposed RCT scheme and validate our analytical development using Monte Carlo simulations.

The simulation set up consists of an  $n_B \times n_A = 3 \times 4$  MIMO SM system with  $P_{A,d} = P_{B,\tau}$ ,  $L_{B,\tau} = 3$  symbols, and  $L_c = 100$  symbols. The MIMO channel is assumed to be Rayleigh flat fading channel with i.i.d. coefficients drawn from  $\mathcal{CN}(0, 1)$ . The AWGN is also modeled as having i.i.d.  $\mathcal{CN}(0, 1)$  components. With conventional orthogonal RCT, we employ the training sequence in (5.8) as in [5] to obtain an estimate of the channel, from which dominant BF vectors are computed using the SVD. The resulting error in the estimated BF vectors is used to compute the MSE, and the CLB is computed using the expression in (5.22). The computed CLB is used to schedule the user for RCT and data transmission, and hence, with conventional estimation also, only the selected user sends the RCT signal. We compare the performance of the proposed RCT scheme with orthogonal RCT both in terms of the MSE and the CLB, in the following subsections.

### 5.4.1 Mean Square Error

Figure 5.2 shows the performance of the training scheme proposed in Sec. 5.2.1 in terms of the MSE in the estimate of  $V_m$  versus  $P_{B,\tau}$ , with  $m = 3$  modes. Since the proposed RCT scheme has multiple parameters, it is of interest to see the gain offered by optimizing each of the components. Towards this, we plot the MSE with the following settings: (a)  $X_{B,\tau}$  in (5.3) with  $D = I_{n_B \times m}$  and  $\phi_c = 1$  (fixed-power RCT), (b)  $X_{B,\tau}$  in (5.3) with optimal  $D$  and  $\phi_c = 1$  (RCT with optimal spatial power control), (c)  $X_{B,\tau}$  in (5.3) with jointly optimal  $D$  and  $\phi_c$  (RCT with the optimal spatio-temporal power control), and (d) Conventional orthogonal training ( $X_{B,\tau}^{(\text{conv})}$  in (5.8)) [5]. From the plot, we observe that the approximate theoretical expression in (5.17) is tight, and that the

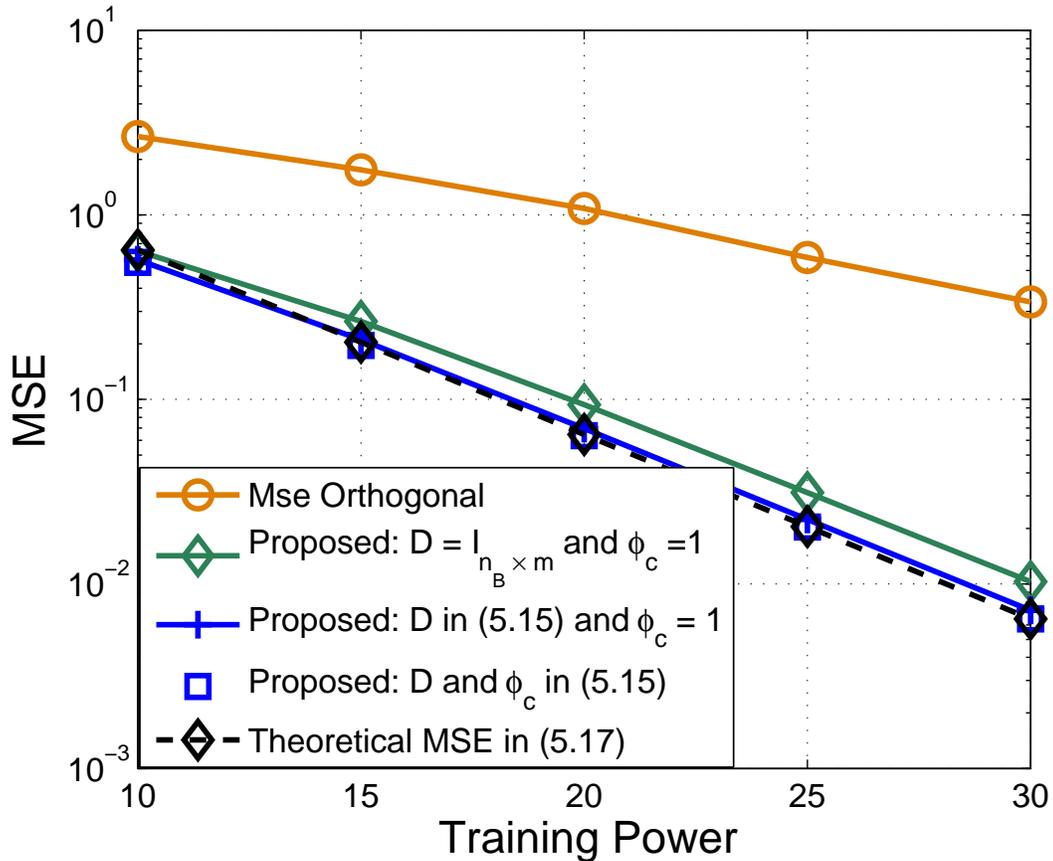


Figure 5.2: MSE versus training power for a  $3 \times 4$  MIMO system with  $m = 3$ .

proposed scheme significantly outperforms orthogonal RCT at all training powers. Using the optimal spatial power allocation during training offers a gain of approximately 1dB at  $P_{B,\tau} = 15$ dB compared to using  $D = I_{n_B \times m}$ . However, temporal power allocation does not further significantly improve the performance compared to pure spatial power allocation.

#### 5.4.2 Data Rate with a Single User

Figure 5.3 shows a plot of the capacity lower bound in (5.22) and its approximate expression in (5.23) versus the training power. It is clear that the approximate expression

for the capacity lower bound is tight at all training powers. Further, the data rate corresponding to  $m = 2$  outperforms  $m = 3$ , i.e., using higher number of modes is not always optimal when the training overhead for acquiring CSIT is taken into account.

For the optimal spatio-temporal power allocation, the value of the Lagrange multiplier  $\lambda > 0$  in (5.27) is required to plot the CLB. To compute this, we start with some  $\lambda > 0$ . We generate a large number of channel instantiations, and compute the power control function  $\phi_c^*$  for each instantiation by inverting (5.27), as it is a monotone function of  $\phi_c$ . We then compute the average of the values of the  $\phi_c$ 's so obtained. Due to the monotonicity of the function  $\mathcal{H}(\phi_c^*)$ , if the average exceeds unity, we increment  $\lambda$ , otherwise we decrement  $\lambda$  by a small step. Repeating this procedure until the average value is sufficiently close to one yields the desired  $\lambda$ , and consequently, the optimal power control function.

Figure 5.4 shows a plot of the exact CLB in (5.22) versus the training power. The proposed training offers an improvement of about 2 bits/channel use over orthogonal training. Also, optimal spatial power allocation during training outperforms the proposed scheme with  $D = I_{n_B \times m}$  by approximately 1 bit/channel use at  $P_{B,\tau} = 20\text{dB}$ . On the other hand, temporal power allocation during training only offers a marginal data rate improvement.

Figure 5.5 shows the capacity lower bound in (5.22) versus training duration with  $P_{A,d} = P_{B,\tau} = 6\text{dB}$  for the proposed and orthogonal training schemes. The figure shows that training for the minimum duration of one symbol is not always optimal. For the proposed training scheme, the optimal training duration is 8 symbols, while for the orthogonal training scheme, it is 12 symbols. Thus, the analysis presented in this

chapter can be used to determine the training duration that optimally trades-off the estimation accuracy with the time overhead due to training.

### 5.4.3 Data Rate with Multiple Users

Figure 5.6 shows a comparison of the proposed training scheme with the orthogonal training scheme in terms of the average data rate versus  $P_{B,\tau}$  for a multi-user system with max-rate based user scheduling,  $m = 3$  modes, and  $M = 2$  and 6 users. We use the capacity lower bound expression in (5.22) while evaluating the average data rate in (5.31). The proposed RCT scheme can lead to a reduction of 2 to 3dB of training power compared to orthogonal RCT for achieving the same data rate.

Finally, in Fig. 5.7, we study the behavior of the average data rate as a function of the number of users. We consider a  $2 \times 2$  multiuser BF system with  $L_{B,\tau} = 1$ , and plot the normalized data rate, defined as  $\frac{R_{\text{avg}}}{\log_2 \log M}$ , where  $R_{\text{avg}}$  is as in (5.31), versus  $M$ . We see that the approximate expression in (5.32) matches well with the exact expression in (5.31). Further, the upper bound in (5.34) also captures the  $\log_2 \log M$  scaling of the average data rate very well.

## 5.5 Conclusions

This chapter considered a multiuser SM based TDD-MIMO system with perfect CSIR. First, for a single user system, a novel *power-controlled* Reverse Channel Training (RCT) scheme that adapts to the time-varying channel was proposed. This was used by the **BS** to estimate the dominant beamforming vectors of the channel. The spatial and temporal allocation of the training matrix were optimized using the following two metrics:

(i) a capacity lower bound, and (ii) Mean Square Error (MSE), subject to an average power constraint. We then extended the training scheme and the data rate analysis to a multiuser case. Further, for a BF system, we derived a closed-form expression for the average sum data rate and its upper bound. We showed that the upper bound scales as  $\frac{L_c - L_{B,\tau}}{L_c} \log_2 \log M$  with the number of users  $M$ , where  $L_c$  and  $L_{B,\tau}$  are the channel coherence time and the training duration, respectively. Using simulation results, we demonstrated the significant performance gain offered by the proposed training sequence over conventional orthogonal RCT. We also illustrated that the spatial power allocation during training outperforms its equal power allocation counterpart, while temporal power allocation only offers a marginal improvement in performance.

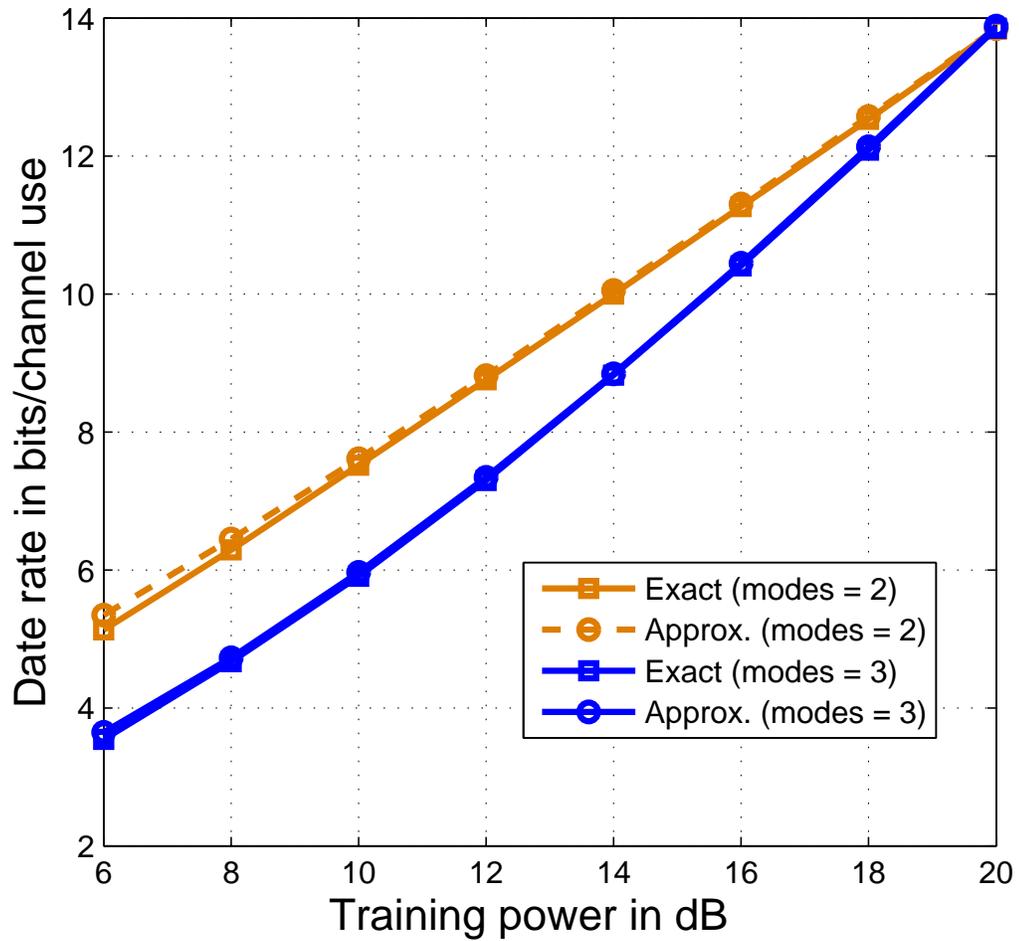


Figure 5.3: The figure demonstrates the tightness of the proposed approximation in (5.23) for the capacity lower bound in (5.22) for a  $3 \times 4$  MIMO system with the data power of  $P_{A,d} = P_{B,\tau}$ ,  $L_{B,\tau} = 3$  symbols.

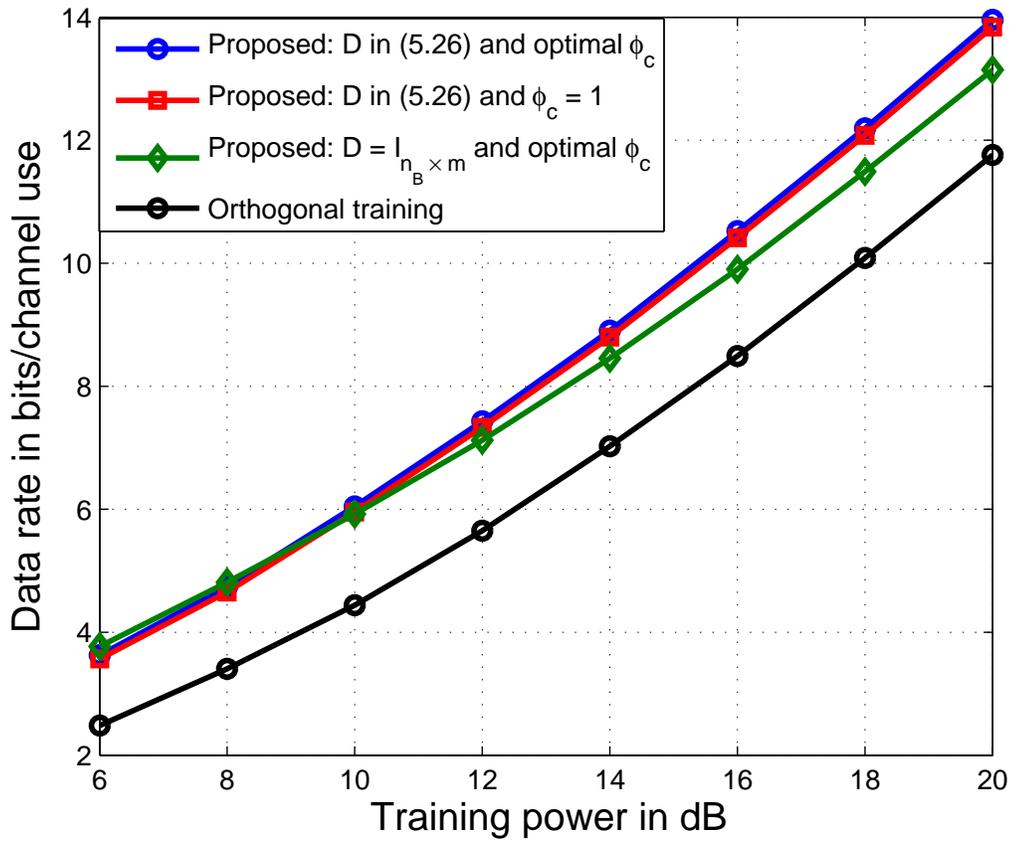


Figure 5.4: Capacity lower bound in (5.22) for a  $3 \times 4$  MIMO system versus RCT power  $P_{B,\tau}$ , with  $P_{A,d} = P_{B,\tau}$ ,  $L_{B,\tau} = 3$  symbols and  $m = 3$  modes.

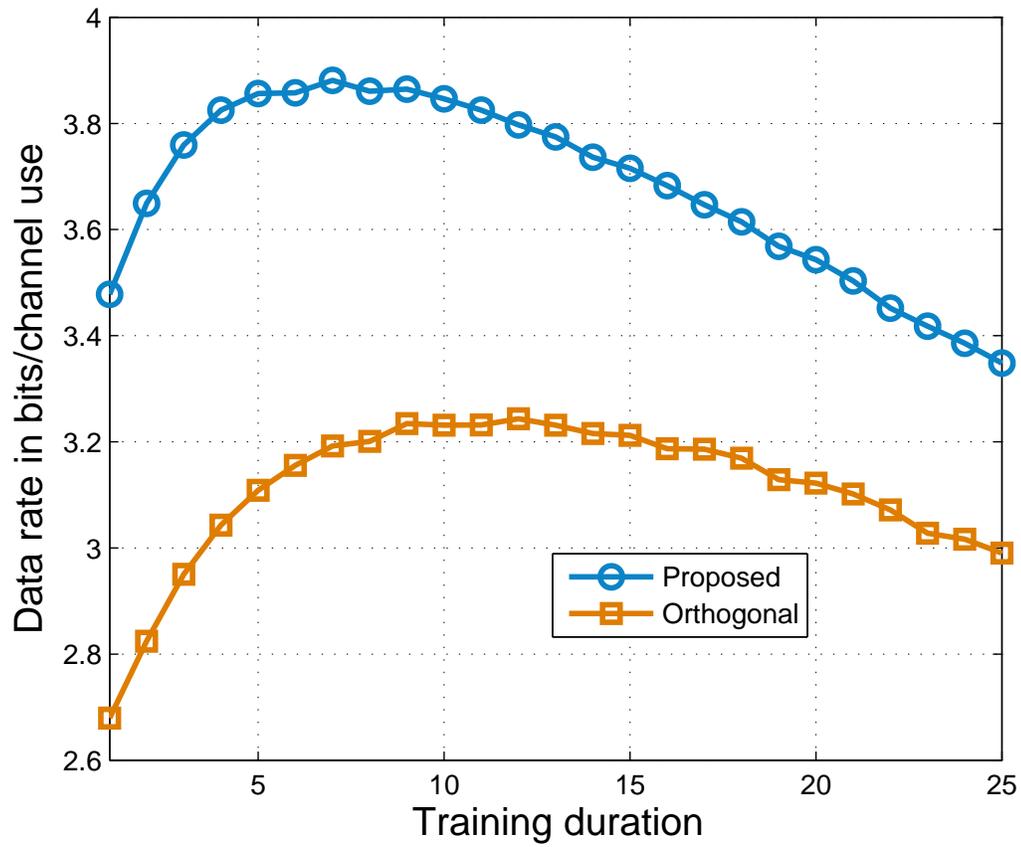


Figure 5.5: Capacity lower bound in (5.22) for a  $3 \times 4$  MIMO system versus training duration with  $P_{A,d} = P_{B,\tau} = 6\text{dB}$ .

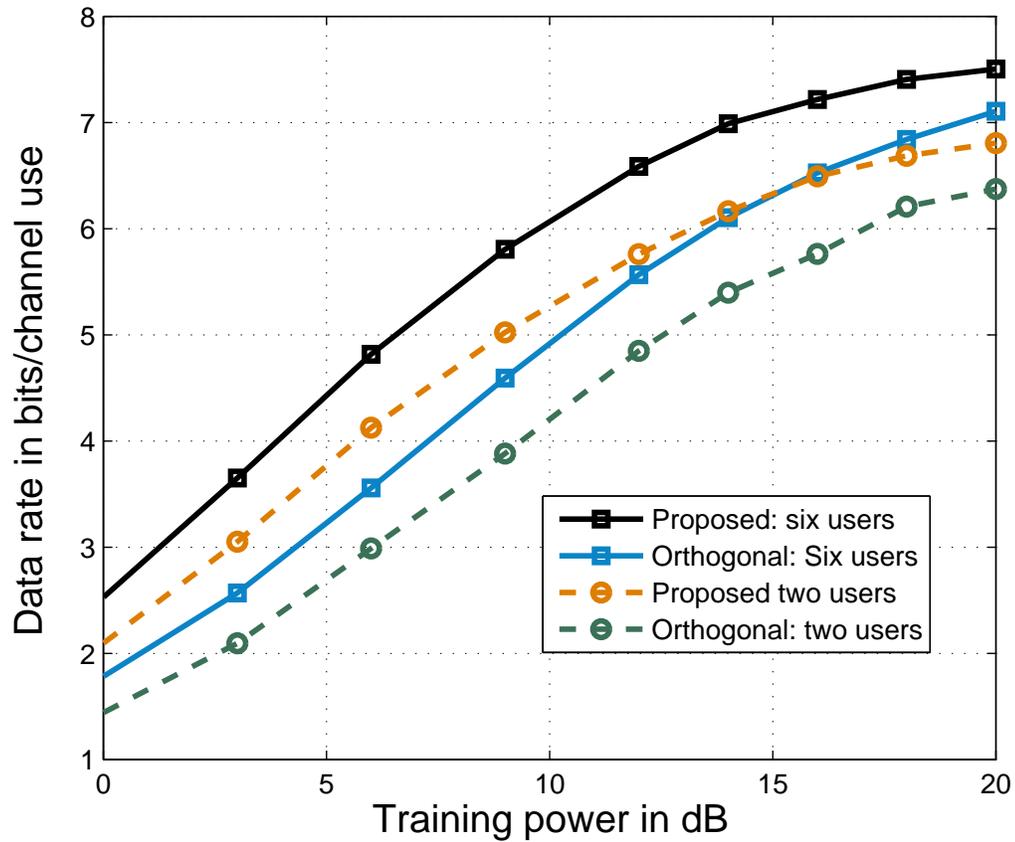


Figure 5.6: Capacity lower bound in (5.22) for a  $3 \times 4$  multiuser MIMO system, versus training power  $P_{B,\tau}$ , with the data power  $P_{A,d} = P_{B,\tau}$ , and for the scheduling scheme described in Sec. 5.3.

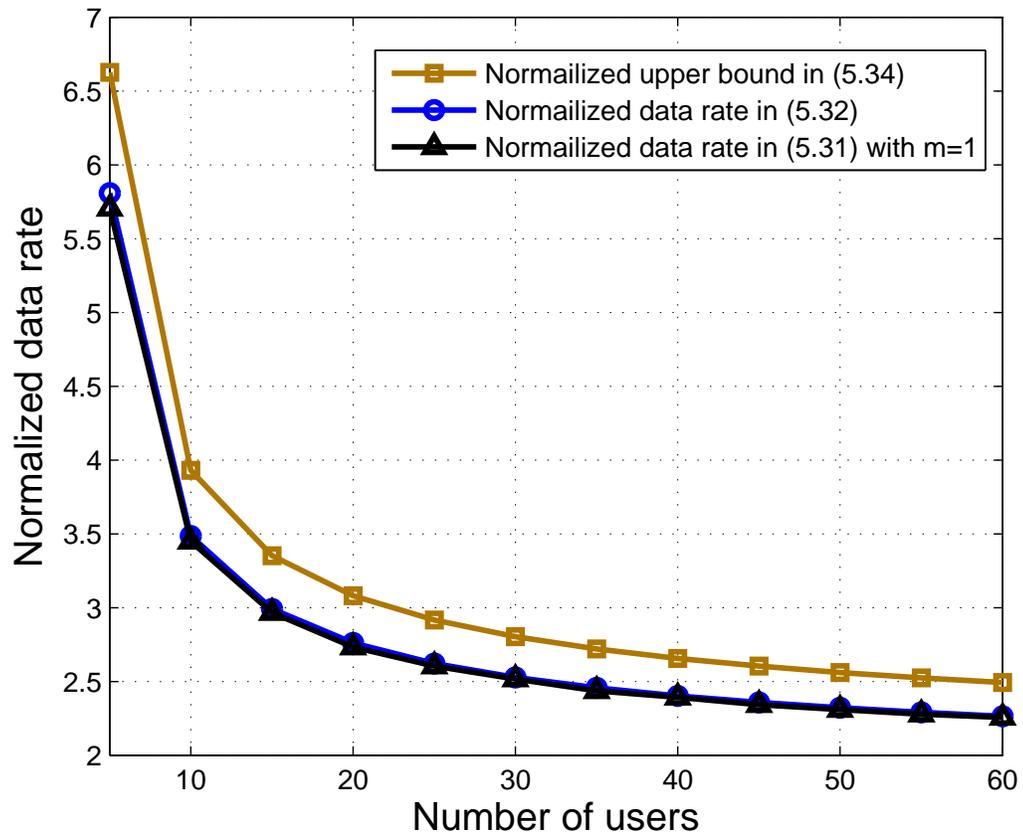


Figure 5.7: Normalized data rate versus number of users  $M$  for a  $2 \times 2$  multiuser MIMO BF system with  $P_{A,d} = P_{B,\tau} = 6\text{dB}$ . Here, the data rate is normalized by  $\log_2 \log M$ .

# Chapter 6

## Conclusions and Future Work

In this thesis, we investigated the problem of designing RCT sequences in reciprocal MIMO systems that exploit channel knowledge at the receiver to enable fast and efficient channel estimation at the transmitter. The main contributions of this thesis are summarized below.

### 6.1 Conclusions

Chapter 2 proposed reverse training and data power control schemes for a TDD-SIMO system with perfect/imperfect CSIR and investigated its DMT performance. It was shown that a diversity order of  $d(g_m) = r \left( s + 1 - \frac{g_m}{\alpha} \right)$  is achievable for  $l \geq s + 1$ ,  $1 \leq s < r$  and  $0 \leq g_m < \alpha$ , where  $\alpha$  represents the fractional data transmit duration. The diversity order thus increases monotonically with  $r$  at nonzero multiplexing gain. This is a significant improvement over channel-agnostic orthogonal training schemes, where the diversity order saturates with the number of receive antennas. The DMT analysis was extended to a more practical situation where the training is done in both directions. In this case also, it was shown that the DMT performance can improve quadratically

with the number of receive antennas, and nearly the same DMT can be achieved as that with perfect CSIR and a genie-aided receiver. For fast varying channels, the proposed RCT and data transmission scheme can lead to a significant advantage in DMT performance, which, at finite SNR, can translate to a large improvement in outage probability performance compared to orthogonal training schemes.

In chapter 3, we proposed a channel-dependent PCRCT and a data power control scheme in a TDD-SIMO system with perfect CSIB, and analyzed its DMT performance. The key ingredient of the RCT sequence presented in chapter 3 that was not present in the RCT sequence of chapter 2 is that the power in the training sequence was varied depending on the current CSI such that an average power constraint is satisfied. We showed that the proposed scheme achieves an infinite diversity order for  $0 \leq g_m < \alpha \min(\gamma, 1)$ . Also, at high SNR, the derived upper bound on the outage probability goes to zero approximately as  $\exp(-\bar{P}^{\gamma - \frac{g_m}{\alpha}})$ , where  $\gamma > 0$  is the exponent of the RCT power. We also showed that there exists an uncoded data transmission scheme for which the probability of error exhibits an infinite diversity order for  $0 \leq g_m < \alpha \min(\gamma, 1)$ . The proposed scheme can thus achieve a significantly better DMT performance compared to fixed-power, channel-agnostic orthogonal RCT schemes as well as the RCT sequence of chapter 2.

In chapter 3, the training scheme proposed in chapter 2 was extended to a more general MIMO channel. In particular, we considered the design and analysis of a channel-dependent training signal for a TDD-MIMO system with two nodes, *node A* and *node B*, and with data transmission over the dominant mode of the channel from *node A* to *node B*. We assumed that one of the nodes has perfect channel knowledge, and proposed a

scheme for estimating the dominant singular vector at the other node. We derived a lower bound on the forward-link capacity and used it to optimize the training power and duration. We showed that when *node B* has perfect CSI, with independent power constraints at *node A* and *node B*, it is optimal to spend the minimum possible time for training (i.e.,  $L_{B,\tau} = 1$ ), provided the training power is at least of the order of one quarter of the total data transmit power. Otherwise, the optimal  $L_{B,\tau} > 1$ . From an energy efficiency perspective, regardless of the available energy,  $L_{B,\tau} = 1$  is optimal, and the fractional power spent on data transmission is  $\frac{\sqrt{2(L_c-1)}}{\sqrt{2(L_c-1)+1}}$ , where  $L_c$  is the channel coherence time. Also, when *node A* has perfect CSI, at low SNR,  $L_{A,\tau} = 1$  is optimal. We showed that if  $L_c \geq 2n_A - n_B + 1$ , which is typically the case in practical systems, initiating training at *node A* is better than initiating training at *node B* in terms of an approximate capacity lower bound. Monte Carlo simulations validated the theoretical expressions and illustrated the significant performance benefits offered by the proposed channel-dependent training scheme compared to the conventional channel-agnostic orthogonal training scheme.

Chapter 5 extended the training scheme proposed in chapter 4 to a general multi-user Spatial Multiplexing (SM) based TDD-MIMO system with perfect CSIR. First, for a single user system, a novel *power controlled* RCT scheme that adapts to the time-varying channel was proposed. This was used by the **BS** to estimate the dominant beamforming vectors of the channel. The singular values of the training matrix were optimized using the following two metrics: (i) a capacity lower bound, and (ii) Mean Square Error (MSE). We showed that it is optimal to allocate lesser power to the dominant modes of the channel. We then extended the training scheme and the data rate analysis to

a multiuser case. We showed that, in the multiuser case, the proposed scheme offers a significant performance improvement over the conventional orthogonal training scheme. Further, for a BF system, we derived a closed form expression for the average sum data rate and its upper bound. Using the upper bound, we showed that the data rate scales as  $\frac{L_c - L_{B,\tau}}{L_c} \log \log M$  with the number of users  $M$ , where  $L_c$  and  $L_{B,\tau}$  are the channel coherence time and the training duration, respectively. Using simulation results, we demonstrated the significant performance gain offered by the proposed training sequence over the conventional orthogonal RCT sequence. We also illustrated that the spatial power allocation during training outperforms its equal power allocation counterpart while temporal power allocation only offers a marginal improvement in performance.

## 6.2 Future Work

Future work in the design of RCT sequences could include the following issues:

- RCT sequence design proposed in this thesis, for the most part, assumed perfect CSI at *node B*. Although we considered the impact of imperfect CSI at *node B* on the DMT performance in chapter 2, it would be interesting to conduct a more detailed study of the impact of imperfect CSIB on the design of RCT sequences.
- In this thesis, the channel was assumed to be perfectly reciprocal. Extending the RCT design to the case where the forward channel is correlated with the reverse channel is an interesting challenge.

- Throughout this thesis, the channel was assumed to be quasi-static, i.e., it is assumed to be evolving in an i.i.d. fashion. It would be useful to analyze the performance of channel-dependent RCT when the channel is correlated across time. Some initial work in this direction can be found in [51].

# Appendix A

## Appendix for Chapter 2

### A.0.1 Useful Lemmas

**Lemma 5.** *If the random variable  $\sigma^2$  is a chi-square distributed with  $2r$  degrees of freedom, then*

$$\Pr\{\sigma^2 < z\} \leq \frac{z^r}{r!}, z \geq 0.$$

*Proof:* The result follows from

$$\Pr\{\sigma^2 < z\} = \frac{1}{(r-1)!} \int_0^z e^{-x} x^{r-1} dx \quad (\text{A.1})$$

$$\leq \frac{1}{(r-1)!} \int_0^z x^{r-1} dx \quad (\text{A.2})$$

$$= \frac{z^r}{r!}. \blacksquare \quad (\text{A.3})$$

**Lemma 6.** *For the system in (2.3),  $|\hat{\sigma}| \leq \hat{\sigma}_U$ , where  $\hat{\sigma}_U^2 \triangleq (\sigma + |\bar{w}_{A,\tau}|)^2$ , with  $\bar{w}_{A,\tau} \triangleq \frac{\Re\{w_{A,\tau}\}}{\sqrt{PL_{B,\tau}}}$ .*

*Proof:* We upper bound the absolute value of (2.3) as follows:

$$|\hat{\sigma}| \stackrel{(a)}{\leq} \sigma |\Re\{\mathbf{v}^H \hat{\mathbf{v}}\}| + \left| \frac{\Re\{w_{A,\tau}\}}{\sqrt{PL_{B,\tau}}} \right| \quad (\text{A.4})$$

$$\stackrel{(b)}{\leq} \sigma + |\bar{w}_{A,\tau}|, \quad (\text{A.5})$$

where (a) follows from the triangle inequality and (b) follows since  $|\Re\{\mathbf{v}^H \hat{\mathbf{v}}\}| \leq 1$ .  $\blacksquare$

## A.0.2 Proof of Lemma 1

Consider the following constraint on the data power

$$\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \int_{-\infty}^{\infty} \mathcal{P}(\hat{\sigma}) f_{\hat{\sigma}}(\hat{\sigma}; \bar{P}) d\hat{\sigma} = \bar{P}, \quad (\text{A.6})$$

where  $f_{\hat{\sigma}}(\hat{\sigma}; \bar{P})$  is the pdf of  $\hat{\sigma}$ . Substituting (2.10) in (A.6), we get

$$\mathbb{E}[\mathcal{P}(\hat{\sigma})] = \kappa_{\bar{P}} \left[ \exp\left(\frac{L_c R_{\bar{P}}}{L_c - L_{B,\tau}}\right) - 1 \right] F(\bar{P}) + I_{\bar{P}}, \quad (\text{A.7})$$

where  $R_{\bar{P}}$  is the target data rate and the data transmit power is  $\bar{P}$ ,

$$F(\bar{P}) \triangleq \int_{\theta_{\bar{P}}}^{\infty} \frac{1}{x^{2s}} f_{\hat{\sigma}}(x; \bar{P}) dx \quad \text{and} \quad I_{\bar{P}} \triangleq \bar{P}^l \int_{-\infty}^{\theta_{\bar{P}}} f_{\hat{\sigma}}(x; \bar{P}) dx. \quad (\text{A.8})$$

The proof is complete by choosing

$$\kappa_{\bar{P}} = \frac{(\bar{P} - I_{\bar{P}})}{\left(\exp\left(\frac{L_c R_{\bar{P}}}{L_c - L_{B,\tau}}\right) - 1\right) F(\bar{P})}, \quad (\text{A.9})$$

and showing that  $I_{\bar{P}} < \bar{P}$  and that  $F(\bar{P})$  is bounded for large  $\bar{P}$  when  $0 \leq l \leq r + 1$  and  $n = 1/2$ . From (A.8),  $I_{\bar{P}} = \bar{P}^l \Pr\{\sigma + \bar{w}_{A,\tau} < \theta_{\bar{P}}\}$  can be bounded as,

$$I_{\bar{P}} \stackrel{(a)}{\leq} \frac{\bar{P}^l}{r!} \mathbb{E}(\theta_{\bar{P}} - \bar{w}_{A,\tau})^{2r} \quad (\text{A.10})$$

$$\stackrel{(b)}{=} \frac{\bar{P}^l}{r!} \mathbb{E} \sum_{j=0}^r \theta_{\bar{P}}^{2(r-j)} \binom{2r}{2j} \bar{w}_{A,\tau}^{2j}$$

$$\stackrel{(c)}{=} \bar{P}^l \max_{j \in \{0,1,\dots,r\}} \frac{1}{\bar{P}^{2(r-j)n+j}} \quad (\text{A.11})$$

$$\stackrel{(d)}{=} \frac{1}{\bar{P}^{r-l}}, \quad (\text{A.12})$$

where (a) follows from Lemma 5 above, and the expectation is with respect to the distribution of  $\bar{w}_{A,\tau}$ , (b) follows from the binomial expansion and the fact that  $\mathbb{E}\bar{w}_{A,\tau}^i = 0$  when  $i$  is odd, (c) follows from  $\theta_{\bar{P}} \doteq \frac{1}{\bar{P}^n}$  and  $\mathbb{E}\bar{w}_{A,\tau}^{2j} \doteq \frac{1}{\bar{P}^j}$ , and (d) follows by substituting  $n = 1/2$  in the left hand side. From (A.12), clearly,  $I_{\bar{P}} < \bar{P}$  for large  $\bar{P}$  if  $l < r + 1$  and  $n = 1/2$ . When  $l = r + 1$  and  $n = 1/2$ , we have  $I_{\bar{P}} \preceq \bar{P}$ , and therefore we can ensure that  $I_{\bar{P}} < \bar{P}$  for large  $\bar{P}$  by scaling  $I_{\bar{P}}$  by an appropriately chosen constant scaling factor.

Next, we show that  $F(\bar{P})$  is bounded. Note that

$$F(\bar{P}) = \int_{\theta_{\bar{P}}}^1 \frac{1}{x^{2s}} f_{\hat{\sigma}}(x; \bar{P}) dx + \int_1^{\infty} \frac{1}{x^{2s}} f_{\hat{\sigma}}(x; \bar{P}) dx. \quad (\text{A.13})$$

Now, it is sufficient to show that the first integral in (A.13) is bounded, since the second integral is clearly  $< 1$ . To this end, we need the distribution of  $\hat{\sigma}$ , i.e.,  $\Pr(\sigma + \bar{w}_{A,\tau} \leq x)$ , where  $\bar{w}_{A,\tau} \sim \mathcal{N}(0, \sigma_{var}^2)$ , and  $\sigma_{var}^2 \triangleq \frac{1}{2PLB,\tau}$ . Consider

$$G(x) \triangleq \Pr(\sigma + \bar{w}_{A,\tau} \leq x) \quad (\text{A.14})$$

$$= \int_0^{\infty} f_{\sigma}(y) \int_{-\infty}^{x-y} \frac{1}{\sqrt{2\pi\sigma_{var}}} e^{-z^2/2\sigma_{var}^2} dz dy, \quad (\text{A.15})$$

where  $f_{\sigma}(y)$  is the pdf of  $\sigma$ , which is chi distributed with  $2r$  degrees of freedom. Taking the derivative of (A.14) with respect to  $x$ , we get

$$\frac{\partial G(x)}{\partial x} = \frac{J}{\sqrt{2\pi\sigma_{var}}} \int_0^{\infty} y^{2r-1} e^{-\frac{y^2}{2}} e^{-\frac{(x-y)^2}{2\sigma_{var}^2}} dy \quad (\text{A.16})$$

$$= \frac{J e^{-\beta_3}}{\sqrt{2\pi\sigma_{var}}} \int_0^{\infty} y^{2r-1} e^{\left\{-\frac{(y-\beta_1)^2}{2\beta_2}\right\}} dy, \quad (\text{A.17})$$

where  $J$  is the constant term in the standard chi pdf,  $\beta_1 \triangleq \frac{x}{1+\sigma_{var}^2}$ ,  $\beta_2 \triangleq \frac{\sigma_{var}^2}{1+\sigma_{var}^2} \doteq \frac{1}{P}$  and

$\beta_3 \triangleq \beta_2 x^2 / (2\sigma_{var}^2)$ . Let  $t = \frac{y-\beta_1}{\sqrt{\beta_2}}$  and using the binomial expansion, it can be shown that

$$\frac{\partial G(x)}{\partial x} = \frac{J \exp(-\beta_3)}{\sqrt{2\pi}\sigma_{var}} \sum_{j=0}^{2r-1} \binom{2r-1}{j} (\sqrt{\beta_2})^{2r-j} \times \frac{x^j}{(1 + \sigma_{var}^2)^j} \int_{-\beta_1/\sqrt{\beta_2}}^{\infty} t^{2r-1-j} e^{-\frac{t^2}{2}} dt. \quad (\text{A.18})$$

Now, using  $\exp(-\beta_3) \leq 1$ , we can upper bound the first term in (A.13) as

$$\int_{\theta_{\bar{P}}}^1 \frac{1}{x^{2s}} \frac{\partial G(x)}{\partial x} dx \leq \frac{J}{\sqrt{2\pi}\sigma_{var}} \sum_{j=0}^{2r-1} \binom{2r-1}{j} C_j \times \frac{(\sqrt{\beta_2})^{2r-j}}{(1 + \sigma_{var}^2)^j} \int_{\theta_{\bar{P}}}^1 x^{j-2s} dx, \quad (\text{A.19})$$

where  $s < r$ , and  $C_j \doteq 1$  is some constant that does not scale with  $\bar{P}$ . Now, the behavior of the terms above with  $\bar{P}$  is governed by

$$\frac{\beta_2^{r-j/2}}{\sigma_{var}} \int_{\theta_{\bar{P}}}^1 x^{j-2s} dx \doteq \frac{1}{j-2s+1} \left[ \frac{1}{\bar{P}^{a_1}} - \frac{1}{\bar{P}^{a_2}} \right], \quad (\text{A.20})$$

where  $a_1 \triangleq r - j/2 - 1/2$ , and  $a_2 \triangleq (-2s + j + 1)n + r - j/2 - 1/2$ . The exponent corresponding to the first term above is  $r - j/2 - 1/2 \geq 0$  for all  $0 \leq j \leq 2r - 1$ . Also, when  $n = 1/2$ , the exponent corresponding to the second term above is  $r - s > 0$  for all  $0 \leq j \leq 2r - 1$ , and hence the integral is bounded for  $1 \leq s < r$ .

Finally, let  $R_{\bar{P}} = g_m \log(\bar{P})$ . Since  $I_{\bar{P}} < \bar{P}$  and  $F(\bar{P})$  are bounded when  $0 \leq l \leq r + 1$ , using  $\left( \exp\left(\frac{L_c R_{\bar{P}}}{L_c - L_{B,\tau}}\right) - 1 \right) \doteq \bar{P}^{\frac{g_m}{\alpha}}$  in (A.9), we get  $\kappa_{\bar{P}} \doteq \frac{1}{\bar{P}^{\frac{g_m}{\alpha}-1}}$ , where  $\alpha \triangleq \frac{L_c - L_{B,\tau}}{L_c}$ . This completes the proof of Lemma 1. ■

### A.0.3 Proof of Theorem 1

Using the power control in (2.10), the outage probability in (2.8) can be written as

$$P_{out} = \Pr_{\{\hat{\sigma} \leq \theta_{\bar{P}}\}} \left\{ \alpha \log(1 + \bar{P}^l \sigma^2) < R_{\bar{P}} \right\} \quad (\text{A.21})$$

$$+ \Pr_{\{\hat{\sigma} > \theta_{\bar{P}}\}} \left\{ \alpha \log(1 + \kappa_{\bar{P}} \Phi(\hat{\sigma}^{2s}) \sigma^2) < R_{\bar{P}} \right\}$$

$$\leq \Pi_1 + \Pi_2, \quad (\text{A.22})$$

where

$$\Pi_1 \triangleq \Pr \left\{ \alpha \log(1 + \bar{P}^l \sigma^2) < R_{\bar{P}} \right\}, \quad (\text{A.23})$$

$$\Pi_2 \triangleq \Pr \left\{ \alpha \log(1 + \kappa_{\bar{P}} \Phi(\hat{\sigma}^{2s}) \sigma^2) < R_{\bar{P}} \right\}. \quad (\text{A.24})$$

In the above, we have used  $\Pr_{\{A\}}\{\cdot\}$  to mean  $\Pr\{\cdot \cap \{A\}\}$ . Using  $R_{\bar{P}} = g_m \log \bar{P}$ , we have

$\Pi_1 = \Pr \left\{ \sigma^2 < \frac{1}{\bar{P}^{l - \frac{g_m}{\alpha}}} \right\}$  for large  $\bar{P}$  and  $0 \leq l \leq r + 1$  from Lemma 1. From Lemma 5 in

Appendix A, we have,

$$\Pi_1 \preceq \frac{1}{\bar{P}^{(l - \frac{g_m}{\alpha})r}}.$$

Next, substituting for  $\Phi(\hat{\sigma}^{2s})$  from (2.9),  $\Pi_2$  can be written as,  $\Pi_2 = \Pr \left\{ \sigma^2 < \hat{\sigma}^{2s} / \kappa_{\bar{P}} \right\}$ .

Using  $\hat{\sigma}^2 \leq \hat{\sigma}_U^2 \triangleq (\sigma + |\bar{w}_{A,\tau}|)^2$  from Lemma 6 in Appendix A with  $\hat{\sigma}^2 = \sigma^2$ , we get

$$\Pi_2 \leq \Pr \left\{ \sigma^2 < \frac{1}{\kappa_{\bar{P}}} (\sigma + |\bar{w}_{A,\tau}|)^{2s} \right\} \quad (\text{A.25})$$

$$\leq \Pr \left\{ \sigma^2 < \frac{(2\sigma)^{2s}}{\kappa_{\bar{P}}} \cap \sigma^2 > |\bar{w}_{A,\tau}|^2 \right\}$$

$$+ \Pr \left\{ \sigma^2 < \frac{(2|\bar{w}_{A,\tau}|)^{2s}}{\kappa_{\bar{P}}} \cap \sigma^2 \leq |\bar{w}_{A,\tau}|^2 \right\}. \quad (\text{A.26})$$

It is straightforward to show that provided  $\kappa_{\bar{P}}$  is strictly increasing with  $\bar{P}$ , the first term in the above goes to zero exponentially with  $\bar{P}$  for  $1 \leq s < r$ . This implies that  $g_m < \alpha$ , since  $\kappa_{\bar{P}} \doteq \bar{P}^{(1-\frac{g_m}{\alpha})}$  from Lemma 1. The second term in (A.26) is upper-bounded as

$$\Pr \left\{ \sigma^2 < \frac{|\bar{w}_{A,\tau}|^{2s} 2^{2s}}{\kappa_{\bar{P}}} \right\} \stackrel{(a)}{\leq} \frac{2^{2sr} \mathbb{E} |\bar{w}_{A,\tau}|^{2sr}}{\kappa_{\bar{P}}^r r!} \quad (\text{A.27})$$

$$\stackrel{(b)}{=} \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}, \quad (\text{A.28})$$

where (a) follows from Lemma 5 in Appendix A, and the  $\doteq$  in (b) uses the fact that  $\kappa_{\bar{P}} \doteq \bar{P}^{(1-\frac{g_m}{\alpha})}$  and  $\mathbb{E} |\bar{w}_{A,\tau}|^{2sr} \doteq 1/\bar{P}^{sr}$ . Hence, we have

$$\Pr \left\{ \sigma^2 < \frac{|\bar{w}_{A,\tau}|^{2s} 2^{2s}}{\kappa_{\bar{P}}} \right\} \preceq \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}, \quad (\text{A.29})$$

which implies

$$\Pi_2 \preceq \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}}. \quad (\text{A.30})$$

Using this and  $\Pi_1 \preceq \frac{1}{\bar{P}^{r(l-\frac{g_m}{\alpha})}}$  in (A.22), we have

$$P_{out} \preceq \max \left( \frac{1}{\bar{P}^{r(l-\frac{g_m}{\alpha})}}, \frac{1}{\bar{P}^{r(s+1-\frac{g_m}{\alpha})}} \right) \quad (\text{A.31})$$

$$= \frac{1}{\bar{P}^{r(\min\{l,s+1\}-\frac{g_m}{\alpha})}}, \quad (\text{A.32})$$

for  $0 \leq l \leq r+1$ ,  $1 \leq s < r$  and  $0 \leq g_m < \alpha$ . This ends the proof of Theorem 1. ■

### A.0.4 Proof of Lemma 2

Note that  $\tilde{\mathbf{p}}_c$  can be written as

$$\tilde{\mathbf{p}}_c = \mathbf{p}_c - \tilde{\mathbf{y}}_{B,\tau_2} - \mathbb{E}\{\mathbf{p}_c - \tilde{\mathbf{y}}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\} \quad (\text{A.33})$$

$$= \frac{1}{\sqrt{\bar{P}L_{A,\tau_2}}} [\mathbb{E}\{\mathbf{w}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\} - \mathbf{w}_{B,\tau_2}]. \quad (\text{A.34})$$

Now,

$$\mathbb{E}\|\tilde{\mathbf{p}}_c\|_2^{2z} = \frac{1}{\bar{P}^z L_{A,\tau_2}^z} \mathbb{E}_{\mathbf{w}_{B,\tau_2}, \tilde{\mathbf{y}}_{B,\tau_2}} \{\mathcal{A}\} \quad (\text{A.35})$$

$$\stackrel{(a)}{\leq} \frac{1}{\bar{P}^z L_{A,\tau_2}^z} \left[ \mathbb{E}_{\tilde{\mathbf{y}}_{B,\tau_2}} \left\{ 2^{2z} \|\mathbb{E}\{\mathbf{w}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\}\|_2^{2z} \right\} \right. \\ \left. + 2^{2z} \mathbb{E}_{\mathbf{w}_{B,\tau_2}} \left\{ \|\mathbf{w}_{B,\tau_2}\|_2^{2z} \right\} \right] \quad (\text{A.36})$$

$$\stackrel{(b)}{\leq} \frac{2^{2z+1}}{\bar{P}^z L_{A,\tau_2}^z} \mathbb{E} \|\mathbf{w}_{B,\tau_2}\|_2^{2z} \quad (\text{A.37})$$

$$\doteq \frac{1}{\bar{P}^z}, \quad (\text{A.38})$$

where  $\mathcal{A} \triangleq \|\mathbb{E}\{\mathbf{w}_{B,\tau_2} | \tilde{\mathbf{y}}_{B,\tau_2}\} - \mathbf{w}_{B,\tau_2}\|_2^{2z}$ . In the above, (a) follows from the triangle inequality and using  $(a + b)^n \leq (2a)^n + (2b)^n$  for even  $n > 0$ , and (b) follows from the Jensen's inequality and the fact that  $\mathbb{E} \|\mathbf{w}_{B,\tau_2}\|_2^{2z} < \infty$  as  $\bar{P} \rightarrow \infty$ . The subscripts on the expectation in the above denote the random variables over which the expectation is taken; and  $\mathbb{E}\{X|y\}$  denotes the expectation of the random variable  $X$  conditioned on the instantiation  $Y = y$ . This completes the proof. ■

### A.0.5 Proof of Theorem 2

Using the capacity lower bound in (2.17), the outage probability can be upper bounded as

$$P_{out} \leq \Pr \{C_{AB} < R_{\bar{P}}\}, \quad (\text{A.39})$$

where  $R_{\bar{P}} \triangleq g_m \log \bar{P}$  is the target data rate. We choose  $\eta < 1$ , and arbitrarily close to 1.

We split the event in the expression for  $P_{out}$  as

$$\begin{aligned} P_{out} &\leq \Pr \left\{ C_{AB} < R_{\bar{P}} \cap \mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] \leq \frac{1}{\bar{P}^\eta} \right\} \\ &\quad + \Pr \left\{ C_{AB} < R_{\bar{P}} \cap \mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] > \frac{1}{\bar{P}^\eta} \right\} \end{aligned} \quad (\text{A.40})$$

$$\begin{aligned} &\stackrel{(a)}{\leq} \Pr \left\{ \alpha \log \left( 1 + \frac{\bar{P} \|\hat{\mathbf{p}}_c\|_2^2}{\frac{\bar{P}^{(1-\eta)}}{r} + 1} \right) < R_{\bar{P}} \right\} \\ &\quad + \Pr \left\{ \mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}] > \frac{1}{\bar{P}^\eta} \right\}, \end{aligned} \quad (\text{A.41})$$

where (a) follows by substituting  $1/\bar{P}^\eta$  in place of  $\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B,\tau_2}]$  in the first term, and removing one of the events in the intersection. Define  $\bar{R}_{\bar{P}} \triangleq \frac{(\exp\{R_{\bar{P}}/\alpha\}-1)}{\bar{P}} \left( \frac{\bar{P}^{(1-\eta)}}{r} + 1 \right)$ , and note that  $\bar{R}_{\bar{P}} \doteq \frac{1}{\bar{P}^{(\eta-\frac{gm}{\alpha})}}$ . Then, the first term in (A.41) can be written as:

$$\Pr \{ \|\hat{\mathbf{p}}_c\|_2^2 < \bar{R}_{\bar{P}} \} \stackrel{(a)}{\leq} \Pr \left\{ | \|\mathbf{p}_c\|_2 - \|\tilde{\mathbf{p}}_c\|_2 | < \sqrt{\bar{R}_{\bar{P}}} \right\} \quad (\text{A.42})$$

$$\begin{aligned} &\leq \Pr \{ E_1 \cap E_2 \} + \Pr \{ E_1 \cap E_2^c \} \\ &\leq \Pr \left\{ \|\tilde{\mathbf{p}}_c\|_2 > \sqrt{\bar{R}_{\bar{P}}} \right\} + \Pr \{ \|\mathbf{p}_c\|_2^2 < 4\bar{R}_{\bar{P}} \}, \end{aligned} \quad (\text{A.43})$$

where  $E_1 \triangleq \{ \|\mathbf{p}_c\|_2 < \|\tilde{\mathbf{p}}_c\|_2 + \sqrt{\bar{R}_{\bar{P}}} \}$  and  $E_2 \triangleq \{ \|\tilde{\mathbf{p}}_c\|_2 > \sqrt{\bar{R}_{\bar{P}}} \}$ . In the above, (a) follows from the reverse triangle inequality, and the last two inequalities follow by ignoring one

of the events in the intersection. The first term in (A.43) can be written as

$$\Pr \left\{ \|\tilde{\mathbf{p}}_c\|_2^{2\delta} > \bar{R}_{\bar{P}}^\delta \right\} \stackrel{(a)}{\leq} \frac{\mathbb{E} \|\tilde{\mathbf{p}}_c\|_2^{2\delta}}{\bar{R}_{\bar{P}}^\delta} \quad (\text{A.44})$$

$$\stackrel{(b)}{\leq} \frac{1}{\bar{P}^\delta} \frac{1}{\bar{P}^{\left(\frac{gm}{\alpha} - \eta\right)\delta}}, \quad (\text{A.45})$$

where (a) follows from the Markov inequality and (b) follows from Lemma 2. Letting

$\delta = r \frac{1}{\frac{gm}{\alpha} - \eta + 1} \left( s + 1 - \frac{gm}{\alpha} \right) > 0$ , we have

$$\Pr \left\{ \|\tilde{\mathbf{p}}_c\|_2 > \sqrt{\bar{R}_{\bar{P}}} \right\} \leq \frac{1}{\bar{P}^{r\left(s+1-\frac{gm}{\alpha}\right)}}, \quad 1 \leq s < r. \quad (\text{A.46})$$

In order to solve for the second term in (A.43), we need to handle two cases of the singular value estimate at *node A* separately; the good estimated channel case  $g \triangleq \{\hat{\sigma} \geq \theta_{\bar{P}}\}$  and the bad estimated channel case  $b \triangleq \{\hat{\sigma} < \theta_{\bar{P}}\}$ .

### Good Estimated Channel $\{\hat{\sigma} \geq \theta_{\bar{P}}\}$

When  $\hat{\sigma} \geq \theta_{\bar{P}}$ , substituting for  $\mathbf{p}_c \triangleq \sqrt{\mathcal{P}(\hat{\sigma})} \mathbf{h}$  and  $\kappa_{\bar{P}} \doteq \bar{P}^{-\frac{gm}{\alpha}}$ , and defining  $\hat{\sigma}_U \triangleq (\sigma + |\bar{w}_{A,\tau}|)$  as the upper bound on  $\hat{\sigma}$  from Lemma 6 in Appendix A, the second term in (A.43) leads to:

$$\begin{aligned} \Pr_{\{\hat{\sigma} \geq \theta_{\bar{P}}\}} \{E_3\} &\stackrel{(a)}{\leq} \Pr \left\{ \sigma^2 < 2^{2(s+1)} \sigma^{2s} \bar{R}_{\bar{P}} \cap E_4 \right\} \\ &\quad + \Pr \left\{ \sigma^2 < 2^{2(s+1)} |\bar{w}_{A,\tau}|^{2s} \bar{R}_{\bar{P}} \cap E_4^c \right\} \\ &\leq \Pr \left\{ \sigma^{2(s-1)} > \frac{2^{-2(s+1)}}{\bar{R}_{\bar{P}}} \right\} \\ &\quad + \Pr \left\{ \sigma^2 < 2^{2(s+1)} |\bar{w}_{A,\tau}|^{2s} \bar{R}_{\bar{P}} \right\}, \end{aligned} \quad (\text{A.47})$$

where  $E_3 \triangleq \left\{ \frac{\|\mathbf{h}\|_2^2}{\hat{\sigma}_U^{2s}} < 4\bar{R}_{\bar{P}} \right\}$ , and  $E_4 \triangleq \{\sigma^2 > |\bar{w}_{A,\tau}|^2\}$ . In the above, we have used  $\Pr_{\{A\}}\{\cdot\}$  to mean  $\Pr\{\cdot \cap \{A\}\}$ , as before; and (a) follows by ignoring the event  $g$ . It can be shown that first term in (A.47) decreases exponentially with  $\bar{P}^{\frac{\eta-g_m/\alpha}{s-1}}$ , as follows:

$$\Pr\{\mathcal{B}\} \stackrel{(a)}{=} \int_{1/\bar{R}_{\bar{P}}^{1/(s-1)}}^{\infty} e^{-x} x^{r-1} dx \quad (\text{A.48})$$

$$\begin{aligned} &\stackrel{(b)}{=} \exp\left\{-1/\bar{R}_{\bar{P}}^{1/(s-1)}\right\} \sum_{k=0}^{r-1} \frac{1}{(\bar{R}_{\bar{P}}^{1/(s-1)})^{r-k-1}} \\ &\stackrel{(b)}{=} e^{-\mathcal{Z}}, \end{aligned} \quad (\text{A.49})$$

where  $\mathcal{B} \triangleq \{\sigma^{2(s-1)} > \frac{1}{2^{2(s+1)}\bar{R}_{\bar{P}}}\}$ , and  $\mathcal{Z} \triangleq \bar{P}^{\frac{\eta-g_m/\alpha}{s-1}}$ . In the above, (a) follows by ignoring the constant factors and substituting for the chi-square pdf of  $\sigma^2$ . Since  $1/\bar{R}_{\bar{P}} \doteq \bar{P}^{(\eta-g_m/\alpha)}$  when  $g_m < \eta\alpha$ , and since the exponential term outside the summation dominates the polynomial terms inside the summation, we obtain (b). Note that the special case of  $s = 1$  is trivial, since this corresponds to the probability that  $\bar{R}_{\bar{P}}$  exceeds a constant, which becomes 0 for sufficiently large  $\bar{P}$ . The second term in (A.47) becomes:

$$\begin{aligned} \Pr\{\sigma^2 < 2^{2(s+1)}|\bar{w}_{A,\tau}|^{2s}\bar{R}_{\bar{P}}\} &\leq \frac{2^{2r(s+1)}\bar{R}_{\bar{P}}^r \mathbb{E}\{|\bar{w}_{A,\tau}|^{2sr}\}}{r!} \\ &\doteq \frac{1}{\bar{P}^{(\eta-\frac{gm}{\alpha})r} \bar{P}^{rs}} \end{aligned} \quad (\text{A.50})$$

$$= \frac{1}{\bar{P}^{r(s+\eta-g_m/\alpha)}} \quad (\text{A.51})$$

for  $0 \leq g_m < \eta\alpha$ . In the above, we have used Lemma 5 in Appendix A and  $\mathbb{E}|\bar{w}_{A,\tau}|^{2s} = \frac{1}{\bar{P}^s}$ . Thus, in the good estimated channel case, we have

$$\Pr_{\{\hat{\sigma} \geq \theta_{\bar{P}}\}} \left\{ \|\mathbf{p}_c\|_2^2 \leq 4\bar{R}_{\bar{P}} \right\} \preceq \frac{1}{\bar{P}^{r(s+\eta-\frac{gm}{\alpha})}}, \quad 0 \leq g_m < \eta\alpha. \quad (\text{A.52})$$

**Bad Estimated Channel**  $\{\hat{\sigma} < \theta_{\bar{P}}\}$ 

Recall that when  $\hat{\sigma} < \theta_{\bar{P}}$ , the composite channel is given by  $\mathbf{p}_c = \sqrt{\bar{P}^l} \mathbf{h}$ . With this, the second term in (A.43) becomes

$$\begin{aligned} \Pr_{\{\hat{\sigma} < \theta_{\bar{P}}\}} \{ \|\mathbf{p}_c\|_2^2 < 4\bar{R}_{\bar{P}} \} &= \Pr \left\{ \|\mathbf{h}\|_2^2 < \frac{4\bar{R}_{\bar{P}}}{\bar{P}^l} \cap b \right\} \\ &\leq \Pr \left\{ \sigma^2 < \frac{4\bar{R}_{\bar{P}}}{\bar{P}^l} \right\} \end{aligned} \quad (\text{A.53})$$

$$\stackrel{\dagger}{=} \frac{1}{\bar{P}^{rl}} \frac{1}{\bar{P}^{r(-\frac{gm}{\alpha} + \eta)}} \quad (\text{A.54})$$

$$\stackrel{\dagger}{=} \frac{1}{\bar{P}^{r(l+\eta-\frac{gm}{\alpha})}}, \quad (\text{A.55})$$

where  $0 \leq l \leq r$ . This completes the analysis of the first term in (A.41).

Now, the second term in (A.41) can be bounded as:

$$\begin{aligned} \Pr \left\{ (\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B, \tau_2}])^\zeta > \frac{1}{\bar{P}^{\zeta\eta}} \right\} &\stackrel{(a)}{\leq} \mathbb{E}(\mathbb{E}[\|\tilde{\mathbf{p}}_c\|_2^2 | \tilde{\mathbf{y}}_{B, \tau_2}])^\zeta \bar{P}^{\zeta\eta} \\ &\stackrel{(b)}{\leq} \mathbb{E}([\|\tilde{\mathbf{p}}_c\|_2^{2\zeta}]) \bar{P}^{\zeta\eta} \end{aligned} \quad (\text{A.56})$$

$$\stackrel{(c)}{\leq} \frac{1}{\bar{P}^{\zeta(1-\eta)}}, \quad (\text{A.57})$$

where  $\zeta > 0$  is an arbitrary number. In the above, (a) and (b) follow from the Markov inequality and Jensen's inequality, respectively, and (c) follows from Lemma 2. Since  $\eta < 1$ , and  $\zeta$  can be chosen arbitrarily large, the second term in (A.41) goes to zero with an arbitrarily large exponent as  $\bar{P}$  goes to infinity.

Putting (A.46), (A.52), (A.55) and (A.57) together, a DMT of  $d(g_m) = r \left( \min\{l, s\} + \eta - \frac{gm}{\alpha} \right)$  is achievable, for  $0 \leq l \leq r$ ,  $1 \leq s < r$  and  $0 \leq g_m < \eta\alpha$ . Noting that  $\eta$  is arbitrarily close to 1 completes the proof of Theorem 2. ■

# Appendix B

## Appendix for Chapter 3

### B.0.6 Proof of Theorem 4

We analyze the probability of error performance for the only real part of (3.4), as the imaginary part is statistically similar to the real part. Stacking a sequence of  $L_d \geq 1$  consecutive received symbols, from (3.4), the system model for data transmission becomes

$$\mathbf{y} = \beta \mathbf{x} + \mathbf{n}_{eff}, \quad (\text{B.1})$$

where  $\mathbf{y} \in \mathbb{R}^{L_d}$  is the received signal,  $\mathbf{x} \triangleq [x_1, \dots, x_{L_d}]$  is the transmitted signal of length  $L_d \triangleq L_c - L_{B,\tau}$  symbols,  $\beta \triangleq c_{\bar{P}} \sqrt{\bar{P}^\gamma L_{B,\tau}} \sqrt{(r-1)(r-2)} \doteq \bar{P}^{1/2}$ , and  $\mathbf{n}_{eff} \triangleq \|\mathbf{h}\|_2 c_{\bar{P}} \mathbb{R}\{w_{A,\tau}\} \mathbf{x} + \mathbf{w}$ . Here, the noise  $\mathbf{w} \in \mathbb{R}^{L_d}$  is distributed as  $\mathcal{N}(0, I_{L_d \times L_d}/2)$ . Now, we propose a simple data transmission scheme with a rate of  $\frac{g_m}{2\alpha} \log \bar{P}$  per real dimension that achieves an infinite diversity order. We partition the interval  $[0, 1]$  into  $\lceil \bar{P}^{g_m/2\alpha} \rceil$  bins of equal length, i.e., each bin is of length at least  $\bar{P}^{-g_m/2\alpha}$ . Each component of  $\mathbf{x}$  belongs to the set  $\mathcal{X}$  whose components are the bin centroids, i.e.,  $\mathcal{X} \triangleq \left\{ \frac{2(i-1)+1}{2} \bar{P}^{-g_m/2\alpha}; i = 1, 2, \dots, \lceil \bar{P}^{g_m/2\alpha} \rceil \right\}$ . Now, transmitting  $\mathbf{x} \subseteq \mathcal{X}^{L_d}$  conveys  $\frac{g_m L_d}{2\alpha} \log \bar{P}$  bits of information, which implies that a multiplexing gain of  $g_m/2\alpha$  is achieved per

channel use per real dimension, as desired. Also, note that there are roughly  $2^{L_d R} = \bar{P}^{L_d g_m / 2\alpha}$  number of codewords, denoted  $\mathbf{x}_1, \dots, \mathbf{x}_M$ ,  $M \triangleq \lceil \bar{P}^{L_d g_m / 2\alpha} \rceil$ . Further, assume that the codewords (message symbols) are equally likely. Now, we derive an upper bound on the probability of error.

### Probability of Error Analysis

The decoding rule is as follows. Upon receiving  $y$ , choose  $\mathbf{x}_i$  as the transmitted codeword if  $\|\mathbf{y} - \beta \mathbf{x}_i\|_2^2 < \|\mathbf{y} - \beta \mathbf{x}_j\|_2^2$  for all  $j \neq i, j = 1, 2, \dots, M$ . The pairwise error probability is defined as

$$P_e \triangleq \mathbb{E} \Pr \left\{ \|\mathbf{y} - \beta \mathbf{x}_i\|_2^2 > \|\mathbf{y} - \beta \mathbf{x}_j\|_2^2 \mid \mathbf{h}, \mathbf{x}_i, \mathbf{x}_j \right\}, \quad (\text{B.2})$$

where the expectation is with respect to  $\mathbf{x}_i, \mathbf{x}_j$  and  $\mathbf{h}$ . Now, (B.2) can be simplified as

$$P_e = \mathbb{E} \Pr \left\{ -(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{n}_{eff} \geq \frac{\beta \|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2} \mid \mathbf{x}_i, \mathbf{x}_j, \mathbf{h} \right\}. \quad (\text{B.3})$$

Now, given  $\mathbf{x}_i, \mathbf{x}_j$ , and  $\mathbf{h}$ ,  $-(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{n}_{eff}$  is distributed as

$$\mathcal{N} \left( 0, \frac{c_P^2 \|\mathbf{h}\|_2^2 |(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{x}_i|^2 + \|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2} \right).$$

Hence, it can be shown that

$$P_e = \mathbb{E} Q \left( \frac{\beta \|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sqrt{2} \sqrt{c_P^2 \|\mathbf{h}\|_2^2 |(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{x}_i|^2 + \|\mathbf{x}_i - \mathbf{x}_j\|_2^2}} \right) \quad (\text{B.4})$$

Using  $Q(x) \leq \exp(-x^2/2)$ , we can upper bound the pairwise error probability as

$$P_e \leq \mathbb{E} \exp \left\{ -\frac{1}{4} \frac{\beta^2 \|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{1 + c_P^2 \|\mathbf{h}\|_2^2 A} \right\}, \quad (\text{B.5})$$

where  $A \triangleq \frac{|(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{x}_i|^2}{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}$ . Using the Cauchy-Schwarz inequality, we have  $A \leq \|\mathbf{x}_i\|_2^2$ .

Substituting this in (B.5), we get

$$P_e \leq \mathbb{E} \exp \left\{ -\frac{1}{4} \frac{\beta^2 \|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{1 + c_{\bar{P}}^2 \|\mathbf{h}\|_2^2 \|\mathbf{x}_i\|_2^2} \right\}. \quad (\text{B.6})$$

Now, using the fact that  $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \geq d_{min}^2 \triangleq \bar{P}^{-g_m/\alpha}$ , and  $\|\mathbf{x}_i\|_2^2 \leq L_d$ , we get the following bound:

$$P_e \leq \mathbb{E} \exp \left\{ -\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \|\mathbf{h}\|_2^2 L_d} \right\}. \quad (\text{B.7})$$

From the law of total expectation, we can write the above as

$$\begin{aligned} P_e &\leq \mathbb{E} \left[ \exp \left\{ -\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \|\mathbf{h}\|_2^2 L_d} \right\} \middle| \mathcal{E} \right] \Pr\{\mathcal{E}\} \\ &\quad + \mathbb{E} \left[ \exp \left\{ -\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \|\mathbf{h}\|_2^2 L_d} \right\} \middle| \mathcal{E}^c \right] \Pr\{\mathcal{E}^c\} \end{aligned} \quad (\text{B.8})$$

$$\leq \exp \left\{ -\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \bar{P}^\theta} \right\} + \Pr\{\mathcal{E}^c\} \quad (\text{B.9})$$

$$\preceq \exp \left\{ -\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \bar{P}^\theta} \right\} + \exp\{-\bar{P}^\theta\}. \quad (\text{B.10})$$

where  $\mathcal{E} \triangleq \{\|\mathbf{h}\|_2^2 \leq \bar{P}^\theta/L_d\}$  for some  $\theta > 0$  to be chosen later. By substituting for  $d_{min}^2 \doteq \bar{P}^{-g_m/\alpha}$  and  $c_{\bar{P}}^2 \doteq \bar{P}^{1-\gamma}$ , it is easy to see that  $\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \bar{P}^\theta} \doteq \bar{P}^{1-g_m/\alpha}$  if  $\theta \leq \gamma - 1$ , else  $\frac{1}{4} \frac{\beta^2 d_{min}^2}{1 + c_{\bar{P}}^2 \bar{P}^\theta} \doteq \bar{P}^{\gamma-\theta-g_m/\alpha}$ . Using this in (B.10), we get

$$P_e \preceq \exp\{-\bar{P}^{E(\theta)}\}, \quad (\text{B.11})$$

where

$$E(\theta) = \begin{cases} \min\{\theta, 1 - g_m/\alpha\}, & \text{if } \theta \leq \gamma - 1 \\ \min\{\theta, (\gamma - \frac{g_m}{\alpha} - \theta)\}, & \text{if } \theta > \gamma - 1. \end{cases} \quad (\text{B.12})$$

Now, since we are free to choose  $\theta > 0$ , with a little algebra, it is easy to show that when  $\gamma \geq 2 - g_m/\alpha$ , setting  $\theta = \gamma - 1$  maximizes  $E(\theta)$ , resulting in  $E(\theta) = 1 - g_m/\alpha$ ; and when  $\gamma < 2 - g_m/\alpha$ , setting  $\theta = \frac{1}{2}(\gamma - \frac{g_m}{\alpha})$  maximizes  $E(\theta)$ , resulting in  $E(\theta) = \frac{1}{2}(\gamma - \frac{g_m}{\alpha})$ . We also require the exponent of  $\bar{P}$  in the probability of error terms in (B.10) to be positive, which leads to  $g_m < \gamma\alpha$ . Since there are  $2^{L_d R} = \bar{P}^{L_d g_m/2\alpha}$  codewords, the upper bound on  $P_e$  must be multiplied by  $\bar{P}^{L_d g_m/2\alpha}$  to upper bound the average probability of error. However, this does not change the probability of error exponent as  $\bar{P}^{L_d g_m/2\alpha}$  is polynomial in  $\bar{P}$  while the upper bound on the pairwise error probability in (B.11) is exponential in  $\bar{P}$  for all  $L_d \geq 1$ . Putting the above together and writing the constraint in (B.12) in terms of  $g_m$  leads to (3.14); and this completes the proof. ■

# Appendix C

## Appendix for Chapter 4

### C.0.7 Constrained Cramér-Rao Lower Bound

**Theorem 14.** *The estimate of the dominant singular vector,  $\hat{\mathbf{v}}_1$  in (4.6), conditioned on  $\sigma_1$ , asymptotically achieves the CCRLB =  $\frac{(2n_A-1)}{2\sigma_1^2 P_{B,\tau} L_{B,\tau}}$  as  $P_{B,\tau} L_{B,\tau} \rightarrow \infty$ , and the corresponding mean square error in the estimate is given by*

$$\mathbb{E}[\|\mathbf{v}_1 - \hat{\mathbf{v}}_1\|_2^2 | \sigma_1] = \frac{(2n_A - 1)}{2\sigma_1^2 P_{B,\tau} L_{B,\tau}} + \mathcal{O}_{\frac{3}{2}}, \quad (\text{C.1})$$

where  $\mathcal{O}_r \triangleq \mathcal{O}\left(\frac{1}{(P_{B,\tau} L_{B,\tau})^r}\right)$  for rational  $r$ .

*Proof:* Using (4.5), for large  $P_{B,\tau} L_{B,\tau}$ ,  $\hat{\mathbf{v}}_1$  can be written as

$$\hat{\mathbf{v}}_1 = \frac{\mathbf{y}_{A,\tau}}{\|\mathbf{y}_{A,\tau}\|_2} = \mathbf{v}_1 - \frac{1}{\sigma_1} \Re \left\{ \frac{\mathbf{v}_1^H \mathbf{w}_{A,\tau}}{\sqrt{P_{B,\tau} L_{B,\tau}}} \right\} \mathbf{v}_1 + \frac{1}{\sigma_1} \frac{\mathbf{w}_{A,\tau}}{\sqrt{P_{B,\tau} L_{B,\tau}}} + \mathcal{O}_1.$$

From the above, conditioned on  $\sigma_1$ , the MSE is given by  $\text{MSE} \triangleq \mathbb{E}_{\sigma_1} \|\mathbf{v}_1 - \hat{\mathbf{v}}_1\|_2^2$ . In turn,

this can be simplified as:

$$\text{MSE} = \frac{\mathbb{E}_{\sigma_1} \left\| \Re \{ \mathbf{v}_1^H \mathbf{w}_{A,\tau} \} \mathbf{v}_1 - \mathbf{w}_{A,\tau} \right\|_2^2}{\sigma_1^2 P_{B,\tau} L_{B,\tau}} + \mathcal{O}_{\frac{3}{2}} \quad (\text{C.2})$$

$$= \frac{(2n_A - 1)}{2\sigma_1^2 P_{B,\tau} L_{B,\tau}} + \mathcal{O}_{\frac{3}{2}}, \quad (\text{C.3})$$

where we have used  $\mathbb{E}_{\sigma_1} \left( \left\{ \frac{\Re\{\mathbf{v}_1^H \mathbf{w}_{A,\tau}\}}{\sqrt{P_{B,\tau} L_{B,\tau}}} \right\}^2 \right) = \frac{1}{2P_{B,\tau} L_{B,\tau}}$  and  $\frac{\mathbb{E}(\mathbf{w}_{A,\tau}^H \mathbf{w}_{A,\tau})}{P_{B,\tau} L_{B,\tau}} = \frac{n_A}{P_{B,\tau} L_{B,\tau}}$  in obtaining (C.3) from (C.2). Next, we show that the first term in (C.3) coincides with the CCRLB. Writing (4.5) by stacking real and imaginary parts, we have,

$$\bar{\mathbf{y}}_{A,\tau} \triangleq \begin{pmatrix} \Re\{\mathbf{y}_{A,\tau}\} \\ \Im\{\mathbf{y}_{A,\tau}\} \end{pmatrix} = \sigma_1 \underbrace{\begin{pmatrix} \Re\{\mathbf{v}_1\} \\ \Im\{\mathbf{v}_1\} \end{pmatrix}}_{\bar{\mathbf{v}}_1} + \underbrace{\begin{pmatrix} \Re\{\mathbf{w}_{A,\tau}\} \\ \Im\{\mathbf{w}_{A,\tau}\} \end{pmatrix}}_{\bar{\mathbf{w}}_A}. \quad (\text{C.4})$$

The unitary constraint equation is  $f(\mathbf{y}) = \mathbf{y}^T \mathbf{y} - 1 = 0$ , where  $\mathbf{y} \in \mathbb{R}^{2n_A \times 1}$ . The derivative of  $f(\cdot)$  is given by  $\frac{\partial f}{\partial \mathbf{y}} = 2\mathbf{y}$ . Define a matrix  $U_1 \in \mathbb{R}^{2n_A \times 2n_A - 1}$  such that  $U_1^T U_1 = I_{2n_A - 1 \times 2n_A - 1}$  and  $U_1^T \mathbf{y} = 0$ . Since  $p(\bar{\mathbf{y}}_{A,\tau} | \bar{\mathbf{v}}_1, \sigma_1) \sim \mathcal{N}(\sigma_1 \bar{\mathbf{v}}_1, \frac{1}{2P_{B,\tau} L_{B,\tau}} I_{2n_A \times 2n_A})$ , it is straightforward to see that

$$\Delta \triangleq \frac{\partial \log p(\bar{\mathbf{y}}_{A,\tau} | \bar{\mathbf{v}}_1, \sigma_1)}{\partial \bar{\mathbf{v}}_1} = 2\sigma_1 P_{B,\tau} L_{B,\tau} \bar{\mathbf{w}}_A. \quad (\text{C.5})$$

The Fisher information matrix  $J$  is given by

$$J \triangleq \mathbb{E}(\Delta \Delta^T) = 2\sigma_1^2 P_{B,\tau} L_{B,\tau} I_{2n_A \times 2n_A}.$$

Now, from Theorem 1 of [65], the CCRLB is given by

$$\mathbb{E} \{ \|\mathbf{v}_1 - \hat{\mathbf{v}}_1\|_F^2 | \sigma_1 \} \geq \text{Tr} \{ U_1 (U_1^T J U_1)^{-1} U_1^T \} \quad (\text{C.6})$$

$$= \frac{(2n_A - 1)}{2\sigma_1^2 P_{B,\tau} L_{B,\tau}}, \quad (\text{C.7})$$

which concludes the proof. ■

### C.0.8 Derivation of (4.12)

For a BF system, the estimate of the dominant singular vector can be written as

$$\hat{\mathbf{v}}_1 = \frac{\mathbf{y}_{A,\tau}}{\|\mathbf{y}_{A,\tau}\|_2} = \mathbf{v}_1 - \frac{1}{\sigma_1} \frac{\Re\{\mathbf{v}_1^H \mathbf{w}_{A,\tau}\}}{\sqrt{P_{B,\tau} L_{B,\tau}}} \mathbf{v}_1 + \frac{1}{\sigma_1} \frac{\mathbf{w}_{A,\tau}}{\sqrt{P_{B,\tau} L_{B,\tau}}} + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right),$$

where  $\frac{1}{\sqrt{1+x}} = 1 - \frac{x}{2} + \mathcal{O}(x^2)$  has been used. Note that at high training powers and/or duration, the higher order term in (C.8) becomes negligible. Thus, at high training power, we get

$$\tilde{\mathbf{v}}_1 = \mathbf{v}_1 - \hat{\mathbf{v}}_1 \approx \frac{1}{\sigma_1} \frac{\Re\{\mathbf{v}_1^H \mathbf{w}_{A,\tau}\}}{\sqrt{P_{B,\tau} L_{B,\tau}}} \mathbf{v}_1 - \frac{1}{\sigma_1} \frac{\mathbf{w}_{A,\tau}}{\sqrt{P_{B,\tau} L_{B,\tau}}}. \quad (\text{C.8})$$

From the above equation,  $\mathbb{E}\{\tilde{\mathbf{v}}_1|H\} \approx 0$ , ignoring higher order terms, where the expectation is taken with respect to the distribution of  $\mathbf{w}_{A,\tau}$ . Moreover,

$$\mathbb{E}\{|\mathbf{v}_1^H \tilde{\mathbf{v}}_1|^2\} \approx \frac{1}{2\sigma_1^2 P_{B,\tau} L_{B,\tau}}. \quad (\text{C.9})$$

Using (C.9) and  $\mathbb{E}\{\tilde{\mathbf{v}}_1|H\} \approx 0$  in (4.10), we get the approximation for the lower bound on the capacity given by (4.12).

### C.0.9 Proof of Theorem 5

Let  $P_{A,d} L_{A,d} \triangleq \alpha \rho L_c$ , where  $0 \leq \alpha \leq 1$ . Then,  $P_{B,\tau} L_{B,\tau} \triangleq (1 - \alpha) \rho L_c$ . Clearly, since  $C_{B,A,\text{approx}}$  is a monotonic function of  $\rho_{\text{eff}} \triangleq \frac{P_{A,d}}{1 + \frac{P_{A,d}}{2P_{B,\tau} L_{B,\tau}}}$ , maximizing  $C_{B,A,\text{approx}}$  amounts to maximizing  $\rho_{\text{eff}}$  with respect to  $\alpha$  [8]. Substituting for  $P_{A,d}$ ,  $P_{A,\tau}$ , and  $L_{B,\tau}$ ,  $\rho_{\text{eff}}$  can be written as a function of  $\alpha$  as

$$\rho_{\text{eff}} = \left( \frac{2\rho L_c}{2L_{A,d} - 1} \right) \left[ \frac{\alpha(1 - \alpha)}{-\alpha + \frac{2L_{A,d}}{2L_{A,d} - 1}} \right]. \quad (\text{C.10})$$

Since the function inside the square brackets in (C.10) is a concave function in  $\alpha$ , differentiating it with respect to  $\alpha$ , equating it to zero and solving, we get  $\alpha^* = \frac{\sqrt{2L_{A,d}}}{\sqrt{2L_{A,d}+1}}$ . Substituting this in (C.10) and simplifying, we get the optimal capacity lower bound in (4.20). Finally, we get the expression in (4.19) by substituting for  $\alpha^*$  in  $P_{A,d}^* L_{A,d} = \alpha^* \rho L_c$  and  $P_{A,\tau}^* L_{B,\tau} = (1 - \alpha^*) \rho L_c$ , respectively. ■

### C.0.10 Proof of Theorem 6

Differentiating (4.20) with respect to  $L_{A,d}$ , we get

$$\frac{\partial C_{B,A,\text{approx}}}{\partial L_{A,d}} = \frac{1}{L_c} \mathbb{E} \log(1 + \rho_{\text{eff}}^* \sigma_1^2) + \frac{L_{A,d}}{L_c} \mathbb{E} \left[ \frac{\sigma_1^2}{1 + \sigma_1^2 \rho_{\text{eff}}^*} \frac{\partial \rho_{\text{eff}}^*}{\partial L_{A,d}} \right], \quad (\text{C.11})$$

$$\stackrel{(a)}{=} \frac{1}{L_c} \mathbb{E} \log(1 + \rho_{\text{eff}}^* \sigma_1^2) - \frac{1}{L_c} \frac{\sqrt{2L_{A,d}}}{1 + \sqrt{2L_{A,d}}} \mathbb{E} \left[ \frac{\sigma_1^2 \rho_{\text{eff}}^*}{1 + \sigma_1^2 \rho_{\text{eff}}^*} \right], \quad (\text{C.12})$$

where  $\rho_{\text{eff}}^* \triangleq \frac{2\rho L_c}{(\sqrt{2L_{A,d}+1})^2}$ , and (a) is obtained by substituting for  $\frac{\partial \rho_{\text{eff}}^*}{\partial L_{A,d}}$  and simplifying. Since  $\frac{\sqrt{2L_{A,d}}}{1 + \sqrt{2L_{A,d}}} < 1$ ,  $\frac{\partial C_{B,A,\text{approx}}}{\partial L_{A,d}}$  in (C.12) can be lower bounded as follows:

$$\frac{\partial C_{B,A,\text{approx}}}{\partial L_{A,d}} > \frac{1}{L_c} \mathbb{E} \log(1 + \rho_{\text{eff}}^* \sigma_1^2) - \frac{1}{L_c} \mathbb{E} \left[ \frac{\sigma_1^2 \rho_{\text{eff}}^*}{1 + \sigma_1^2 \rho_{\text{eff}}^*} \right]. \quad (\text{C.13})$$

In the right hand side above,  $\log(1 + y) - y/(1 + y) \geq 0$  for  $y \geq 0$  and since  $\sigma_1^2 \geq 0$ , (C.13) implies that  $C_{B,A,\text{approx}}$  is a monotonically increasing function of  $L_{A,d}$ . Hence, the optimal data duration is  $L_{A,d}^* = L_c - 1$ . ■

### C.0.11 Derivation of (4.25)

At low data SNR, using  $\log(1 + x) \approx x$ , we have  $C_{AB,L} \approx \frac{L_c - L_{A,\tau}}{L_c} \mathbb{E}(\text{SNR}_{\text{eff}})$ . Also, using a first order approximation,  $\mathbb{E}(\text{SNR}_{\text{eff}})$  can be written as the ratio of the means of the

numerator and denominator [31, 66].<sup>1</sup> Thus, we get

$$C_{AB,L} \approx \frac{L_c - L_{A,\tau}}{L_c} \frac{P_{A,d} \mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2}{\frac{P_{A,d}}{n_B} \mathbb{E} \left\| \tilde{\mathbf{b}}_{\text{mmse}} \right\|_2^2 + 1}, \quad (\text{C.14})$$

$$\geq \frac{L_c - L_{A,\tau}}{L_c} \frac{P_{A,d} (\mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2)}{\frac{P_{A,d}}{n_B} \mathbb{E} \left\| \tilde{\mathbf{b}}_{\text{mmse}} \right\|_2^2 + 1}. \quad (\text{C.15})$$

The inequality in (C.15) is obtained using  $\mathbb{E} \left\| \tilde{\mathbf{b}}_{\text{mmse}} \right\|_2^2 \leq \mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2$ , where  $\tilde{\mathbf{b}}_{\text{mmse}}$  is the error in the *Linear* MMSE (LMMSE) estimate of  $\mathbf{b}$ ,  $\hat{\mathbf{b}}_{\text{mmse}}$  is the LMMSE estimate of  $\mathbf{b}$ , and

$$\mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2 = \mathbb{E} \left\| \mathbf{b} \right\|_2^2 - \mathbb{E} \left\| \tilde{\mathbf{b}}_{\text{mmse}} \right\|_2^2 \quad (\text{C.16})$$

$$\geq \mathbb{E} \left\| \mathbf{b} \right\|_2^2 - \mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2 \quad (\text{C.17})$$

$$= \mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2. \quad (\text{C.18})$$

Now, the LMMSE estimate is given by  $\hat{\mathbf{b}}_{\text{mmse}} = a \mathbf{y}_{B,\tau}$ , where<sup>2</sup>

$$a \triangleq \frac{\sqrt{P_{A,\tau} L_{A,\tau}} \mathbb{E} \sigma_1^2}{n_B + P_{A,\tau} L_{A,\tau} \mathbb{E} \sigma_1^2}. \quad (\text{C.19})$$

A simple substitution yields

$$\mathbb{E} \left\| \hat{\mathbf{b}}_{\text{mmse}} \right\|_2^2 = \frac{P_{A,\tau} L_{A,\tau} (\mathbb{E} \sigma_1^2)^2 \mathbb{E} \left\{ \mathbf{y}_{B,\tau}^H \mathbf{y}_{B,\tau} \right\}}{(n_B + P_{A,\tau} L_{A,\tau} \mathbb{E} \sigma_1^2)^2} \quad (\text{C.20})$$

$$= \frac{P_{A,\tau} L_{A,\tau} (\mathbb{E} \sigma_1^2)^2}{n_B + P_{A,\tau} L_{A,\tau} \mathbb{E} \sigma_1^2}. \quad (\text{C.21})$$

<sup>1</sup>The authors in [37] have commented in detail about this approximation. In particular, when  $\mathbf{b}$  is Gaussian (e.g., when  $n_A = 1$  and the channel undergoes Rayleigh distributed fading), the approximation error is zero. For non-Gaussian  $\mathbf{b}$ , it corresponds to replacing the mean of  $\left\| \tilde{\mathbf{b}}_{\text{mmse}} \right\|_2$  conditioned on  $\mathbf{y}_{B,\tau}$  in the denominator of  $\text{SNR}_{\text{eff}}$  with its unconditional expectation.

<sup>2</sup>Note that the estimate above satisfies  $\mathbb{E} \left\{ \hat{\mathbf{b}}_{\text{mmse}}^H \tilde{\mathbf{b}}_{\text{mmse}} \right\} = 0$ .

where the last equality is obtained using  $\mathbb{E} \{ \mathbf{y}_{B,\tau}^H \mathbf{y}_{B,\tau} \} = n_B + P_{A,\tau} L_{A,\tau} \mathbb{E} \sigma_1^2$ . Substituting this in (C.15), we get the expression in (4.25).

### C.0.12 Proof of Theorem 7

Let  $\alpha$  denote the fraction of the total power allocated to data transmission. We have  $P_{A,d} L_{A,d} = \alpha \rho L_c$  and  $P_{A,\tau} L_{A,\tau} = (1 - \alpha) \rho L_c$ . Using this, when  $L_{A,d} > 1$ , the  $\text{SNR}_L$  defined after (4.25) becomes

$$\text{SNR}_L = \frac{\frac{\alpha \rho L_c}{L_{A,d}} (1 - \alpha) \rho L_c}{\mathbb{E} \sigma_1^2 \left[ \frac{\alpha \rho L_c}{L_{A,d}} + (1 - \alpha) \rho L_c \right] + n_B}, \quad (\text{C.22})$$

$$= \frac{\rho L_c}{(L_{A,d} - 1) \mathbb{E} \sigma_1^2} \frac{\alpha(1 - \alpha)}{(-\alpha + \theta)}, \quad (\text{C.23})$$

where

$$\theta \triangleq \frac{L_{A,d}}{L_{A,d} - 1} + \frac{n_B L_{A,d}}{\mathbb{E} \sigma_1^2 \rho L_c (L_{A,d} - 1)} > 1. \quad (\text{C.24})$$

Thus, the problem in (4.27) is equivalent to

$$\max_{\alpha: \alpha \in [0,1]} \frac{\alpha(1 - \alpha)}{-\alpha + \theta}. \quad (\text{C.25})$$

Since the function  $\frac{\alpha(1-\alpha)}{\theta-\alpha}$  is concave function of  $\alpha \in [0, 1]$ , differentiating it with respect to  $\alpha$  and setting it equal to zero, and using the fact that  $\theta > 1$ , we obtain the optimal  $\alpha$  as

$$\alpha = \theta - \sqrt{\theta(\theta - 1)}. \quad (\text{C.26})$$

Substituting the above in the expression for  $\text{SNR}_L$ , we have

$$\text{SNR}_L = \frac{\rho L_c}{(L_{A,d} - 1) \mathbb{E} \sigma_1^2} \left( \sqrt{\theta} - \sqrt{\theta - 1} \right)^2. \quad (\text{C.27})$$

When  $L_{A,d} = 1$ , it is easy to see from the definition of  $\text{SNR}_L$  that

$$\text{SNR}_L = \frac{(\rho L_c)^2(1-\alpha)\alpha}{\mathbb{E}\sigma_1^2\rho L_c + n_B}. \quad (\text{C.28})$$

Thus, maximizing  $\frac{1}{L_c}\text{SNR}_L$  amounts to maximizing  $\alpha(1-\alpha)$ ; the solution is  $\alpha = 0.5$ . The corresponding lower bound in (4.30) easily follows by substituting for  $\alpha = 0.5$  above and using the result in (4.25). ■

### C.0.13 Proof of Theorem 8

We prove this by showing that the derivative  $\frac{\partial C_{A,B,\text{approx}}^*}{\partial L_{A,d}} > 0$  for all  $L_{A,d} > 1$ . Taking the derivative of (4.30) with respect to  $L_{A,d}$ ,  $\frac{1}{\rho\mathbb{E}\sigma_1^2} \frac{\partial C_{A,B,\text{approx}}^*}{\partial L_{A,d}}$  can be written as

$$\begin{aligned} & \frac{L_{A,d}X^2}{\sqrt{\theta(\theta-1)}(L_{A,d}-1)^3} \left[ 1 + \frac{n_B}{\rho L_c \mathbb{E}\sigma_1^2} \right] - \frac{X^2}{(L_{A,d}-1)^2} \\ &= \frac{X^2}{(L_{A,d}-1)^2} \left( \frac{L_{A,d} \left( 1 + \frac{n_B}{\rho L_c \mathbb{E}\sigma_1^2} \right)}{\sqrt{\theta(\theta-1)}(L_{A,d}-1)} - 1 \right), \end{aligned} \quad (\text{C.29})$$

where  $X \triangleq (\sqrt{\theta} - \sqrt{\theta-1})$ . The above is greater than zero if  $\left( \frac{L_{A,d} \left( 1 + \frac{n_B}{\rho L_c \mathbb{E}\sigma_1^2} \right)}{\sqrt{\theta(\theta-1)}(L_{A,d}-1)} \right) > 1$ . Since  $\sqrt{\theta(\theta-1)} < \theta$ , we have

$$\left( \frac{L_{A,d} \left( 1 + \frac{n_B}{\rho L_c \mathbb{E}\sigma_1^2} \right)}{\sqrt{\theta(\theta-1)}(L_{A,d}-1)} \right) > \left( \frac{L_{A,d} \left( 1 + \frac{n_B}{\rho L_c \mathbb{E}\sigma_1^2} \right)}{\theta(L_{A,d}-1)} \right) = 1, \quad (\text{C.30})$$

where the last equality is obtained by substituting for  $\theta$  from (4.29). Thus,  $\frac{\partial C_{A,B,\text{approx}}^*}{\partial L_{A,d}} > 0$ .

Finally, comparing  $C_{A,B,\text{approx}}^*$  with  $L_{A,d} = 1$  and  $L_{A,d} = 2$ , we can eliminate the special case of  $L_{A,d} = 1$  from the solution. Therefore,  $L_{A,d}^* = L_c - 1$ . ■

# Appendix D

## Appendix for Chapter 5

### D.0.14 Proof of Theorem 9

In order to compute an approximate expression for  $\sum_{k=1}^m \mathbb{E} \|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2$ , we first find the Taylor series expansion of the estimate  $\hat{\mathbf{v}}_k$ , as follows. Substituting for  $\bar{\mathbf{y}}_{k,A,\tau}$ , the  $k^{\text{th}}$  column of  $\bar{Y}_{A,\tau}$ , in (5.4), the estimate of the  $k^{\text{th}}$  singular vector in (5.5) becomes:

$$\hat{\mathbf{v}}_k = \frac{(\sigma_k d_k \sqrt{\phi_c} \mathbf{v}_k + \bar{\mathbf{w}}_{k,A,\tau})}{\|(\sigma_k \sqrt{\phi_c} d_k \mathbf{v}_k + \bar{\mathbf{w}}_{k,A,\tau})\|_2}, \quad (\text{D.1})$$

$$= \frac{\mathbf{v}_k + \frac{\bar{\mathbf{w}}_{k,A,\tau}}{\sigma_k d_k \sqrt{\phi_c}}}{\sqrt{1+x}}, \quad (\text{D.2})$$

$$= \mathbf{v}_k - \frac{\Re\{\mathbf{v}_k^H \bar{\mathbf{w}}_{k,A,\tau}\} \mathbf{v}_k}{\sigma_k d_k \sqrt{\phi_c}} + \frac{\bar{\mathbf{w}}_{k,A,\tau}}{\sigma_k d_k \sqrt{\phi_c}} + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right), \quad (\text{D.3})$$

where the last expression follows by using  $\frac{1}{\sqrt{1+x}} = 1 - x/2 + \mathcal{O}(x^2)$ , where

$$x \triangleq \frac{2\Re\{\mathbf{v}_k^H \bar{\mathbf{w}}_{k,A,\tau}\}}{\sigma_k d_k \sqrt{\phi_c}} + \frac{\bar{\mathbf{w}}_{k,A,\tau}^H \bar{\mathbf{w}}_{k,A,\tau}}{\sigma_k^2 d_k^2 \phi_c},$$

and retaining only the terms of the order strictly less than  $\mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right)$ . Thus,  $\hat{V}_m = \hat{V}_{m,\text{approx}} + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right)$ , with  $\hat{V}_{m,\text{approx}} = V_m + E$ , where the error matrix  $E$  is as defined in the theorem. Let  $\mathbf{e}_k$  denote the  $k^{\text{th}}$  column of  $E$ . Let  $\hat{\mathbf{v}}_{k,\text{approx}}$  be the  $k^{\text{th}}$  column of the

matrix  $\hat{V}_{m,\text{approx}}$ . From (D.3), we have

$$\mathbb{E} \left\{ \|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2 \right\} = \mathbb{E} \left\| \mathbf{e}_k + \mathcal{O} \left( \frac{1}{P_{B,\tau} L_{B,\tau}} \right) \right\|_2^2, \quad (\text{D.4})$$

$$= \mathbb{E} \left\{ \|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2 \right\} + \mathcal{O} \left( \frac{1}{(P_{B,\tau} L_{B,\tau})^2} \right), \quad (\text{D.5})$$

where  $\mathcal{O} \left( \frac{1}{(P_{B,\tau} L_{B,\tau})^2} \right)$  follows from the fact that  $\mathcal{O} \left( \frac{1}{P_{B,\tau} L_{B,\tau}} \right)$  contains random variables of the form  $x^2$ , which results in the cross terms  $\mathbb{E}\{\mathbf{e}_k^H x^2\} = 0$  due to the Gaussian distribution of the noise. Subtracting  $\mathbb{E} \left\{ \|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2 \right\}$  on both sides of (D.5), and summing over  $k$  and taking the absolute value, we get

$$\begin{aligned} \left| \mathbb{E} \left\| V_m - \hat{V}_m \right\|_F^2 - \mathbb{E} \left\| V_m - \hat{V}_{\text{approx}} \right\|_F^2 \right| &= \left| \sum_{k=1}^m \mathbb{E} \left\{ \|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2 \right\} - \sum_{k=1}^m \mathbb{E} \left\{ \|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2 \right\} \right|, \\ &= \mathcal{O} \left( \frac{1}{(P_{B,\tau} L_{B,\tau})^2} \right). \end{aligned} \quad (\text{D.6})$$

The proof is complete once we evaluate  $\sum_{k=1}^m \mathbb{E} \left\{ \|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2 \right\} = \sum_{i=1}^m \mathbb{E} \|\mathbf{e}_k\|_2^2$ , which is done as follows:

$$\mathbb{E} \|\mathbf{e}_k\|_2^2 = \mathbb{E} \left[ \mathbb{E}_{|\sigma_1, \dots, \sigma_m} \left\{ \left\| \frac{-\Re\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\}}{\sigma_k d_k \sqrt{\phi_c}} \mathbf{v}_k + \frac{\mathbf{w}_{k,A,\tau}}{\sigma_k d_k \sqrt{\phi_c}} \right\|_2^2 \right\} \right], \quad (\text{D.7})$$

$$= \mathbb{E} \frac{1}{\sigma_k^2 d_k^2 \phi_c} \left[ \mathbb{E} \|\mathbf{w}_{k,A,\tau}\|_2^2 - \mathbb{E} |\Re\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\}|^2 \right], \quad (\text{D.8})$$

$$= \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \mathbb{E} \frac{1}{\sigma_k^2 d_k^2 \phi_c}. \quad (\text{D.9})$$

The facts that  $\mathbb{E} |\Re\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\}|^2 = \frac{1}{2P_{B,\tau} L_{B,\tau}}$  and  $\mathbb{E} \|\mathbf{w}_{k,A,\tau}\|_2^2 = \frac{n_A}{P_{B,\tau} L_{B,\tau}}$  have been used to obtain the above. This completes the proof. ■

### D.0.15 Proof of Lemma 4

Recall that the approximate MSE  $\mathbb{E} \left\| V_m - \hat{V}_{m,\text{approx}} \right\|_F^2 = \frac{2n_A-1}{2P_{B,\tau}L_{B,\tau}} \mathbb{E} \left[ \frac{1}{\phi_c} \sum_{i=1}^m \frac{1}{\sigma_i^2 d_i^2} \right]$ . Now, the problem in (5.14) can be rewritten as

$$\min_{\phi_c > 0, d_i \geq 0} \mathbb{E} \left\| V_m - \hat{V}_{m,\text{approx}} \right\|_F^2, \text{ such that } \sum_{i=1}^m d_i^2 = 1, \text{ and } \mathbb{E} \phi_c = 1. \quad (\text{D.10})$$

Now, without loss of optimality, we can first optimize  $d_i$  first for a given  $\phi_c$ , substitute the optimal  $d_i$  into the objective function, and then optimize  $\phi_c$  subject to the average power constraint. Using the Lagrange multiplier method, the solution for the optimal  $d_i$  is given by

$$d_i = \sqrt{\frac{\sigma_i^{-1}}{\sum_{k=1}^m \sigma_k^{-1}}}, \quad (\text{D.11})$$

and the corresponding approximate MSE is given by  $\frac{2n_A-1}{2P_{B,\tau}L_{B,\tau}} \mathbb{E} \left( \frac{\sum_{i=1}^m \sigma_i^{-1}}{\sqrt{\phi_c}} \right)^2$ . Note that, when  $\phi_c = 1$ , this corresponds to the MSE for temporally-constant power training.

Now, we solve the following problem:

$$\min_{\phi_c: \mathbb{E} \phi_c \leq 1} \mathbb{E} \left\{ \frac{(\sum_{i=1}^m \sigma_i^{-1})^2}{\phi_c} \right\}. \quad (\text{D.12})$$

Due to the convexity of the problem, using variational calculus [67], the solution is

$$\phi_c = \frac{\sum_{i=1}^m \sigma_i^{-1}}{\mathbb{E} \sum_{i=1}^m \sigma_i^{-1}}, \quad (\text{D.13})$$

and the corresponding approximate MSE is given by

$$\mathbb{E} \left\| V_m - \hat{V}_{m,\text{approx}} \right\|_F^2 = \frac{2n_A-1}{2P_{B,\tau}L_{B,\tau}} \left( \mathbb{E} \sum_{i=1}^m \sigma_i^{-1} \right)^2. \quad \blacksquare \quad (\text{D.14})$$

### D.0.16 Proof of Theorem 10

We need to show that as  $P_{B,\tau}, P_{A,d} \rightarrow \infty$  with  $\frac{P_{A,d}}{P_{B,\tau}} = \mu$ ,  $|C_{\text{LB}} - C_{\text{LB,a}}| \rightarrow 0$ , where  $C_{\text{LB}}$  and  $C_{\text{LB,a}}$  are as defined in (5.22) and (5.23), respectively. From (5.22), we have

$$\frac{1}{m} \mathbb{E}_{|H} \{ \|\tilde{\mathbf{w}}_{\text{eff}}\|_F^2 \} = 1 + \frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\{ \left\| \sum_m V^H \mathbb{E}_{|H} \{ V_e \} - \sum_m V^H V_e \right\|_F^2 \right\}. \quad (\text{D.15})$$

Using the above and the definition of  $G \triangleq \Sigma_{m,m} - \sum_m V^H \mathbb{E}_{|H} \{ V_e \}$ , after some algebraic manipulation,  $C_f \triangleq \frac{C_{\text{LB}}}{\alpha}$  can be written as

$$C_f = C_1 - m \mathbb{E} \log_2 \left( 1 + \frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\{ \left\| \sum_m V^H \mathbb{E}_{|H} \{ V_e \} - \sum_m V^H V_e \right\|_F^2 \right\} \right), \quad (\text{D.16})$$

where

$$C_1 \triangleq \mathbb{E} \log_2 |I_m + X| + \mathbb{E} \log_2 |I_m + (I_m + X)^{-1} \Gamma|. \quad (\text{D.17})$$

In the above, for the ease of presentation, we have defined

$$\Gamma \triangleq \frac{P_{A,d}}{m} \left\{ -2\Re \left[ \sum_{m,m} \mathbb{E}_{|H} \{ V_e^H \} V \Sigma_m^H \right] + \sum_m V^H \mathbb{E}_{|H} \{ V_e \} \mathbb{E}_{|H} \{ V_e^H \} V \Sigma_m^H \right\} \quad (\text{D.18})$$

and

$$X \triangleq \frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\{ \left\| \sum_m V^H \mathbb{E}_{|H} \{ V_e \} - \sum_m V^H V_e \right\|_F^2 \right\} I_m + \frac{P_{A,d}}{m} \sum_{m,m} \Sigma_{m,m}^H, \quad (\text{D.19})$$

where  $\mathbb{E}_{|H}\{\cdot\}$  denotes the expectation conditioned on the channel  $H$ . Using (D.17) in (D.16), we get

$$C_f = \mathbb{E} \log_2 \left| I_m + \frac{\frac{P_{A,d}}{m} \Sigma_{m,m} \Sigma_{m,m}^H}{1 + \frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\| \Sigma_m V^H \mathbb{E}_{|H} \{V_e\} - \Sigma_m V^H V_e \right\|_F^2} \right| + \mathbb{E} \log_2 |I_m + (I_m + X)^{-1} \Gamma|. \quad (\text{D.20})$$

Now, the proof would be complete, if we could establish that:

$$\textbf{Claim 1: } \frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\| \Sigma_m V^H \mathbb{E}_{|H} \{V_e\} - \Sigma_m V^H V_e \right\|_F^2 \rightarrow \sigma_{\text{eff}}^2 \quad (\text{D.21})$$

as  $P_{A,d}, P_{B,\tau} \rightarrow \infty$ , with  $\frac{P_{A,d}}{P_{B,\tau}} = \mu > 0$ .

$$\textbf{Claim 2: } \mathbb{E} \log_2 |I_m + (I_m + X)^{-1} \Gamma| \rightarrow 0 \quad (\text{D.22})$$

as  $P_{A,d}, P_{B,\tau} \rightarrow \infty$  with  $\frac{P_{A,d}}{P_{B,\tau}} = \mu > 0$ .

### Proof of Claim 1

Recall from (5.9) of Theorem 9 that  $V_e \triangleq V_m - \hat{V}_m = E + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right)$ . Since  $\mathbb{E}_{|H} \{E\} = 0$ , we have  $\mathbb{E}_{|H} \{V_e\} = \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right)$ . Using this in the expression in **Claim 1**, we get,

$$\frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\| \Sigma_m V^H \mathbb{E}_{|H} \{V_e\} - \Sigma_m V^H V_e \right\|_F^2 = \frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \|\Delta\|_F^2 + \frac{P_{A,d}}{m^2} \mathcal{O}\left(\frac{1}{(P_{B,\tau} L_{B,\tau})^{3/2}}\right),$$

where  $\Delta \triangleq \Sigma_m V^H E \in \mathbb{C}^{m \times m}$ . First, we compute an expression for  $\frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \{\|\Delta\|_F^2\}$  as follows. The  $(i, j)^{\text{th}}$  entry of  $\Delta$  is  $\Delta_{ij} = \sigma_i \mathbf{v}_i^H \mathbf{e}_j$ , with  $\mathbf{e}_j$  representing the  $j^{\text{th}}$  column of  $E$  defined in Theorem 9. Using this, it can be shown that the  $(i, j)^{\text{th}}$  element of  $\Delta$  can

be written as

$$\Delta_{ij} = \begin{cases} \frac{1}{d_i \sqrt{\phi_c}} \sqrt{-1} \Im \{ \mathbf{v}_i^H \mathbf{w}_{i,A,\tau} \} & i = j, 1 \leq i, j \leq m, \\ \frac{\sigma_i}{\sigma_j d_j \sqrt{\phi_c}} \mathbf{v}_i^H \mathbf{w}_{j,A,\tau} & i \neq j, 1 \leq i, j \leq m. \end{cases} \quad (\text{D.23})$$

From (D.23),

$$\mathbb{E}_{|H} \{ |\Delta_{ij}|^2 \} = \begin{cases} \frac{1}{2P_{B,\tau} L_{B,\tau} d_j^2 \phi_c} & i = j, 1 \leq i, j \leq m, \\ \frac{1}{P_{B,\tau} L_{B,\tau}} \frac{\sigma_i^2}{\sigma_j^2 d_j^2 \phi_c} & i \neq j, 1 \leq i, j \leq m. \end{cases} \quad (\text{D.24})$$

From the above, it follows that  $\frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \{ \|\Delta\|_F^2 \} = \sigma_{\text{eff}}^2$ . Since

$$\frac{P_{A,d}}{m} \mathcal{O} \left( \frac{1}{(P_{B,\tau} L_{B,\tau})^{3/2}} \right) = \mathcal{O} \left( \frac{\mu}{\sqrt{P_{B,\tau} L_{B,\tau}}} \right) \rightarrow 0$$

as  $P_{A,d}, P_{B,\tau} \rightarrow \infty$  such that  $\frac{P_{A,d}}{P_{B,\tau}} = \mu > 0$ , we get

$$\frac{P_{A,d}}{m^2} \mathbb{E}_{|H} \left\{ \left\| -\sum_m V^H \mathbb{E}_{|H} \{ V_e \} + \sum_m V^H V_e \right\|_F^2 \right\} \rightarrow \sigma_{\text{eff}}^2 = \frac{\mu}{L_{B,\tau} m^2} \sum_{i=1}^m \frac{\beta_i}{d_i^2 \phi_c},$$

which proves **Claim 1**.

### Proof of Claim 2

First, note that  $\Gamma$  converges to a finite constant as  $P_{A,d}, P_{B,\tau} \rightarrow \infty$  such that  $\frac{P_{A,d}}{P_{B,\tau}} = \mu > 0$  since  $\mathbb{E}_{|H} \{ V_e \} = \mathcal{O} \left( \frac{1}{P_{B,\tau} L_{B,\tau}} \right)$ . From **Claim 1**, and from the definition of  $X$ , it is clear that  $(I_m + X)^{-1} \rightarrow 0$ , as  $P_{A,d}, P_{B,\tau} \rightarrow \infty$ . Thus, we have  $(I_m + X)^{-1} \Gamma \rightarrow 0$  at the rate of  $\frac{1}{P_{A,d}}$ . Since  $\log_2 |\cdot|$  is continuous, from (D.20), we have,

$$\left| C_f - \mathbb{E} \log_2 \left| I_m + \frac{P_{A,d}}{m} \frac{\Sigma_{m,m} \Sigma_{m,m}^H}{1 + \sigma_{\text{eff}}^2} \right| \right| \rightarrow 0 \quad (\text{D.25})$$

as  $P_{A,d}, P_{B,\tau} \rightarrow \infty$ , with  $\frac{P_{A,d}}{P_{B,\tau}} = \mu > 0$ , which completes the proof. ■

### D.0.17 Proof of Theorem 11

Maximizing  $C_{\text{LB},a}$  given in Theorem 10 with respect to  $D$  is equivalent to solving the following optimization problem:

$$\min_{d_i \geq 0: \sum_{i=1}^m d_i^2 \leq 1} \sum_{i=1}^m \frac{\beta_i}{d_i^2}. \quad (\text{D.26})$$

The solution in (5.26) now follows directly by noting that the above objective function is convex in  $(d_1, \dots, d_m)$ , and using the Lagrangian multiplier method. The resulting expression for  $C_{\text{LB},a}$  is given by

$$C_{\text{LB},a} = \frac{L_c - L_{B,\tau}}{L_c} \sum_{i=1}^m \mathbb{E} \log_2 \left( 1 + \frac{P_{A,d}}{m} \frac{\sigma_i^2 \phi_c}{\tau + \phi_c} \right), \quad (\text{D.27})$$

where  $\tau \triangleq \frac{P_{A,d}}{P_{B,\tau} L_{B,\tau} m^2} (\sum_{k=1}^m \sqrt{\beta_k})^2$ . Since the objective functional in (D.27) is concave in  $\phi_c$ , and the constraint is convex, we get the necessary and sufficient condition in (5.27) by differentiating the Lagrangian and equating it to zero, and solving for  $\lambda$ . ■

### D.0.18 Proof of Theorem 12

Substituting for  $R_1$  to  $R_M$  from (5.29) with  $m = 1$ ,  $\mathbb{E} \max\{R_1, \dots, R_M\}$  can be written as

$$\mathbb{E} \max\{R_1, \dots, R_M\} = \int_0^\infty \Pr \left\{ \alpha \log_2 \left( 1 + \frac{P_{A,d}}{1 + \sigma_{\text{eff}}^2} \max_{1 \leq i \leq M} \sigma_{1,i}^2 \right) > x \right\} dx \quad (\text{D.28})$$

$$= \int_0^\infty \Pr \left\{ \max_{1 \leq i \leq M} \sigma_{1,i}^2 > \omega \right\} dx \quad (\text{D.29})$$

$$= \int_0^\infty (1 - (\Pr \{\sigma_{1,1}^2 \leq \omega\})^M) dx, \quad (\text{D.30})$$

where  $\omega$  is as defined in Theorem 12, and the last equality follows since the singular values are i.i.d. across users. Now, we need to find an expression for  $\Pr \{\sigma_{1,1}^2 \leq \omega\}$ .

From [62], the pdf of  $\sigma_{1,1}^2$ , denoted  $f_{\sigma_{1,1}^2}(y)$ , is given by

$$f_{\sigma_{1,1}^2}(y) = \frac{1}{\prod_{k=1}^{n_B} (n_B - k)!(n_A - k)!} \sum_{j=1}^{n_B} \sum_{p=n_A-n_B}^{(n_A+n_B)j-2j^2} c_{j,p} p! e^{-jy} y^p, \quad (\text{D.31})$$

where the coefficients  $c_{j,p}$  are as described in [62]. Using the identity [68]

$$\int e^{ax} x^q dx = e^{ax} \sum_{l=0}^q (-1)^l \frac{q! x^{q-l}}{(q-l)! a^{l+1}}, \quad (\text{D.32})$$

we get the desired result. ■

### D.0.19 Proof of Theorem 13

Substituting for  $R_k$  from (5.29), and with  $m = 1$ ,  $\mathbb{E} \max\{R_1, \dots, R_M\}$  can be written as

$$\mathbb{E} \max\{R_1, \dots, R_M\} = \alpha \mathbb{E} \log_2 \left( 1 + \frac{P_{A,d}}{1 + \sigma_{\text{eff}}^2} \max\{\sigma_{1,1}^2, \dots, \sigma_{1,M}^2\} \right) \quad (\text{D.33})$$

$$\leq \alpha \log_2 \left( 1 + \frac{P_{A,d}}{1 + \sigma_{\text{eff}}^2} \mathbb{E} \max\{\sigma_{1,1}^2, \dots, \sigma_{1,M}^2\} \right), \quad (\text{D.34})$$

where  $\sigma_{\text{eff}}^2$  is as defined in (5.30), and (D.34) follows from the Jensen's inequality. Now,

we find an upper bound on  $\mathbb{E} \max\{\sigma_{1,1}^2, \dots, \sigma_{1,M}^2\}$  as follows. Pick  $s > 0$ , and consider

$$\exp \left\{ s \mathbb{E} \max_{1 \leq i \leq M} \sigma_{1,i}^2 \right\} \stackrel{(a)}{\leq} \mathbb{E} \exp \left\{ s \max_{1 \leq i \leq M} \sigma_{1,i}^2 \right\} \quad (\text{D.35})$$

$$= \int_0^\infty \Pr \left\{ \exp \left\{ s \max_{1 \leq i \leq M} \sigma_{1,i}^2 \right\} > x \right\} dx \quad (\text{D.36})$$

$$\stackrel{(b)}{\leq} \int_0^\infty \sum_{k=1}^M \Pr \left\{ \exp \left\{ s \sigma_{1,k}^2 \right\} > x \right\} dx \quad (\text{D.37})$$

$$\stackrel{(c)}{=} M \int_0^\infty \Pr \left\{ \exp \left\{ s \sigma_{1,1}^2 \right\} > x \right\} dx \quad (\text{D.38})$$

$$\leq M \mathbb{E} \left\{ \exp \left\{ s \sigma_{1,1}^2 \right\} \right\} \quad (\text{D.39})$$

$$\stackrel{(d)}{\leq} M \mathbb{E} \left\{ \exp \left\{ s \|H_1\|_F^2 \right\} \right\}. \quad (\text{D.40})$$

In the above, (a) follows from the Jensen's inequality, (b) follows from the union bound, (c) follows from the fact that  $\sigma_{1,i}$ 's are i.i.d., and finally (d) follows from the fact that  $\sigma_{1,1}^2 \leq \|H_1\|_F^2$ . Now, we evaluate  $\mathbb{E} \left\{ \exp \left\{ s \|H_1\|_F^2 \right\} \right\}$ . Using the fact that  $\|H\|_2^2$  is a chi-square random variable with  $2n_A n_B$  degrees of freedom, we have

$$\mathbb{E} \left\{ \exp \left\{ s \|H_1\|_F^2 \right\} \right\} = \frac{1}{(n_A n_B - 1)!} \int_0^\infty x^{n_A n_B - 1} e^{-(1-s)x} dx \quad (\text{D.41})$$

$$= \frac{1}{(1-s)^{n_A n_B}}, s \in (0, 1) \quad (\text{D.42})$$

Using the above in (D.40), and taking the logarithm on both sides, we get

$$\mathbb{E} \max_{1 \leq i \leq M} \sigma_{1,i}^2 \leq \inf_{s \in (0,1)} \left[ \frac{\log M}{s} - \frac{n_A n_B \log(1-s)}{s} \right]. \quad (\text{D.43})$$

Substituting the above in (D.34), we get (5.34). Choosing  $s = \frac{1}{2}$  in (5.34), we get

$$R_{\text{avg}} \leq R_{\text{avg}}^u \triangleq \alpha \log_2 \left( 1 + \frac{P_{A,d}}{1 + \sigma_{\text{eff}}^2} [2 \log M + 2n_A n_B \log 2] \right). \quad (\text{D.44})$$

For large  $M$ , it is easy to see that  $\lim_{M \rightarrow \infty} \frac{R_{\text{avg}}^u}{\alpha \log_2 \log M} = 1$ . This completes the proof.

# Bibliography

- [1] C. Steger and A. Sabharwal, "Single-input two-way SIMO channel: Diversity-multiplexing tradeoff with two-way training," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4877–4885, Dec. 2008.
- [2] X. J. Zhang, Y. Gong, and K. B. Letaief, "Power control and channel training for MIMO channels: A DMT perspective," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2080–2089, Jul. 2011.
- [3] C. Steger, A. Khoshnevis, A. Sabharwal, and B. Aazhang, "The case for transmitter training," in *International Symp. on Inf. theory (ISIT)*, Sep. 2007.
- [4] D. Samardzija and N. Mandayam, "Impact of pilot design on achievable data rates in multiple antenna multiuser TDD systems," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1370–1379, Sep. 2007.
- [5] R. Venkataramani and T. L. Marzetta, "Reciprocal training and scheduling protocol for MIMO systems," in *Proc. 41st Annual Allerton Conf. Commun., Control and Computing*, Oct. 2003, pp. 304–304.
- [6] T. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.

- [7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecom.*, vol. 10, no. 4, pp. 585–595, Dec. 1999.
- [8] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [9] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, May 1999.
- [10] B. N. Bharath and C. R. Murthy, "On the improvement of diversity-multiplexing gain tradeoff in a training based TDD-SIMO system," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, Mar. 2010, pp. 3366–3369.
- [11] —, "On the DMT of TDD-SIMO systems with channel-dependent reverse channel training," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3332–3341, Nov. 2012.
- [12] —, "Power controlled reverse channel training achieves an infinite diversity order in a TDD-SIMO system with perfect CSIR," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1800–1803, Nov. 2012.
- [13] —, "Channel training signal design for reciprocal multiple antenna systems with beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 140–151, Jan. 2013.
- [14] —, "Channel estimation at the transmitter in a reciprocal MIMO spatial multiplexing system," in *National Conf. on Commun. (NCC)*, 2012, pp. 1–5.
- [15] —, "Reverse Channel Training in a Multi-User TDD-MIMO Spatial Multiplexing System," *Submitted to IEEE Trans. on Veh. Technol.*, Jan. 2013.

- [16] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [17] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1273–1289, Apr. 2001.
- [18] H. El Gamal, G. Caire, and M. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.
- [19] A. Khoshnevis and A. Sabharwal, "On diversity and multiplexing gain of multiple antenna systems with transmitter channel information," in *Proc. Allerton Conf. on Commun., Control and Computing*, 2004.
- [20] H. N. Raghava and V. Sharma, "Diversity-multiplexing trade-off for channels with feedback," in *Allerton Conf. Commun., Control, and Computing*, 2005, pp. 668–677.
- [21] L. Zheng, "Diversity-multiplexing tradeoff: A comprehensive view of multiple antenna systems," Ph.D. dissertation, University of California, Berkeley, 2002.
- [22] A. Lim and V. Lau, "On the fundamental tradeoff of spatial diversity and spatial multiplexing of MIMO links with imperfect CSIT," in *IEEE Int. Symp. on Info. Theory*, 2006, pp. 2704–2708.
- [23] —, "On the fundamental tradeoff of spatial diversity and spatial multiplexing

- of MISO/SIMO links with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 110–117, Jul. 2008.
- [24] T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1618–1626, Apr. 2009.
- [25] X. J. Zhang, Y. Gong, and K. B. Letaief, "On the diversity gain in cooperative relaying channels with imperfect CSIT," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1273–1279, Apr. 2010.
- [26] A. Khoshnevis and A. Sabharwal, "Performance of quantized power control in multiple antenna systems," in *IEEE International Conf. on Commun.*, vol. 2, 2004, pp. 803–807.
- [27] T. Kim and M. Skoglund, "Diversity-multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2743–2759, Aug. 2007.
- [28] V. Aggarwal and A. Sabharwal, "Bits about the channel: multiround protocols for two-way fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3352–3370, Jun. 2011.
- [29] X. J. Zhang, Y. Gong, and K. B. Letaief, "On the diversity gain in MIMO channels with joint rate and power control based on noisy CSITR," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 68–72, Jan. 2011.
- [30] B. N. Bharath and C. R. Murthy, "Reverse channel training for reciprocal MIMO systems with spatial multiplexing," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, Apr. 2009, pp. 2673–2676.

- [31] X. Zhou, T. Lamahewa, P. Sadeghi, and S. Durrani, "Two-way training: optimal power allocation for pilot and data transmission," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 564–569, Feb. 2010.
- [32] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3233–3258, Jul. 2006.
- [33] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869–3884, Sep. 2006.
- [34] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.
- [35] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.
- [36] V. Aggarwal, G. Krishna, S. Bhashyam, and A. Sabharwal, "Two models for noisy feedback in MIMO channels," in *42nd Asilomar Conf. on Signals, Systems and Computers*, oct. 2008, pp. 718–722.
- [37] T. A. Lamahewa, P. Sadeghi, and X. Zhou, "On lower bounding the information capacity of amplify and forward wireless relay channels with channel estimation errors," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 2075–2079, Jul. 2011.

- [38] V. Sharma, K. Premkumar, and R. Swamy, "Exponential diversity achieving spatio-temporal power allocation scheme for fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 188–208, Jan. 2008.
- [39] J. C. Roh and B. D. Rao, "Design and analysis of mimo spatial multiplexing systems with quantized feedback," *IEEE Trans. on Sig. Proc.*, vol. 54, no. 8, pp. 2874–2886, Aug. 2008.
- [40] M. R. B. Shankar and K. V. S. Hari, "On the variations in mutual information of MIMO communication systems due to perturbed channel state information at transmitter," *IEEE Trans. Commun.*, vol. 54, no. 9, pp. 1593–1603, Sep. 2006.
- [41] D. J. Love, "Duplex distortion models for limited feedback MIMO communication," *IEEE Trans. on Sig. Proc.*, vol. 54, no. 2, pp. 766–774, Feb. 2006.
- [42] U. Salim, D. Gesbert, and D. Slock, "Combining training and quantized feedback in multiantenna reciprocal channels," *IEEE Trans. on Signal Proc.*, vol. 60, no. 3, pp. 1383–1396, Mar. 2012.
- [43] U. Salim and D. T. M. Slock, "How much feedback is required for TDD multi-antenna broadcast channels with user selection?" *EURASIP Journal on Advances in Sig. Proc.*, May 2010.
- [44] T. L. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. on Sig. Proc.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.
- [45] L. P. Withers, R. M. Taylor, and M. W. David, "Echo-MIMO: A two-way channel

- training method for matched cooperative beamforming," *IEEE Trans. on Sig. Proc.*, vol. 56, no. 9, pp. 4419–4432, Sep. 2008.
- [46] J. Chen and D. T. M. Slock, "Comparison of two analog feedback schemes for transmit side mimo channel estimation," in *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, 2007, pp. 1–5.
- [47] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 9, pp. 2424–2431, Sep. 2004.
- [48] S. Gazor and K. AlSuhaili, "Communications over the best singular mode of a reciprocal MIMO channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1993–2001, Jul. 2010.
- [49] Y. Tang, B. Vucetic, and Y. Li, "An iterative singular vectors estimation scheme for beamforming transmission and detection in MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 6, pp. 505–507, Jun. 2005.
- [50] T. L. Marzetta and B. M. Hochwald, "Learning the channel at the transmitter," in *The 42nd Annual Allerton Conf. on Commun., Control, and Computing*, Sep. 2004.
- [51] R. Prasad, B. N. Bharath, and C. R. Murthy, "Joint data detection and dominant singular mode estimation in time varying reciprocal MIMO systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, May 2011, pp. 3240–3243.
- [52] E. Martinian *et al.*, "Waterfilling gains o  $(1/\text{SNR})$  at high SNR," *unpublished note*. Available online: [citeseer.ist.psu.edu/646493.html](http://citeseer.ist.psu.edu/646493.html), 2004.

- [53] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.
- [54] C. Chen and L. Wang, "Performance analysis of scheduling in multiuser MIMO systems with zero-forcing receivers," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1435–1445, Sep. 2007.
- [55] M. Guillaud, D. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," *Signal Processing and Its Applications*, 2005.
- [56] T. Marzetta, "How much training is required for multiuser MIMO?" in *Fortieth Asilomar Conf. on Signals, Systems and Computers*, 2006, pp. 359–363.
- [57] —, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [58] V. Shah, N. Mehta, and R. Yim, "Splitting algorithms for fast relay selection: generalizations, analysis, and a unified view," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1525–1535, Apr. 2010.
- [59] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Twenty-third Annual Joint Conf. of the IEEE Computer and Commun. Societies.*, vol. 3, 2004, pp. 1662–1672.
- [60] V. Shah, N. Mehta, and R. Yim, "Optimal timer based selection schemes," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1814–1823, Jun. 2010.

- [61] A. Lozano, A. Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.
- [62] P. Dighe, R. Mallik, and S. Jamuar, "Analysis of transmit-receive diversity in rayleigh fading," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 694–703, Apr. 2003.
- [63] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [64] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 11–15, Jan. 2007.
- [65] P. Stoica and B. C. Ng, "On the Cramer-Rao bound under parametric constraints," *IEEE Sig. Proc. Letters*, vol. 5, no. 7, pp. 177–179, Jan. 1998.
- [66] A. Stuart and K. Ord, *Kendall's Advanced Theory of Statistics, Volume 1, Distribution Theory*, 6th ed. John Wiley & Sons Inc., 1994.
- [67] D. Bertsekas, "Nonlinear programming". Athena Scientific, 1999.
- [68] E. W. Beyer, "Standard Mathematical Tables and Formulae". CRC Press London, 1991.